Ana Claudia Patrocínio

**CLASSIFICADOR AUTOMÁTICO DE ACHADOS
MAMOGRAFICOS EM IMAGENS DIGITAIS DE MAMAS
DENSAS UTILIZANDO TÉCNICAS HÍBRIDAS**

DEDALUS - Acervo - EESC

31100052153

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

Orientador: Prof. Associado Dr. Homero Schiabel
Co-Orientadora: Prof^a. Associada Dra. Roseli Ap. Francelin Romero

São Carlos
2004

Class. Tese-EESC
 Num. 4184
 Tombo T093/05
 Syano 1444881

Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca – EESC/USP

P314c Patrocínio, Ana Claudia
 Classificador automático de achados mamográficos em
 imagens digitais de mamas densas utilizando técnicas
 híbridas / Ana Claudia Patrocínio. -- São Carlos, 2004.

Tese (Doutorado) -- Escola de Engenharia de São
 Carlos-Universidade de São Paulo, 2004.

Área: Engenharia Elétrica.

Orientador: Prof. Assoc. Dr. Homero Schiabel.

Co-orientadora: Profa. Assoc. Dra. Roseli Ap.
 Francelin Romero.

1. Esquemas CAD. 2. Mamografia. 3. Classificadores.
4. Extração de atributos. 5. Processamentos de imagens.
6. Câncer de mama. I. Título.

FOLHA DE JULGAMENTO

Candidata: Tecnóloga ANA CLAUDIA PATROCINIO

Tese defendida e julgada em 22-12-2004 perante a Comissão Julgadora:

Prof. Associado **HOMERO SCHIABEL** (Orientador)
(Escola de Engenharia de São Carlos/USP) APROVADA

Profa. Associada **ROSELI APARECIDA FRANCELIN ROMERO** (Co-Orientadora)
(Instituto de Ciências Matemáticas e de Computação/USP) APROVADA

Prof. Dr. **JOÃO FERNANDO MARAR**
(UNESP/Bauru) APROVADA

Prof. Dr. **LEONARDO VIDAL BATISTA**
(Universidade Federal da Paraíba/UFPB) APROVADA

Prof. Dr. **JOSÉ MORCELI**
(UNESP/Botucatu) APROVADA

Prof. Associado **HOMERO SCHIABEL**
Vice-Coordenador em exercício do Programa de Pós-Graduação
em Engenharia Elétrica

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

Este trabalho teve o apoio financeiro da FAPESP – Fundação de Amparo à
Pesquisa do Estado de São Paulo - processo nº 00/08965-5

*Este trabalho é dedicado com amor aos meus avós Odila e Orlando, que
também me ensinaram a caminhar e a amar.*

*E a meus pais, Neusa e Guilherme, e minha irmã Carla, por
caminharmos juntos.*

AGRADECIMENTOS

Meus sinceros agradecimentos a meus orientadores Prof. Dr. Homero Schiabel e Profa. Dra. Roseli Ap. Francelin Romero, pela colaboração e dedicação para a realização deste trabalho, confiança, apoio e amizade durante todos esses anos - Vocês foram fundamentais no cumprimento desta minha jornada. *"Os princípios mais importantes podem e devem ser inflexíveis."* (Abraham Lincoln).

A meus pais, meus avós e minha irmã, por me amarem e compreenderem minhas ausências, obrigada pelo incentivo e apoio - Vocês são responsáveis pela minha perseverança e minhas conquistas

A toda minha família especialmente ao meu tio, o BOIA, Letícia e Piqui, Flávia e Tania, pela alegria e boa convivência, me fortalecendo para enfrentar a vida durante todos estes anos longe de casa.

Ao Paulo A. Ferreira, pelo carinho e prontidão em ajudar sempre, também por ajudar nas correções da tese - Obrigada de coração.

Ao Leandro Patrocínio de Freitas, meu primo querido, por sua alegria, agradável companhia, auxílio durante os momentos de trabalho árduo e dedicação em minhas longas noites de trabalho - Lelê, você é muito especial e tem sido meu anjo da guarda.

A todos meus colegas do Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas (LAPIMO), especialmente Michele, Vivian, Ana (Embrapa), Cláudio, César, Marcelo (Mogi), Maurício, Fabinho, Patrícia, Hamilton, Adriana, Eveline, Anderson e Hellien. *"... As únicas coisas que importam são as feitas de verdades e alegrias e não as de lata e vidro..."* (Richard Bach). E o último que sair apague a luz e desligue a cafeteira....

À minha amiga Vivian Toledo Santos, pela amizade, por tudo que passamos juntas nas noites de trabalho no LAPIMO, contatos e dedicação na coleta de dados em Botucatu e por me auxiliar na elaboração dos testes e da própria tese - Vi, a gente só não divide namorado porque está difícil, mas até "veneninho" entra na partilha.

À minha amiga Michele Fúlvia Ângelo, pelo incentivo, desabafos, ótima convivência e pelas muitas risadas em nossos almoços e suas estadias em minha casa - Amiga, "tô boba", acabei!

A meu amigo, "praticamente Dr.", Marcelo Andrade da Costa Vieira, pela ótima convivência nos alegrando no LAPIMO, discussões ao longo do mestrado e doutorado e por me socorrer com as burocracias dos relatórios FAPESP - Mogi, você sempre "firmou o gorpe no 1... 2... e... já!".

A meu amigo César H. M. Santaella, por trabalharmos juntos nas *Wavelets* e pelos bons momentos compartilhados.

A meu amigo Fábio Lagoeiro, que não me deixava render pelo cansaço nas noites compridas de LAPIMO, pelas discussões, palavras certas nas horas exatas e por sua fé.

A meus amigos Maurício, Juliana e Matheus, por todos estes anos de convivência e bons momentos compartilhados - Vocês são especiais!

A meu amigo Cláudio Eduardo Goes, por sua prontidão em me socorrer durante o desenvolvimento deste trabalho e pelos momentos agradáveis no LAPIMO.

Ao Dr. José Morceli, do Hospital das Clínicas de Botucatu, por facilitar nossa coleta de dados, permitir nossa participação em reuniões de discussão de casos junto ao corpo clínico do hospital e as valiosas informações sobre o diagnóstico do câncer de mama.

À Profa. Dra. Regina Bitelli Medeiros da UNIFESP, pela digitalização das imagens fornecidas pelo Hospital São Paulo.

À Profa. Dra. Ana Maria Marques Silva, pela atenção, sugestões e longas discussões por e-mail.

À Profa. Dra. Ana Maria Marques Silva, pela atenção, sugestões e longas discussões por e-mail. *"Um sorriso caloroso é a linguagem universal da gentileza"* (William Arthur Ward).

À minha amiga Dra. Evelina M. A. Neves, pela amizade, carinho e incentivo - Eve, você é meu exemplo de garra e perseverança.

À minha amiga Sandra C. Pereira Lopes, pelas imensas contribuições com a língua inglesa e demonstrações de amizade e companheirismo na minha estadia em São Carlos. Muito Obrigada de Coração.

À Dra. Renata A T. Arantes: pela boa convivência dividindo apartamento durante tantos anos em São Carlos.

Às minhas amigas irmãs Frederica B. Lorenzetti e Patricia B. Grizotto, pelo amor, respeito e carinho que sentimos. Obrigada por nossa amizade. *"Amigo é aquele que te diz eu te amo sem qualquer medo de má interpretação; amigo é quem te ama e ponto. É verdade e razão, sonho e sentimento. Amigo é para sempre, mesmo que o sempre não exista"*.

A meus amigos de Batatais, especialmente: Juliana, Adriana (Tata) e Eduardo (Gambá), Zilda e Bosco, Luciana, Taisa e Breno, Marcelão, Graziela, Serginho e minha afilhada Isabela, que me proporcionam tantos momentos divertidos e agradáveis, renovando minha energia. Obrigada pelo carinho. *"O mundo está nas mãos daqueles que têm coragem de sonhar, e correr o risco de viver seus sonhos. Cada qual com seu talento"*.

Às minhas amigas Veridiana, Valéria e Mércia, obrigada pela admiração e carinho especial que sentem por mim, hospitalidade em Bauru e por acreditarem que eu chegaria até aqui - Vocês fazem parte desta minha conquista.

A todos meus amigos conquistados em São Carlos, especialmente ao Dr. Paulo W. Tardioli e ao Dr. Edmário Belatti, pelos momentos de descontração e amizade. *"A vida é muito curta para ser pequena"*.

Aos professores e funcionários do Departamento de Engenharia Elétrica da EESC - USP, especialmente Cida, Rui Bertho, José Luis, Jussara, Vera, Lia, Maria e ao Prof. Adilson. Vocês estarão para sempre em minha memória. *"O prazer dos grandes homens consiste em tornar os outros mais felizes"* (Pascal).

Às secretárias da Pós-graduação: Marisa e Denise, por deixarem nossos compromissos menos árduos, pela amizade e atenção. *"Comece fazendo o que é necessário, depois o que é possível, e de repente você estará fazendo o impossível"* (São Francisco de Assis).

Aos cientistas aos quais atribuo a responsabilidade por minha escolha pela área acadêmica: Dra. Berenice Borges Lorenzetti e Dr. João Fernando Marar, pelos exemplos de amor a vida e a pesquisa. *"... O Mestre na arte da vida faz pouca distinção entre o seu trabalho e o seu lazer, entre a sua mente e o seu corpo, entre a sua educação e a sua recreação, entre o seu amor e a sua religião. Ele dificilmente sabe distinguir um corpo do outro. Ele simplesmente persegue sua visão de excelência em tudo que faz, deixando para os outros a decisão de saber se está trabalhando ou se divertindo. E ele adia que está sempre fazendo as duas coisas simultaneamente"* (Texto Zen Budista).

À FAPESP, pelo apoio financeiro - proc. Nº.: 00/08965-5.

A DEUS, por minha Vida e capacidade de crescer, amar e usar todas minhas habilidades para meu bem estar e felicidade.

SUMÁRIO

<i>Lista de Figuras</i>	<i>iv</i>
<i>Lista de Tabelas</i>	<i>vi</i>
<i>RESUMO</i>	<i>viii</i>
<i>ABSTRACT</i>	<i>ix</i>
<i>1 Introdução</i>	<i>1</i>
1.1 Esquema CAD e o Câncer de Mama	1
1.2 Motivação	3
1.3 Objetivos	4
1.4 Disposição do trabalho	5
<i>2 Imagem Mamográfica, Achados e a Interpretação Diagnóstica</i>	<i>6</i>
2.1 Mamografia	6
2.2 Interpretação mamográfica segundo especialistas	11
2.2.1 A questão das mamas densas	19
2.3 Padronização BI-RADS	23
2.3.1 Nódulo	25
2.3.2 Calcificações	26
2.3.3 Achados Associados	28
2.3.4 Localização da Lesão	28
2.4 A Imagem Mamográfica Digital	30
2.5 Conclusão	32
<i>3 Extração e Análise de Atributos</i>	<i>33</i>
3.1 Atributos extraídos da imagem	35
3.1.1 A teoria dos Momentos	36
3.1.1.1 Os Momentos invariantes de Hu	42
3.1.1.2 Interpretação geométrica dos Momentos	45
3.1.2 Outros atributos de forma	49
3.1.3 Atributos de intensidade e descrição de Textura	50
3.1.3.1 Atributos utilizados para descrição de imagens mamográficas	53

3.2 Análise e seleção de características	56
3.3 Métodos para análise e processamento de imagens médicas	58
3.4 Conclusão	61
4 Classificadores Automáticos	62
4.1 Análise Discriminante	64
4.2 Redes Neurais Artificiais	65
4.2.1 Algoritmo <i>Backpropagation</i>	67
4.2.2 ART	72
4.2.3 Modelo de "Kohonen"	73
4.3 Redes Bayesianas	74
4.4 "Clusterização" – Classificador <i>K-Means</i>	78
4.4.1 Matrizes de espalhamento e medidas de dispersão	78
4.4.2 "Clusterização" Não-Hierárquica	79
4.4.3 <i>K-Means</i>	80
4.5 Análise de resultados da classificação	81
4.6 Classificadores aplicados em esquemas CAD	86
4.7 Conclusão	106
5 Materiais e Métodos	108
5.1 Base de Dados	109
5.2 Imagens geradas a partir de pré-processamentos e processamento de nível médio	113
5.2.1 Sistema para detecção de <i>clusters</i> de microcalcificações	113
5.2.2 Sistema para detecção de nódulos	115
5.3 Etapas de extração de atributos	119
5.3.1 Extração de atributos de intensidade	122
5.3.2 Extração dos descritores geométricos	129
5.4 Análise de atributos	134
5.5 Etapas dos Processos de Classificação	141
5.5.1 Experimentos com redes Bayesianas para classificação de <i>clusters</i> de microcalcificações	144
5.5.2 Classificação de <i>clusters</i> de microcalcificações e nódulos por Rede Neurais <i>Backpropagation</i>	146
5.5.2.1 Treinamento e testes da RNA	146
5.5.3 Classificação de nódulos por grau de intensidade	147
5.6 Apresentação dos resultados da classificação	149
6 Resultados e Discussões	153
6.1 Resultados Obtidos com as Redes Bayesianas na Classificação de <i>Clusters</i> de Microcalcificações	153
6.2 Resultados Obtidos com as Redes <i>Backpropagation</i> na Classificação de <i>Clusters</i> de Microcalcificações	156
6.3 Resultados Obtidos com Redes <i>Backpropagation</i> na	161

Classificação de Nódulos

7	<i>Considerações Finais</i>	171
7.1	Investigação da análise de lesões mamográficas segundo especialistas	171
7.2	Atributos Extraídos das Imagens Mamográficas	172
7.3	Classificação automática híbrida das lesões	173
7.4	Conclusões Finais	176
8	<i>Referências Bibliográficas</i>	179

Lista de Figuras

Figura 2.1 - Posições de exposição para exames mamográficos: a) incidência CC; b) incidência MLO	10
Figura 2.2- Sistema mamográfico	11
Figura 2.3 - Quadro de tipos lesões com suas respectivas características morfológicas (Heywang-Köbrunner, 1999).	13
Figura 2.4 Formatos de massas de acordo com o grau de suspeitabilidade (Kopans, 2000).	14
Figura 2.5 – (a)Quadro de resumo das lesões focais e (b)Continuação do quadro de resumo (Heywang-Köbrunner, 1999).	16
Figura 2.6 – Quadro de resumo da análise de lesão mal definida (Heywang-Köbrunner, 1999).	17
Figura 2.7 – (a)Quadro de resumo da análise de microcalcificações e calcificações mamárias; e (b) Continuação do quadro de resumo (Heywang-Köbrunner, 1999).	21
Figura 2.8 – (a)Quadro de resumo da análise de mama radiograficamente densa; e (b) Continuação do quadro de resumo (Heywang-Köbrunner, 1999).	24
Figura 2.9 Classificação BI-RADS da Quarta edição com as respectivas taxas de risco de câncer (Quadros, 2003).	32
Figura 3.1 - Monômios de até terceira ordem que multiplicam a função $f(x,y)$ no cálculo dos momentos (Hall, 1979).	40
Figura 3.2 – Posição padrão para Invariância à Rotação (Masters, 1994).	43
Figura 3.3 - Gráfico representando uma matriz X com $n = 3$ observações e $p=2$ variáveis. O vetor $\bar{X}$ denota as médias	61
Figura 4.1 – LDA separa o espaço em duas regiões	68
Figura 4.2 - Rede <i>Backpropagation</i>	74
Figura 4.3 – Distribuições de duas classes sob a análise de um parâmetro.	89
Figura 4.4 – Exemplo de curva ROC	90
Figura 4.5 – Rede Bayesiana para auxílio ao diagnóstico do câncer de mama (Kahn, 1997).	101
Figura 4.6 – Rede Bayesiana (Wang, 1999a)	106
Figura 5.1 - (a) e (b) interfaces do banco de imagens (Benatti, 2003)	116
Figura 5.2 – Diagrama do sistema de processamento para detecção de microcalcificações (Nunes, 2001).	119
Figura 5.3 Exemplos de RIs binarizadas com microcalcificações detectadas (Nunes, 2001)	120
Figura 5.4 (a)Imagem com as linhas de <i>Watershed</i> e junção com a imagem original; (b)imagem segmentada e seu contorno (Santos, 2003).	122
Figura 5.5 – Diagrama esquemático do sistema para detecção de nódulos	123

(Santos, 2002).

Figura 5.6 – Exemplos de RIs binarizadas com massas detectadas (Santos, 2003)	123
Figura 5.7 – Interface do processo de extração de atributos	125
Figura 5.8 – a) Histograma de imagem antes de ser normalizada – escala efetiva de níveis de cinza; b) histograma de imagem normalizada pelo fator de correção.	127
Figura 5.9 – Interface de extração de atributos de intensidade.	130
Figura 5.10 – RIs antes de serem modificadas pela análise dos atributos de intensidade em (a), (c), (e), (g), (i), (l) e (n). RIs depois da modificação em (b), (d), (f), (h), (j), (m) e (o).	133
Figura 5.11 - <i>Cluster</i> delineado dentro de RI binarizada	134
Figura 5.12 – (a) Interface com imagem de nódulo processada e (b) interface com imagem de <i>cluster</i> processada	136
Figura 5.13 – Interface mostrando o processo completo de extração de atributos para nódulos.	137
Figura 5.14 Gráfico dos coeficientes de correlação dos atributos das classes suspeitos e não suspeitos de <i>clusters</i> .	138
Figura 5.15 - Correlações entre as classes por atributo.	140
Figura 5.16 Correlações entre as categorias BI-RADS 2 e BI-RADS 3.	142
Figura 5.17 - Correlações entre as categorias BI-RADS 3 e BI-RADS 4.	142
Figura 5.18 - Correlações entre as categorias BI-RADS 4 e BI-RADS 5.	143
Figura 5.19 (a) Interface para a etapa de treinamento da Rede <i>Backpropagation</i> , configurações; (b) Interface no final do treinamento.	149
Figura 5.20 – Interface de classificação da RNA	153
Figura 5.21 Interface de “clusterização”	153
Figura 5.22 - Interface de apresentação dos resultados	155
Figura 5.28 – Esquema do Classicador Híbrido	156
Figura 6.1 - Curva ROC que representa a rede1.	164
Figura 6.2 - Curva ROC que representa a rede2.	165
Figura 6.3 - Curva ROC que representa a rede3.	165
Figura 6.4 - Curva ROC REDE1	168
Figura 6.5 - Curva ROC REDE2	169
Figura 6.6 - Curva ROC REDE3	169
Figura 6.7 - Curva ROC REDE1_densidade	173
Figura 6.8 - Curva ROC REDE2_densidade	173
Figura 6.9 - Curva ROC REDE3_densidade	174

Lista de Tabelas

Tabela - 2.1 Classificação de Michele Le Gal (1984)	19
Tabela 3.1 – Interpretação geométrica dos Momentos Centrais de segunda e Terceira Ordem	45
Tabela 4.1- Demonstração de resultados obtidos experimentalmente (Kittler, 1997)	100
Tabela 4.2 – Atributos usados na rede Bayesiana (Wang, 1999a)	105
Tabela 4.3 - avaliação dos classificadores testados para problemas médicos	117
Tabela 5.1 (a) Resumo dos dados coletados relativos as imagens da base.	117
Tabela 5.1 (b) – Continuação das informações coletadas respectivas às imagens da base.	135
Tabela 5.2 – Atributos geométricos extraídos das RIs	138
Tabela 5.3 - Medidas de dispersão entre classes	139
Tabela 5.4 - Medidas de dispersão dos atributos intra-classes	141
Tabela 5.5 - Correlações entre as classes por atributos descritores geométricos extraídos de nódulos	143
Tabela 5.6 - Medidas de dispersão intraclassas	158
Tabela 6.1 - Resultados do <i>Cross-Validation</i> da rede gerada com a base de dados contendo todos os 20 atributos.	159
Tabela 6.2 - Resultados do <i>Cross-Validation</i> da rede gerada com a base de dados contendo apenas 11 atributos.	160
Tabela 6.3 - Resultados do <i>Cross-Validation</i> da rede gerada com a base de dados contendo 5 componentes principais.	161
Tabela 6.4 – Configurações das Redes testadas	162
Tabela 6.5 - Resultados de 10 testes para a rede1	162
Tabela 6.6 - Resultados de 10 testes para a rede2	162
Tabela 6.7 - Resultados de 10 testes para a rede3	164
Tabela 6.8 - Resumo dos desempenho das 3 rede	166
Tabela 6.9 - Configurações das Redes testadas	167
Tabela 6.10 - Resultados dos 10 testes da rede1	167
Tabela 6.11 - Resultados dos 10 testes da rede2	167
Tabela 6.12 - Resultados dos 10 testes da rede3	168
Tabela 6.13 - Configurações das Redes testadas	170
Tabela 6.14 - Resultados dos 10 testes da rede1 considerando densidade	170
Tabela 6.15 - Resultados dos 10 testes da rede2 considerando densidade	171

Tabela 6.16 - Resultados dos 10 testes da rede3 considerando densidade	172
Tabela 6.17 - Resumo dos desempenho das 3 rede, mais a área sob a curva ROC	172

RESUMO

Patrocínio, A. C. (2004). *Classificador automático de achados mamográficos em imagens digitais de mamas densas utilizando técnicas híbridas*. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

Esta tese apresenta uma metodologia para classificação automática de achados mamográficos em mamas densas através de uma abordagem híbrida de classificadores e extração de atributos, como parte de um esquema computadorizado de auxílio ao diagnóstico (CAD) para mamografia. Foram implementadas duas redes *Backpropagation*. Uma responde pela classificação de *clusters* de microcalcificações, através de atributos descritores geométricos, em duas classes - suspeito e não suspeito. A outra rede classifica nódulos utilizando descritores geométricos e uma entrada com informação extraída de atributos de intensidade, produzindo na saída dois tipos de informação: presença ou não do nódulo, e constatada a presença do nódulo, classificação da Região de Interesse (RI) entre as categorias BI-RADS. As respostas de um “clusterizador” de RIs através de atributos de intensidade serviram de entrada para essa rede fornecendo uma informação de grau de densidade da RI. Uma interface foi desenvolvida para a apresentação dos resultados a fim de fornecer informações mais detalhadas da classificação e do caso analisado. Os resultados do classificador foram analisados através de análise estatística de sensibilidade e especificidade, e também por curvas ROC. Os resultados obtidos ficaram próximos a 89% de acerto total (verdadeiros-positivos mais verdadeiros-negativos) para nódulos produzindo valor de A_z superior a 0,92 e ultrapassaram 75% de acerto da classificação entre as classes BI-RADS. Na classificação dos *clusters* os acertos totais do classificador ficaram próximos de 90%, com A_z superior a 0,94. Para ambos tipos de lesões, as taxas de respostas falsas-negativas ficaram abaixo de 0,1, o que significa baixo erro em relação à não detecção da doença quando o sinal está presente. O classificador apresentado nesse trabalho é a conclusão de uma etapa importante do esquema CAD que vem sendo desenvolvido pelo grupo, além de possibilitar a disponibilização de mais uma ferramenta automática de auxílio ao diagnóstico do câncer de mama aos serviços de mamografia.

Palavras – Chave: esquemas CAD, mamografia, classificadores, extração de atributos, processamento de imagens, câncer de mama.

ABSTRACT

Patrocinio, A. C. (2004). *Automatic classifier of mammographic findings in dense breast digital images using hybrid techniques*. Tese (Doutorado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

This thesis presents a methodology for automatic classification of mamographic findings in image of dense breast through hybrid approach of classifiers and features extraction techniques, as part of a Computer-Aided Diagnostic (CAD) scheme for mammography. Two Backpropagation neural networks were implemented. One for microcalcifications clustered classification, through geometric descriptors, in two classes - suspect and non-suspect. The other neural network classifies nodules using geometric descriptors and additional information from intensity features extracted, producing in the output two kinds of information: presence or not of the nodule, and if nodule is present in the image, classification among BI-RADS categories. The result of clustering technique by using intensity features is presented as a new input to neural network, supplying density degree of image. An interface was developed for results presentation in order to supply more detailed information from the classifier outputs and of the analyzed case. The results of the classifier were analyzed through sensibility and specificity statistical analysis, and also for ROC curves. The results were close to 89% of total accuracy (positive-true plus negative-true) for nodules producing value of A_z more than 0,92 and 75% of accuracy to classification among BI-RADS categories. In the cluster classification the total accuracy is about 90%, and A_z greater than 0,94. In both kinds of lesions, negative-false result rates were below 0,1, which means low error related to the fail to detect the disease when the signal is present. The classifier presented in this work is the conclusion of an important stage of the CAD scheme that has been developed by the group, besides making possible the availability of one more automatic tool of aid to the breast cancer diagnosis to be used in mammography centers.

Keywords: CAD schemes, mammography, classifiers, features extraction, image processing, breast cancer.

1

INTRODUÇÃO

As vidas de milhares de mulheres podem ser prolongadas com esforços da ciência e com o fundamental apoio da tecnologia. Para isto basta: incentivo, dedicação, criatividade e amor à vida. Tendo em vista a alta mortalidade provocada pelo câncer de mama, o que mais mata mulheres em todo o mundo, torna-se estimulante a luta pelo desenvolvimento de ferramentas tecnológicas que possam auxiliar na prevenção e detecção desta doença.

1.1 Esquema CAD e o Câncer de Mama

O câncer de mama é uma doença de causas desconhecidas e que mata milhares de mulheres todos os anos no mundo inteiro. Ao ser diagnosticado, o câncer de mama provoca um grande impacto psicológico, pois envolve a percepção da sexualidade e a própria imagem pessoal, mais do que qualquer outro tipo de câncer, devido principalmente, às formas de tratamento, que na maioria das vezes é bastante invasivo.

Há tempos, esforços têm sido enviados para a detecção precoce do câncer de mama pois é, ainda, a única maneira de diminuir a mortalidade pela doença. Isto se dá principalmente

através da mamografia por raios-X, pois é considerado o exame de maior sensibilidade para o rastreamento do câncer de mama. Para isso esquemas computadorizados de auxílio ao diagnóstico (CAD - *Computer-Aided Diagnosis*), especiais para mamografias, vem sendo desenvolvidos com intuito de auxiliar no diagnóstico de achados mamográficos, através de digitalização e processamento das imagens para realce e detecção de estruturas, e sua análise e classificação.

Os esquemas CAD são sistemas que podem detectar anomalias e classificar estruturas e achados mamográficos com base em imagens e informações do paciente, com o objetivo principal de servir como uma segunda opinião ao médico na detecção do câncer e na decisão do diagnóstico (Doi, 1996).

Embora recente, já existem alguns estudos que mostram os ganhos com a introdução de CAD para mamografia na prática clínica. Segundo Gur *et. al.*, a taxa de detecção precoce do câncer de mama pode ser aumentada em até 20% com o uso de CAD, além de diminuir o número de indicações de pacientes a repetirem o exame num espaço de tempo menor (Gur, 2004).

Para classificar um achado mamográfico em um esquema CAD é preciso que se considerem alguns tópicos importantes, tais como: a qualidade e as técnicas de aquisição da imagem, o pré-processamento e processamento das imagens, a identificação das Regiões de Interesse (RIs), a extração das características e atributos da imagem e, eventualmente, informações de anamnese; a seleção das características que melhor representem cada padrão e, finalmente, a utilização de técnica de classificação que possa apresentar boa performance para este tipo de problema.

Nos esquemas CAD, para desenvolver um classificador cujo desempenho apresente generalização para uma grande população, é necessário ter uma base de dados com um número de casos representativo das ocorrências no cotidiano clínico. A avaliação do conjunto de casos é geralmente limitada nas pesquisas das imagens médicas. Então, é importante investigar o tamanho do conjunto de exemplos, pois o desempenho de cada classificador depende do conjunto de dados usado no seu treinamento, o que freqüentemente determina a eficiência do sistema.

1.2 Motivação

Diagnosticar lesões em um mamograma é uma tarefa muito árdua para os especialistas, pois a imagem mamográfica possui baixo contraste e diversos fatores influenciam na qualidade da imagem e consequentemente na interpretação das lesões. Estes fatores estão relacionados à calibração do equipamento mamográfico, às condições de revelação, conservação e manipulação dos filmes, à composição do parênquima mamário e, nas imagens digitais, aos fatores da digitalização. Contudo, ferramentas que possam auxiliar numa melhor visualização e interpretação de estruturas mamográficas vêm sendo desenvolvidas e tornando-se uma realidade na prática dos serviços de radiologia.

A discordância na interpretação e classificação de lesões por especialistas quanto a suspeição é bastante grande, o que implica maior número de biópsias desnecessárias, aumento do custo-benefício do exame mamográfico e, nos piores casos, a possibilidade de que a doença deixe de ser detectada (os falso-negativos). Portanto, um diagnóstico preciso depende muito da experiência do especialista, e um esquema CAD pode servir como uma segunda opinião, no sentido de auxiliá-lo no processo investigativo.

Com intuito de favorecer esse processo, nosso grupo tem desenvolvido há alguns anos pesquisas na linha de avaliação de qualidade e processamento de imagens mamográficas objetivando a elaboração de um esquema CAD em mamografia com gerenciamento da qualidade da imagem (Schiabel, 2003 e Escarpinati, 2004). Tal elaboração tem em foco desde os fatores de aquisição da imagem até a etapa de classificação e apresentação de informações sobre eventuais achados mamográficos detectados automaticamente após o processamento das imagens.

1.3 Objetivos

Este trabalho é uma parte importante do esquema CAD em desenvolvimento pelo grupo e refere-se a uma investigação das técnicas de classificação computadorizadas com a finalidade de desenvolver um esquema classificador híbrido de estruturas mamárias para auxiliar no diagnóstico do câncer de mama.

O trabalho visa principalmente classificar imagens de mamas densas, característica de mulheres jovens, antes e durante a menopausa, e aquelas em tratamento de reposição hormonal. Este padrão refere-se às mamas com predominância de tecidos fibroglandulares em relação ao adiposo. Esta característica influencia a qualidade da imagem e o diagnóstico preciso, pois a absorção de raio-X dos tecidos fibroglandulares é bem próxima à absorção das típicas lesões da mama. Isso gera uma imagem de baixo contraste, dificultando a análise da mamografia e a identificação de lesões.

Alguns trabalhos têm utilizado uma estimativa quantitativa ou qualitativa da proporção de densidade mamográfica por área da mama e associa isto com o risco do câncer, quanto maior a densidade maior o risco do câncer de mama (Gram, 1997; Blane, 2002; Zhou, 2001; Byng, 1998). A avaliação subjetiva da densidade feita por especialistas falha em 60% dos casos, portanto a quantificação objetiva é bastante interessante. Dessa maneira, a densidade poderia ser usada como mais uma informação a ser considerada pelo classificador de achados mamográficos.

Para o desenvolvimento do classificador é necessária uma investigação detalhada de técnicas de extração e seleção de características, técnicas de classificação e uma avaliação de suas performances com os conjuntos de imagens disponíveis e representativas do problema.

A partir de outros trabalhos já desenvolvidos por nosso grupo no Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas - LAPIMO, resultados dos esquemas de detecção de lesões são utilizados para o desenvolvimento do esquema de extração de atributos e classificação.

O objetivo final é mostrar a construção do classificador híbrido, usando combinação de técnicas de extração de atributos e classificação, para melhorar a classificação das lesões encontradas no mamograma e aumentar a generalização do sistema, tornando-o mais flexível e com resultados mais detalhados incluindo informações relevantes para a decisão do

diagnóstico através da padronização de laudos BI-RADS que é utilizada nos serviços de mamografia atualmente.

1.4 Disposição do trabalho

Além dessa introdução, os seguintes capítulos são apresentados:

2. Imagem Mamográfica, Achados e a Interpretação Diagnóstica - Conceitos sobre a mamografia e descrição do exame, o que se procura ao analisar uma mamografia, a interpretação das lesões e a padronização de laudos mamográficos, além de aspectos sobre a imagem digital.

3. Extração e Análise de Atributos - Tipos e descrições de diferentes técnicas de extração e análise de atributos. Descrição de técnicas adaptadas à caracterização de imagens mamográficas.

4. Classificadores Automáticos – Conceitos e descrição de processos de classificação, desde supervisionados, não supervisionados e híbridos. Aplicações de classificação automática em esquemas CAD.

5. Materiais e Métodos - Descrição dos materiais utilizados e detalhamento da metodologia desenvolvida no decorrer do trabalho.

6. Resultados e Discussões - Apresentação dos resultados obtidos com avaliação através de curvas ROC e discussões.

7. Considerações Finais - Conclusões do trabalho.

8. Referências Bibliográficas – Bibliografia citada e consultada no decorrer no trabalho.

2

IMAGEM MAMOGRÁFICA, ACHADOS E A INTERPRETAÇÃO DIAGNÓSTICA

A mamografia tem entre seus objetivos, a detecção precoce do câncer de mama. A imagem mamográfica é essencial na detecção de anomalias da mama, porém a tarefa de classificar uma anomalia através da imagem é difícil, exigindo, muitas vezes, exames complementares, como análise histológica através de exames invasivos, como a biópsia.

2.1 Mamografia

A mamografia é a das melhor técnica para rastreamento do câncer de mama e pode reduzir a mortalidade em até 30%. O padrão de mamas densas tem mostrado fortes associações com o risco do câncer de mama, porque as respectivas imagens tendem a ser grosseiras e com baixo contraste, isso tem sido associado também ao risco de erro do diagnóstico. A identificação e a investigação de mulheres que apresentam alto risco de desenvolver o câncer de mama, podem facilitar a detecção da doença (Kopans, 2000).

A sensibilidade da mamografia é alta, embora haja estudos que mostram perda de 10% a 15% dos casos de câncer detectável no exame clínico. Isso se refere geralmente aos casos de mamas densas, que são características, geralmente, de mulheres mais jovens¹: esse tipo de mama ainda é um problema para o diagnóstico precoce do câncer em razão da pequena variação de contraste entre os achados e os tecidos adjacentes que compõem a mama (Kopans, 2000).

A visibilização de estruturas de interesse em mamogramas depende de muitos fatores físicos e técnicos. A iluminação tem que ser adequada, exige negatoscópios próprios com luz mais forte, o ambiente deve estar com pouca iluminação e a imagem deve ser de boa qualidade. O aparelho mamográfico é um dos maiores responsáveis pela boa visibilização das lesões pelos médicos, pois o equipamento deve estar corretamente calibrado segundo a Portaria nº 453 de 1998, da Secretaria de Vigilância Sanitária do Ministério da Saúde, que determina as medidas corretas que garantem o funcionamento adequado do equipamento, possibilitando que se produza uma imagem de boa qualidade. Uma imagem é considerada de boa qualidade quando se consegue a melhor resolução espacial, em conjunto com a melhor resolução de contraste possível. Além disso, os filmes e sua revelação também influenciam na qualidade da imagem.

A maior dificuldade na interpretação das imagens mamográficas é o baixo contraste e também, no caso da detecção do câncer precoce, o tamanho reduzido das características de malignidade de achados como microcalcificações (Shen, 1993).

Existem técnicas de posicionamento da mama para o exame. Diante do conhecimento da anatomia e mobilidade mamária, os equipamentos atualmente permitem a utilização de técnicas de posicionamento, para que se possa explorar melhor a mama. A incidência Médio-Lateral Oblíqua (MLO) oferece a melhor oportunidade de mostrar toda a mama numa única imagem, mas para isso é preciso que o posicionamento do receptor de imagem do aparelho esteja paralelo ao plano muscular, expondo o máximo possível do tecido mamário e axilar. Já a incidência Crânio-Caudal (CC) complementa a MLO, pois o objetivo da CC é incluir todo o tecido pósteromedial, o que não é possível na visualização na MLO. O mamilo na incidência CC deve estar posicionado na linha média, ou seja, no eixo central (sobre o eixo paralelo ao tubo) do campo de exposição (Mitchell, 1988).

¹ Dados fornecidos pelo Instituto Nacional do Câncer (INCA): <http://www.inca.org.br/epidemiologia/>

Figura 2.1 - Posições de exposição para exames mamográficos: a) incidência CC; b) incidência MLO

Os sistemas mamográficos têm um acessório chamado de compartimento de compressão, através do qual a mama é comprimida por uma placa de um material radiotransparente até que se consiga a menor espessura possível.

A compressão da mama tem três finalidades principais que muito contribuem para a qualidade da imagem: com menor espessura, diminui a radiação espalhada e assim o borramento; imobiliza a mama e assim dificulta o movimento no momento da aquisição; e possibilita a região de interesse ficar mais próxima do detector de imagem (chassi com o filme) evitando a sobreposição de tecidos quando feita de maneira adequada.

O equipamento dedicado à mamografia é especial, pois a imagem deve ser de alta resolução para que se possam visibilizar as estruturas mamárias que são compostas de tecidos moles, ou seja, cuja absorção de raios-X é pequena e com densidades menores que a densidade do osso, por exemplo.

Algumas modificações no tubo de raios-X contribuem para melhorar a imagem dos tecidos moles. O alvo do tubo é de um material diferente do equipamento convencional, que geralmente usa tubo com alvo de tungstênio. O do equipamento mamográfico, porém, é de molibdênio, dadas suas características espectrais. O feixe produzido num tubo com alvo de molibdênio é menos energético que um feixe produzido num de tungstênio, o que é conveniente no caso de radiografia de mamas, devido aos tecidos que a constituem.

O campo de radiação no mamógrafo é pouco maior que a metade do campo dos sistemas convencionais e para se conseguir isso utilizam-se colimadores de feixes e restritores,

isso ajuda a evitar radiação ionizante em outras partes do corpo da paciente e também colabora com a melhoria da imagem. Filtros impedem que os fótons do feixe que atrapalham a formação da imagem sejam espalhados, além de atingirem a paciente somando-se à dose de radiação recebida. Os filtros, geralmente, são de molibdênio, com cerca de 0,03mm de espessura, que impedem a passagem de radiação inferior a 20Kev.

No sistema mamográfico, o ponto focal é bem pequeno, pois deve ser possível a visibilização de pequenas estruturas (de até 0,3 mm de diâmetro).

O chassi mamográfico apresenta um écran intensificador que, ao contrário do convencional, se posiciona embaixo do filme, como pode ser visualizado na Figura 2.2. Os fótons atravessam o filme, chegando pela sua base, atingem o écran, transformam-se em luz visível e, refletidos de volta, impressionam o filme. Esse posicionamento evita o efeito *crossover* e ajuda na obtenção de uma melhor resolução da imagem. O efeito *crossover* é o fenômeno em que o filme é impressionado duas vezes pelo mesmo fóton, de modo que isso causa uma certa penumbra na imagem, deteriorando a resolução. Além de evitar o *crossover*, o posicionamento do écran no chassi mamográfico previne uma grande absorção de fótons antes que eles se encontrem com o filme, pois como os raios-X na mamografia são de baixa energia, um simples écran poderia absorver mais que 50% dos fótons que a ele chegam (Wolbarsst, 1993).

Figura 2.2- Sistema mamográfico

O contraste numa imagem mamográfica é justamente o grau de variação de densidade óptica (grau de escurecimento) entre diferentes áreas da imagem, ou entre uma anormalidade e os tecidos adjacentes. Ele pode ser determinado tanto pelo *contraste do objeto*, produzido pelo corpo, como também pelo *contraste do receptor*, relativo ao sistema de registro, que pode

amplificar ou reduzir o contraste do objeto. Com a radiografia padrão, por exemplo, o contraste do objeto pode referir-se ao contraste nos padrões de intensidade dos raios-X que emergem do corpo do paciente. O *contraste do receptor* do sistema écran-filme descreve sua resposta aos diferentes níveis de intensidades dos raios-X. Isto é determinado em parte pelo processo de fabricação e revelação do filme e também pelas propriedades de fluorescência do écran, sendo possível ser observado através de sua curva característica. (Wolbarst, 1993).

O contraste da imagem é importante, pois para que um detalhe, mesmo que grande o bastante, seja visto, é preciso que haja diferentes graus de densidade óptica entre os tecidos ao seu redor. Ele depende do material do ânodo do tubo de raios-X do mamógrafo ou do espectro de radiação produzido pelo tubo, da filtração do feixe primário, da quantidade de radiação espalhada produzida pela mama; da curva característica e condições do filme.

A exposição à radiação resulta no escurecimento do filme ou aumento da densidade óptica, e é proporcional ao produto da corrente do tubo de raios-X pelo tempo de exposição (Curryii, 1990). O contraste é definido pela diferença de densidade óptica entre duas regiões da imagem. Segundo Cameron, o contraste na mamografia é diretamente proporcional às diferenças entre os produtos dos coeficientes de atenuação dos tecidos e as densidades ópticas (Cameron, 1978).

Além do contraste, a resolução espacial da imagem é um fator de influência para a visualização de detalhes. A nitidez depende do tamanho do ponto focal, da distância foco-filme e filme-objeto e da resolução especial do sistema de registro da imagem.

A resolução espacial de uma imagem está relacionada com a sua capacidade de reproduzir detalhes do objeto radiografado. Em outras palavras, pode ser definido como o espaçamento mínimo entre pequenas estruturas (do corpo do paciente) que permite que elas sejam visualizadas e distinguidas na imagem.

A partir das imagens resultantes das duas exposições mostradas na Figura 2.1, o especialista analisa e interpreta estruturas da mama para fins como rastreamentos e detecção precoce do câncer de mama.

2.2 Interpretação mamográfica segundo especialistas

Geralmente, os especialistas são treinados para seguir uma sistemática de análise das imagens. Ao analisar uma imagem eles devem responder as seguintes questões: Achei? É real? Onde está? O que é? O que deve ser feito em relação ao achado?

As lesões procuradas pelos especialistas são várias e cada qual com significados histológicos distintos, embora numa imagem mamográfica, nem sempre distinguir cada tipo de lesão seja uma tarefa simples. A Figura 2.3 mostra um quadro de lesões que devem ser procuradas e suas características radiográficas.

Diagnóstico histológico	Achados mamográficos característicos
Lipoma	Lesão focal, redonda ou oval de densidade da gordura e borda delicada (cápsula)
Hamartoma	Lesão focal, bem delineada contendo áreas lipomatosas e densidade hídrica
Cisto oleoso	Radiolucência focal com borda bem delimitada (cápsula) com/sem calcificações em forma de casca de ovo
Linfonodo	Densidade focal, uniformemente delineada, com — dependendo da projeção — uma radiolucência central ou marginal
Galactocele	Lesão focal, bem delimitada, composta de componentes gordurosos e estruturas equivalentes à água (possivelmente em camadas), história de gestação ou lactação
Fibroadenoma calcificado	Densidade focal, uniformemente delineada com/sem halo, oval ou polilobulada, se contiver calcificações características
Cisto contendo excrescências cálcicas	Densidade redonda, focal, uniformemente delineada com calcificações que se projetam centralmente na incidência craniocaudal e se acumulam inferiormente na incidência médio-lateral
Verruga	Massa bem delimitada que exhibe uma estrutura em couve-flor (multilobulada) e frequentemente bordas precisas (tecido mole — ar)

Figura 2.3 - Quadro de tipos lesões com suas respectivas características morfológicas (Heywang-Köbrunner, 1999).

De acordo com a classificação BI-RADS (*Breast Imaging Reporting and Data System*) (ACR, 1998), protocolo que definiu desde 1992 modo padronizado para a apresentação do laudo mamográfico (descrito a seguir no item 2.3), as massas podem ser classificadas conforme seu formato, como mostrado a seguir.

- a) Circular: uma massa que tem forma esférica, circular ou globular.
- b) Oval: uma massa que é elíptica.
- c) Lobular: uma massa que tem contornos com ondulações.

- d) Irregular ou espiculada: a forma da lesão não pode ser caracterizada por qualquer das anteriores.

A Figura 2.4 mostra um diagrama das classes de formatos de massas.

Figura 2.4 Formatos de massas de acordo com o grau de suspeitabilidade (Kopans, 2000).

Uma densidade com contornos bem definidos tem grandes chances de ser uma lesão benigna.

Os achados localizados com contornos definidos, vistos em mamografia de rastreamento, usualmente representam cistos ou fibroadenomas, mas cerca de 2% a 7% de todos os carcinomas mamários podem se apresentar como um nódulo com margens parcialmente bem circunscritas e 2% com um contorno bem definido.

Estratégias Diagnósticas:

1 – Primeiro passo: a lesão se encontra dentro ou fora do tecido glandular mamário (parênquima)?

2 – A lesão intraparenquimatosa bem delimitada é posteriormente avaliada como uma radiodensidade.

3 - As lesões bem circunscritas ou razoavelmente bem circunscritas de densidades de tecidos moles necessitam ser adicionalmente investigadas.

Um resumo de como se caracterizam as lesões focais é mostrado na Figura 2.5(a) e (b).

(a)

(b)

Figura 2.5 – (a)Quadro de resumo das lesões focais e (b)Continuação do quadro de resumo (Heywang-Köbrunner, 1999).

O espectro das densidades com bordas pouco definidas inclui: lóbulos parenquimatosos, tumores benignos (cistos, fibroadenomas, tumor filóide, papiloma e adenose nodular), alterações pós-traumáticas ou pós-terapêuticas (hematoma, necrose gordurosa, cicatriz) e carcinomas.

Estratégia Diagnóstica:

1 - Inicialmente deve-se decidir se o achado representa uma real lesão ou é causado por superposição ou somação.

2 – Se a presença da massa for confirmada, devem-se obter as mamografias prévias para a comparação do tamanho e da radiodensidade.

- Lesões que forem claramente suspeitas sempre necessitam de investigação adicional (biópsia), mesmo que o seguimento não mostre alteração a curto prazo.
- Um achado estável por mais de três anos favorece um achado benigno.

Um resumo da caracterização de lesões mal definidas pode ser visto na Figura 2.6.

Figura 2.6 – Quadro de resumo da análise de lesão mal definida (Heywang-Köbrunne, 1999).

Além das lesões citadas acima, uma atenção especial deve ser dada às calcificações e microcalcificações.

As calcificações são depósitos de cálcio que podem ser visibilizadas em imagens mamográficas. Mais de 20% de calcificações de mama são produzidas por lesões de doenças malignas. Através de uma avaliação cuidadosa dos formatos, texturas e distribuições destas calcificações, torna-se mais fácil identificar a sua procedência (Näppi, 1999). Já as microcalcificações são consideradas estruturas menores ou iguais a 0,5 mm. Frequentemente são os primeiros indícios da existência de uma lesão na mama, antes que ela possa ser identificada com segurança. Os formatos, configurações e tamanhos podem ser sugestivos da característica - maligna ou benigna - daquela lesão.

De acordo com Lanyi *et. al.* (Lanyi, 1986), a patogênese das microcalcificações intramamárias não é uniforme. Elas podem estar associadas com inflamação, degeneração, e processo de metabolismo tóxico; ou elas podem resultar de lesões inclusive mecânicas. As calcificações de doenças tipicamente malignas geralmente são múltiplas e são uniformemente distribuídas no tecido; elas são granulares e têm uma densidade relativamente baixa; formam-se dentro de ductos dilatados que contêm resíduo celular.

Por outro lado, estudos mostram que depósitos de cálcio em carcinoma de mama não ocorrem apenas entre as células de câncer, mas também dentro de tumores, de células e em estromas vizinhos.

Algumas características radiográficas que levam à suspeita de malignidade (Lanyi, 1986):

- microcalcificações de 0,1 a 0,3 mm de diâmetro;
- pouca densidade;
- formatos irregulares e densidades não uniformes;
- agrupamentos ovais de 2 a 5 mm de diâmetro;
- pelo menos 5 microcalcificações por cm³.

Lanyi (Lanyi, 1986) considera que, quanto maior a microcalcificação, menor a suspeita de malignidade. Por outro lado *clusters* de microcalcificações são frequentemente o primeiro sinal do câncer de mama visível em mamogramas.

Segundo trabalhos publicados por Michèle Le Gal *et. al.* (Le Gal, 1984), as características radiográficas a serem analisadas para que se possam classificar as estruturas devem incluir tamanho de microcalcificação, número de microcalcificações por centímetro cúbico, (que no plano radiográfico é representado em duas dimensões), formas destas microcalcificações, densidade e distribuição.

As formas podem ser representadas pela tabela 2.1 abaixo, segundo seus estudos, que têm sido usados como critério na prática clínica.

Tabela - 2.1 Classificação de Michele Le Gal (1984)

Tipos	Microcalcificações- Morfologia	% de Malignidade
Tipo I	Anelares, redondas, discóides, com centros lucentes	Todas são benignas
Tipo II	Redondas, isodensas, uniformes	22% são malignas
Tipo III	Puntiformes, tipo "poeira", difícil identificação	40% são malignas
Tipo IV	Irregulares, poliédricas, tipo "grão de sal"	66% são malignas
Tipo V	Vermiculares, ramificadas, em forma de letras	Todas são malignas

Na análise das microcalcificações, há três fatores característicos fundamentais para se chegar a um diagnóstico mais preciso:

- formato do *cluster* das microcalcificações;
- formato das microcalcificações individuais;
- número de microcalcificações por *cluster* (informação relevante).

As Figuras 2.7(a) e 2.7(b) mostram um resumo de como estas partículas são analisadas.

(a)

(b)

Figura 2.7 – (a)Quadro de resumo da análise de microcalcificações e calcificações mamárias; e (b) Continuação do quadro de resumo (Heywang-Köbrunner, 1999).

2.2.1 A questão das mamas densas

Os diversos padrões do parênquima mamário têm sido debatidos no contexto da estimativa dos riscos do câncer de mama, através de investigações da etiologia desse câncer, de estratégias de prevenção, bem como de estratégias de aquisição de imagens (Gram, 1997). Um padrão de parênquima mamário denso, significa que ele é formado, predominantemente, por tecidos fibroglandulares, cuja absorção de raio-X é pequena e, mamograficamente, produz uma imagem de contraste ainda menor e de aparência bastante clara.

O tecido denso é encontrado em uma alta porcentagem em mulheres com menos de 40 anos de idade e com uma frequência decrescente em mulheres com mais de 50 anos de idade. O problema da mama mamograficamente densa é a inerente relação inversa entre a sensibilidade mamográfica de detectar carcinomas sem microcalcificações e o aumento da densidade mamária.

Os especialistas na prática clínica têm usado a seguinte estratégia para este tipo de padrão de parênquima mamário:

- Para este tipo de mama, a análise é um pouco mais complexa, há vários tipos de mama densa e para cada um tem-se uma estratégia diferenciada:
- em mulheres assintomáticas com menos de 40 anos, sem risco aumentado, o rastreamento com mamografia não é recomendado por várias razões, entre elas, a baixa sensibilidade do exame causada pelo parênquima mamário denso; e baixa incidência de câncer de mama nesta faixa etária.
- O rastreamento em mulheres com 40 a 50 anos recomenda mamografias a cada 1 ou 2 anos. Mulheres acima de 50 anos devem ser submetidas a mamografias anualmente.

O rastreamento em mamas densas tem valor, pois é possível mesmo nas mamas muito densas detectar microcalcificações e tumores densos. Quando o nódulo não é calcificado ou ainda a sua densidade relativa é menor que o tecido circundante, a detecção torna-se bem mais difícil, e nestes casos o exame de palpação pode ser um grande aliado.

Em pacientes assintomáticas, métodos complementares como ecografia e ressonância não são indicados.

- Se uma paciente assintomática se tornar sintomática por causa de um achado palpável ou mamográfico, é indicado procedimento complementar;
- Mama densa em mulheres assintomáticas de alto risco: se o risco de desenvolvimento de um carcinoma aumentar por causa da história clínica, então é apropriado que se faça uma investigação individualizada;
- Mama densa com achado: os achados palpáveis necessitam ser avaliados para verificação de sua importância clínica; somente 20% das massas palpáveis excisadas são malignas (Heywang-Köbrunner, 1999).

A Figura 2.8 ilustra um esquema de como é analisada a densidade na mama.

(a)

(b)

Figura 2.8 – (a)Quadro de resumo da análise de mama radiograficamente densa; e (b) Continuação do quadro de resumo (Heywang-Köbrunner, 1999).

A densidade da mama é um fator decisivo na qualidade da imagem radiológica e, consequentemente, na exatidão do diagnóstico.

A dificuldade em analisar mamogramas de mamas densas, segundo Jackson *et al.* (Jackson, 1993) se deve a vários fatores físicos: as lesões mamárias têm propriedade de atenuação de raios-X similares àquelas de tecidos glandulares e fibrosos, o que torna sua detecção mais difícil:

- a mama radiograficamente densa produz mais radiação espalhada, resultando em menor contraste na imagem;
- a mama radiograficamente densa tem mais tecidos não homogêneos e um grande intervalo de atenuações, tornando difícil expor, de forma ótima, todas as suas áreas.

O grau de densidade que uma mama apresenta também é um problema para ser determinado e é uma característica importante para os especialistas.

Uma das primeiras classificações de densidade de mama foi proposta do Wolfe *et al.* (Wolfe, 1975) e consiste em 4 categorias: *N1* – mama composta predominantemente por tecido gorduroso; *P1* – mama composta principalmente por tecido gorduroso, com ductos proeminentes representando até 25% do parênquima mamário; *P2* – dutos proeminentes ocupam mais que um quarto do volume da mama e *DY* – que tem alta densidade; (Gram, 1997).

Mais recentemente, a padronização BI-RADS tem classificado em: I. Gordurosa - lipo substituída; II. parcialmente lipo-substituída; III. densa heterogênea; e IV. densa.

De acordo com a literatura, cada um dos padrões de densidade tem um risco associado ao desenvolvimento do câncer de mama (Chan, 2002).

2.3 Padronização BI-RADS

Em 1992, um estudo colaborativo entre membros de vários comitês do *American College of Radiology* (ACR), com a colaboração do *National Cancer Institute*, *Centers for Disease Control and Prevention*, *Food and Drug Administration*, *American Medical Association*, *American College of Surgeons*, *College of Surgeons* e o *College of American Pathologists*, produziu a primeira edição do ***The ACR Breast Imaging Reporting and Data System (BI-RADS®)***. Trata-se do mais completo sistema já elaborado para garantir a padronização do laudo mamográfico. O objetivo da sistematização é padronizar a nomenclatura dos laudos, que devem possuir conclusão diagnóstica e propor conduta, ressaltando que a mamografia deve sempre ser precedida pelo exame físico e comparada com exames anteriores.

Em setembro de 1995, a segunda edição (ACR, 1995) do BI-RADS® foi lançada com algumas modificações, que a prática diária revelou necessárias. É essa edição que é brevemente aqui relatada, embora já exista uma outra de 2003 com pequenas alterações, que serão comentadas mais adiante.

Desde 1992, o Colégio Brasileiro de Radiologia (CBR) com o auxílio da Comissão Nacional de Energia Nuclear vem desenvolvendo o Programa de Controle e Manutenção da

Qualidade em Mamografia, cadastrando inúmeros serviços em todo o País (o Brasil foi o 2º País no mundo a ter um programa semelhante), mas muitas deficiências ainda eram observadas e a padronização do laudo mamográfico foi uma das necessidades. Assim, o CBR promoveu em São Paulo, em abril de 1998, uma reunião com a Sociedade Brasileira de Mastologia (SBM) e com a Federação Brasileira das Sociedades de Ginecologia e Obstetrícia (FEBRASGO), com o objetivo de padronizar os laudos mamográficos também nesse meio.

As diversas recomendações desta reunião de consenso (SBM & FSGO, 1998) são baseadas também no BI-RADS®: o laudo deve inicialmente descrever a densidade radiológica das mamas, lembrando-se que a maior densidade do parênquima mamário reduz a sensibilidade do método, porém não altera a categoria BI-RADS®. Após a descrição da densidade, descrevem-se nódulos porventura existentes (detalhando forma, margens e densidade), calcificações e outros achados. Finalmente, conclui-se relatando a categoria e propondo a conduta. Isso pode ser detalhado conforme os seguintes passos:

- laudo mamográfico deve ser estruturado em quatro fases:
 - I. Fase Clínica: anamnese e exame físico (não deverá constar no laudo);
 - II. Fase Descritiva: baseada no vocabulário do BI-RADS®;
 - III. Fase Diagnóstica: os laudos devem conter conclusão diagnóstica, sem correlação histológica:
 - Categoria 0 - mamografia sem achados, porém exige investigação;
 - Categoria 1 - mamografia normal ou sem achados visíveis;
 - Categoria 2 - achados mamográficos benignos;
 - Categoria 3 - achados mamográficos provavelmente benignos;
 - Categoria 4 - achados mamográficos suspeitos;
 - Categoria 5 - achados mamográficos altamente suspeitos.
 - IV. Fase de Recomendação de Conduta: os laudos devem incluir recomendação de uma ação conforme as categorias:
 - Categorias 1 e 2: seguimento anual ou bienal, ou então seguir com exame complementar;
 - Categoria 3: seguimento mamográfico em menor espaço de tempo (6 meses);
 - Categorias 4 e 5: indicação de investigação cito-histológica e especificar a técnica a ser adotada.

Para definir o que são estes achados e categorizá-los deve-se conhecer o vocabulário da imagem mamográfica.

2.3.1 Nódulo

É uma lesão visualizada em duas diferentes incidências. Se aparece apenas em uma única incidência, é chamada de DENSIDADE, até sua tridimensionalidade ser confirmada.

1. Forma: a) *redondo*: esférico, circular, globular; b) *oval*: elíptico; c) *lobular*: contorno com ondulações; d) *irregular*: forma não característica; e) *distorção arquitetural*: a arquitetura normal está distorcida, sem lesão definida visível. Inclui: espículas radiadas de um ponto de retração focal ou distorção da margem do parênquima. Pode também ser um ACHADO ASSOCIADO.

2. Casos Especiais: a) *densidade tubular ou ducto solitário dilatado*: caso não esteja associado com suspeita clínica ou achados mamográficos, é de pouca importância; b) *linfonodo intramamário*: é tipicamente uniforme ou possui um detalhe radiotranslúcido devido a presença de gordura no hilo. Geralmente mede 1 cm ou menos. Pode ser maior quando a substituição gordurosa for pronunciada. É um diagnóstico apenas para nódulos nos quadrantes superiores laterais - QSL; c) *tecido mamário assimétrico*: relativo à presença de grande proporção de tecido fibroglandular na área correspondente da mama contralateral. Não existe formação nodular, não há hiperdensidade central, distorção de arquitetura ou calcificações associadas. Geralmente representa uma variação NORMAL, mas pode ser significativo quando corresponde a uma assimetria palpável; d) *densidade assimétrica focal*: é uma densidade que não pode ser adequadamente descrita. É vista como uma densidade de formato semelhante em duas incidências e completamente sem margens para caracterizar aspectos benignos específicos, justificando avaliação adicional. Esta pode revelar um nódulo real ou uma distorção arquitetural importante.

3. Margens ou contornos: a) *definidas*: margens bem definidas; transição abrupta entre a lesão e o tecido circunjacente; b) *microlobuladas*: pequenas

ondulações; c) *margens encobertas*: escondida, por sobreposição ou tecido normal adjacente; d) *indefinidas*: a pouca definição das margens sugere infiltração pela lesão; e) *espiculadas*: linhas radiadas partindo das margens das lesões.

4. Densidade (em relação ao tecido mamário): a) *alta densidade*; b) *igual densidade*; c) *baixa densidade*; d) *conteúdo gorduroso* (radiotransparente, radiotranslúcido)

2.3.2 Calcificações

As benignas são grandes, grosseiras, redondas e facilmente visibilizadas. Calcificações associadas com malignidade freqüentemente são pequenas e necessitam do uso de lentes de aumento. Quando a etiologia específica não pode ser definida, sua descrição deve incluir a morfologia e distribuição. As calcificações benignas não precisam sempre ser relatadas, apenas quando possam ser interpretadas erroneamente.

Os tipos das calcificações levam à sua distribuição da seguinte forma:

1. Tipicamente benignas: a) *calcificações de pele*: tipicamente apresentam o centro radiotransparente. Quando apresentam formas atípicas deve-se realizar incidências tangenciais para confirmar sua origem na pele; b) *calcificações vasculares*: traves paralelas ou tubulares e lineares; c) *grosseiras*: (tipo “pipoca”) clássicas no fibroadenoma involuído; d) *bastonetes*: hastes contínuas, que podem ramificar. Medem mais de 1 mm de diâmetro, podem ter o centro translúcido. São as calcificações encontradas na “mastite plasmocitária”, na ectasia ductal e nas doenças com fluxo papilar; e) *redondas*: quando múltiplas, podem ser de variados tamanhos. São geralmente benignas e quando menores de 1 mm são formadas nos ácinos dos lóbulos. Se menores de 0,5 mm o termo PUNTIFORME pode ser utilizado; f) *centro translúcidas*: variam de 1 mm até acima de 1 cm. São redondas, ovais, com superfície lisa e centrotranslúcidas. Incluem: necrose gordurosa, ductos calcificados e, ocasionalmente, fibroadenomas; g) *casca-de-ovo*: formadas pelo depósito de cálcio na superfície de uma esfera. Geralmente são finas e encontradas na necrose gordurosa e nas

paredes dos cistos; h) *leite-de-cálcio*: formadas pelo sedimento de calcificações no interior de cistos. Na incidência crânio-caudal, aparecem como um floco, arredondada ou amorfa, e na médio-lateral, surgem com forma definida, côncava ou quarto crescente; i) *calcificações de sutura*: comum após RT. Podem ser lineares ou tubulares; j) *calcificações distróficas*: comuns em mamas irradiadas ou pós-trauma. São irregulares e geralmente maiores do que 0,5 mm. Frequentemente têm o centro radiotranslúcido; l) *puntiformes*: redondas ou ovais, menores que 0,5 mm e bem definidas.

2. Intermediárias: a) *amorfas*: geralmente arredondadas em flocos, suficientemente pequenas ou duvidosas para que sua morfologia específica seja determinada.

3. Provavelmente malignas: a) *Pleomórficas ou heterogêneas*: irregulares no tamanho e na forma, menores de 0,5mm; b) *ramificadas*: são finas, irregulares e lineares, porém descontínuas e menores de 0,5mm em extensão. Sua aparência sugere conteúdo do lúmen do ducto.

4. Distribuição: a) *agrupadas*: quando múltiplas calcificações ocupam um pequeno volume, inferior a 2 cm²; b) *lineares*: dispostas linearmente, podendo ter pontos ramificados; c) *segmentares*: sua distribuição sugere depósitos intraductais e sua ramificação aumenta a possibilidade de câncer multifocal segmentar. Embora possa ocorrer em alterações benignas, essa distribuição sugere malignidade, principalmente se a morfologia for intermediária ou suspeita; d) *regionais*: dispersas em grande área de tecido mamário, não necessariamente em toda a mama e não respeitando a distribuição ductal. Geralmente benignas; e) *difusas*: distribuídas aleatoriamente.

Obs.: Expressões como: agrupamentos (múltiplos grupos) ou agrupamentos similares (múltiplos grupos similares) podem ser utilizadas quando existirem mais de um grupo de calcificações semelhantes em morfologia e distribuição.

2.3.3 Achados Associados

Usados com nódulos ou calcificações. São descritos apenas como ACHADOS caso não exista outra anormalidade presente:

1. retração de pele
2. retração de mamilo (ou inversão)
3. espessamento de pele
4. espessamento trabecular (espessamento dos septos fibrosos)
5. lesão de pele (observada em duas projeções)
6. adenopatia axilar (linfonodos aumentados, não gordurosos podem ser comentados. O aspecto mamográfico não é confiável)
7. Distorção arquitetural

2.3.4 Localização da Lesão

As lesões significativas devem ser visibilizadas em duas projeções. Devem ser descritas usando a orientação clínica extrapolada para o filme: 1. Localização – quadrante superior ou inferior, lateral ou medial; região retro areolar; região axilar; 2. Lado – esquerdo ou direito.

Uma vez conhecidas as caracterizações dos achados mamográficos é preciso que o laudo seja definido de maneira concisa e objetiva de maneira que a imagem seja classificada em uma das categorias. É preciso que se saiba que em cada categoria está inferido um risco para câncer de mama:

a) avaliação incompleta

Categoria 0: necessária avaliação adicional de imagem. Inclui compressão localizada, ampliação, incidências adicionais, Ultra-sonografia. Essa categoria é freqüentemente utilizada em exames de rastreamento, porém raramente quando se considera o estudo imagenológico completo.

b) avaliação completa

Categoria 1: negativa (0%) – as mamas são simétricas, não há nódulos, distorções de arquitetura ou calcificações suspeitas.

Categoria 2: achados benignos (também um exame negativo) (0%) – são achados que merecem descrição, porém tipicamente benignos, como: fibroadenomas calcificados, calcificações secretórias (benignas), lesões com conteúdo de gordura (cisto oleoso, lipoma), lesões com densidades mistas (hematoma) e implantes.

Categoria 3: achados provavelmente benignos (<2%) – pode-se pedir um controle em intervalos inferiores a 1 ano para avaliar a estabilidade da lesão (são achados dos quais não se esperam alterações, já que considerados benignos; entretanto, é preferível avaliação precoce).

Categoria 4: achados suspeitos - considerar biópsia (2-90%) – lesões que não têm características de câncer de mama mas também não podem ser classificadas como provavelmente benignas. O médico e seu paciente podem tomar a decisão sobre a realização de procedimento.

Categoria 5: altamente sugestivos de malignidade (>90%) prosseguir com biópsia.

Em 2003 foi lançada a 4ª Edição do BI-RADS. As principais modificações da última edição foram a subdivisão da categoria 4 em A, B e C, de acordo com o grau de suspeita, e a inserção da categoria 6, que inclui os casos nos quais já foi feita biópsia com resultado maligno, mas a lesão não foi totalmente retirada ou foi tratada por quimioterapia, radioterapia ou hormoneoterapia. A divisão da categoria 4 em 4A (suspeita leve), 4B (suspeita moderada) e 4C (suspeita forte, porém não sendo a lesão típica de câncer) certamente provocará polêmica por envolver critérios subjetivos. Um exemplo da categoria 4A é um nódulo palpável parcialmente circunscrito que para ultra-sonografia é sólido, sugestivo de fibroadenoma, sendo admissível controle em seis meses após resultado de citologia ou biópsia negativa (Quadros, 2003). A Figura 2.9 mostra o quadro com a nova padronização e os seus significados.

Categorias BI-Rads® (quarta edição, 2003)			
Categoria	Interpretação	Risco de câncer	Conduta
0	Inconclusivo		Exame adicional (ultra-sonografia, magnificação ou compressão localizada)
1	Benigno	0,05%	Controle anual a partir dos 40 anos
2	Benigno	0,05%	Controle anual a partir dos 40 anos
3	Provavelmente benigno	Até 2%	Repetir em seis meses (eventualmente biópsia)
4 (A, B, C)	Suspeito	> 20%	Biópsia
5	Provavelmente maligno	> 75%	Biópsia
6	Lesão já biopsiada e diagnosticada como maligna, mas não retirada ou tratada	100%	

Fonte: Breast imaging reporting and data system: BI-Rads®. 4ª edition. Copyright 1992, 1993, 1995, 1998, 2003 American College of Radiology.

Figura 2.9 Classificação BI-RADS da Quarta edição com as respectivas taxas de risco de câncer (Quadros, 2003).

2.4 A Imagem Mamográfica Digital

A imagem mamográfica digital é constituída por uma unidade básica que é o pixel (*picture element*). Deve apresentar-se com alta resolução espacial (geralmente usa-se de 150 a 600 dpi, pontos por polegada, na digitalização, que indica o tamanho de cada pixel da imagem). Além disso uma resolução de contraste com uma quantidade suficiente de níveis de cinza, variando de 10 a 12 bits, deve ser utilizada a fim de representar as diferenças entre os tecidos e estruturas presentes na imagem (Doi, 1995).

A imagem digital é formada por sensores que captam a energia luminosa atravessando, por exemplo, um filme mamográfico revelado, e a convertem em sinais elétricos que, transformados em pulsos digitais, serão manipulados para a extração de informações. Os requisitos de aquisição da imagem digital para esquemas CAD envolvem vários parâmetros, desde resolução (tanto de contraste quanto espacial) até o próprio *hardware*, a técnica de

varredura do equipamento. Estes parâmetros devem ser cuidadosamente analisados e suas diferentes configurações podem afetar o resultado final do esquema.

Para a digitalização de uma imagem mamográfica, o filme pode ser varrido por um feixe de luz branca com o sinal captado por transmissão. Nos digitalizadores laser, o filme é varrido por um feixe de laser. Neste caso, a varredura é mais rápida e a transmissão dos sinais também, com menos ruído devido ao baixo espalhamento provocado pelo feixe de laser.

Basicamente, nesse processo, amostra-se imagem analógica contínua $f(x,y)$ convertendo-a em uma matriz discreta de M por N pontos, igualmente espaçados. Cada ponto corresponde a um elemento da imagem digital (pixel). Quanto maiores os valores de M e N utilizados na digitalização de uma imagem, maior será sua resolução espacial e consequentemente, melhor será sua qualidade. Além da forma $M \times N$ (linhas por colunas da imagem), a resolução espacial de uma imagem digital pode ser apresentada na forma de quantidade de pontos por polegada ou também pelo tamanho do pixel.

A quantização faz com que cada um destes pixels assuma um valor inteiro, na faixa entre 0 e $(2^n - 1)$ valores, sendo n o número de bits do conversor A/D. Quanto maior o valor de n , maior a quantidade de níveis de cinza presentes na imagem digitalizada e melhor será sua qualidade.

Uma vez representada a imagem matematicamente, é então possível a manipulação dos valores discretos da função que a representam, de acordo com as necessidades de cada aplicação. É possível o processamento desses valores, a extração de informações importantes de seu conteúdo e sua transformação (Ballard, 1982; Duda, 1973; Gonzales, 1992; Schalkoff, 1989).

No caso de imagens digitais, as resoluções espacial e de contraste são tratadas de forma um pouco diferente que no sistema mamográfico comentado na seção 2.1. A dimensão de cada ponto ou elemento constituinte da matriz-imagem, o pixel, depende da resolução com a qual a imagem foi adquirida. O pixel é, então, a menor unidade sobre a qual podemos realizar operações, e o seu tamanho define a resolução espacial da imagem digital. O contraste é definido pela diferença existente entre os níveis de cinza dos pixels, especialmente entre um determinado pixel e seus vizinhos mais próximos. Quanto maior for essa diferença de níveis de cinza, maior será a probabilidade de destaque deste pixel em uma inspeção visual. Chama-se de resolução de contraste o intervalo de níveis de cinza que um pixel pode apresentar.

2.5 Conclusão

Além do exame clínico, a mamografia é fundamental para análise da mama e detecção do câncer, pois é o exame mais eficaz para esta finalidade possibilitando a avaliação da arquitetura anatômica, disposição e alteração dos tecidos no parênquima mamário.

A interpretação dos achados mamográficos requer experiência dos especialistas, e depende bastante da qualidade da imagem. O baixo contraste da imagem é um fator altamente prejudicial e se deve principalmente aos tecidos da mama apresentar graus de absorção de raios-X muito próximos. O contraste piora drasticamente quando o parênquima mamário é predominantemente composto por tecidos fibroglandulares, pois isto pode esconder estruturas patológicas.

Para um diagnóstico preciso e objetivo, uma padronização é de grande importância, uma vez que o laudo incorreto de uma imagem ou mal interpretado pode levar a paciente a se submeter a exames invasivos (biópsias) desnecessariamente, o que é traumático e tem um custo mais alto, ou então deixar de tratar um tumor, acarretando num diagnóstico de câncer tardio, o que pode levar à morte.

3

EXTRAÇÃO E ANÁLISE DE ATRIBUTOS

Na área médica, o reconhecimento de padrões é utilizado para caracterização de tumores, quantificação de deformidades de estruturas anatômicas, visualização de contornos e análises morfométricas, para auxílio ao diagnóstico.

Raramente se usa apenas uma medida de cada padrão para que o investigador decida a qual classe pertence um caso observado. O mesmo ocorre quando se procura classificar automaticamente estruturas detectadas numa imagem médica digital. O mais comum é extrair várias medidas de cada padrão e representá-las através de um vetor, o qual servirá de entrada para um classificador automático. Os atributos, ou características, nesse caso, são extraídos geralmente através de técnicas de processamento de imagens.

Assim, as imagens são submetidas a etapas de pré-processamento e processamento de nível médio, para detecção efetiva das estruturas de interesse. A segmentação, que é uma das técnicas utilizadas em processamento de imagens em aplicações de esquemas CAD, é o processo onde as estruturas de interesse devem ser reforçadas e os ruídos e artefatos eliminados.

Os atributos de amplitude da imagem, ou características espectrais, são obtidos a partir do valor do pixel em análise. Essa medida pode ser em um ponto específico ou em relação a sua vizinhança média, com a vantagem de reduzir os ruídos (Pratt, 1991).

Os atributos de textura são também muito importantes, pois ajudam a definir regiões de interesse e selecionar um conjunto de propriedades para que se possam identificar duas ou mais regiões. Propriedades texturais são dadas não para um ponto isolado da imagem, mas para uma região (Haralick, 1979). Os atributos de textura são características do domínio espacial da imagem, já que também podem trazer informações como sua altura e largura (Schalkoff, 1989). Ainda podem-se extrair características do domínio temporal, que serão medidas a informar eventuais movimentos da imagem.

Para imagens mamográficas, no caso da extração de atributos de microcalcificações isoladas ou agrupadas e também de nódulos, é preciso obter informações que possibilitem a identificação da malignidade ou não (ou ainda quantificar a suspeita). Por isso, as informações do domínio espacial são relevantes, uma vez que as medidas de tamanho, forma, distribuição espacial, distribuição interpartículas, contraste, entre outras, são significativas para associação das estruturas a lesões malignas ou benignas.

O processo de seleção de atributos tem como função mapear as medidas extraídas dos padrões (imagens) em características significativas (Gonzales, 1992; Mascarenhas, 1989). Esse processo tem duas ênfases:

- transformar o espaço de atributos, o que reduz a dimensão dos vetores de características, em busca de torná-los o mais significativos possíveis na representação de uma classe;
- calcular a distância entre as classes e medidas de informações.

Para a transformação de atributos, técnicas como transformada de Karhunen Loève e a função discriminante de Fisher são bastante utilizadas (Fukunaga, 1990). No cálculo da distância entre as classes, técnicas como medida de Divergência e Distância Jeffries- Matusita (JM) são empregadas (Young, 1986). As técnicas estão descritas detalhadamente em (Mascarenhas, 1989).

A distância JM é usada para medir separabilidade envolvendo múltiplas classes (Duda, 1973; Mascarenhas, 1989). Já o conceito da “Divergência” é definido apenas para duas classes, fornecendo informação média para a discriminação em favor de uma classe contra a outra (Mascarenhas, 1989).

O Erro de Bayes é uma técnica pela qual se calcula a função densidade de probabilidade, que resultará numa medida que quantifica o significado do atributo e o reorganiza. Em seguida, deve ser utilizada uma técnica para se calcular as distâncias entre este subgrupos de atributos selecionados pelo Erro de Bayes (Fukunaga, 1990).

3.1 Atributos Extraídos da Imagem

O significado da palavra formato pode ser simplesmente forma, mas não pode ser facilmente formalizado matematicamente. Às vezes, não se sabe dizer qual o formato de um determinado objeto, mas pode-se identificá-lo quando se o vê (Dryden, 1998). Portanto, para se analisar o formato de um objeto é preciso saber quais são suas propriedades e levar em consideração como eles são identificados pela percepção humana, especialmente quando se trata de objetos de formatos não geométricos e não bem definidos, como ocorre na maioria dos objetos de estudo em imagens médicas.

Formato pode ser definido como qualquer conjunto de pontos conectados (Costa, 2001).

Os métodos existentes na literatura para descrever um formato são diversos. Uma sequência de coordenadas cartesianas é um exemplo de descrição de forma quando se está analisando formas biológicas e não há um número suficiente de marcos anatômicos identificáveis na imagem. Os formatos também podem ser descritos pelos momentos angulares, momentos invariantes de Hu, coeficientes gerados pelas transformadas *wavelets*, entre outros.

Um destaque especial deve ser dado à teoria dos momentos, uma vez que tal técnica foi utilizada para alguns dos experimentos que serão mostrados a seguir.

Uma vez conhecido o método mais eficaz para descrição do objeto em estudo, uma série de outros passos farão parte da análise do objeto. Na verdade, a análise geométrica é um passo intermediário de geração de dados, com propriedades diferentes das simples distâncias dos sistemas convencionais. Os resultados desta análise geométrica devem ser posteriormente analisados por métodos multivariados.

3.1.1 A teoria dos Momentos

O método dos Momentos proporciona uma técnica robusta para decomposição de uma forma arbitrária em um conjunto de características. O aspecto principal dessa abordagem é que ela se baseia em transformações lineares diretas, sem necessidade de se determinarem parâmetros específicos. Ela apresenta simplicidade matemática e é muito versátil, sendo de uso viável em vários tipos de imagens e aplicações.

Momentos são muito úteis para a análise de forma. Momentos de ordem zero e de primeira, segunda e terceira ordens podem fornecer informações úteis em várias aplicações de visão computacional, como localização dos objetos, reconhecimento de formato, orientação e tamanho, além das características invariantes dos momentos quanto à rotação, translação e escala. A forma é freqüentemente representada por uma imagem binária e seus momentos podem ser obtidos através de algoritmos rápidos (Andersen, 1985).

As técnicas baseadas em momento descrevem as propriedades globais de uma imagem, isto é, a maneira com que as características de pixels brancos e pretos representam o padrão.

Em geral, funções de diversas variáveis podem ser caracterizadas por uma seqüência de números, como, por exemplo, coeficientes de Taylor e coeficientes de Fourier. Momentos representam uma técnica descritiva e intuitiva baseada no estudo da Mecânica dos Corpos. Hu, em 1961 (Hu, 1961) desenvolveu a teoria dos Momentos Invariantes (MI) e sugeriu sua aplicação para o reconhecimento invariante de padrões visuais, baseado na teoria dos Invariantes Algébricos (Elliott, 1913). O termo invariante denota uma quantidade (ou um valor) que permanece inalterada sob certas condições. A teoria dos invariantes algébricos de transformação linear, em conexão com uma série de problemas em Geometria Analítica, foi formulada primeiramente no século XIX no trabalho de Carley & Sylvester (Gurevich, 1964). Baseado nesse trabalho, Hu apresentou fórmulas para funções invariantes contendo momentos de segunda e terceira ordens, como exemplos da teoria algébrica de Momentos Invariantes.

Os Momentos Invariantes representam um conjunto completo de descritores de imagem, possuindo, portanto, uma propriedade de fundamental importância para a teoria de Reconhecimento de Padrões (Duda, 1973; Hall, 1979; Ballard, 1982; Gonzáles, 1992; Pratt, 1991; Jain, 1995).

As imagens são representadas matematicamente por funções bidimensionais de intensidade de luz $f(x,y)$, onde x e y são as coordenadas espaciais cartesianas e o valor da função f em qualquer ponto (x,y) é proporcional ao brilho da imagem naquele ponto (luminância).

O teorema da *Unicidade* estabelece que existem momentos de todas as ordens se $f(x,y)$ for contínua e possuir valores não nulos em apenas uma região do plano cartesiano (x,y) (Hu, 1962). O conjunto de momentos $\{m_{p,q}\}$ é determinado de uma maneira *única* por $f(x,y)$ e $f(x,y)$ é determinada unicamente por $\{m_{p,q}\}$.

Para representar toda a informação em uma região de uma imagem é necessário um número infinito de valores de momentos. Na prática isso é impossível, assim um dos objetivos no projeto de um sistema de reconhecimento que utilize essa técnica é a definição da ordem n dos momentos. Essa ordem deve fornecer informação suficiente para representar a imagem de uma maneira única, respeitando-se os critérios de velocidade de reconhecimento e o poder discriminante dos padrões a serem classificados.

A definição de momentos² (ou algébricos) geométricos regulares tem a forma de uma projeção da função $f(x,y)$, que representa a imagem em uma função monomial $x^p y^q$. O momento $(p+q)$ é definido pela equação (3.1) a seguir:

$$m_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q \cdot f(x,y) dx dy \quad (3.1)$$

($p, q = 0, 1, 2, \dots$)

A definição acima tem a forma de uma projeção da função $f(x,y)$ dentro do monômio $x^p y^q$; entretanto, a base, embora completa, não é ortogonal. O cálculo de m_{pq} consiste em se multiplicar a função $f(x,y)$ por um monômio $x^p y^q$ e integrar o resultado. Os monômios até ordem 3 podem ser vistos na Figura 3.1 $x^0 y^0, x^0 y^1, x^0 y^2, x^1 y^0, x^1 y^1, x^1 y^2, x^2 y^0, x^2 y^1, x^2 y^2, x^3 y^0, x^3 y^1, x^3 y^2$. Os momentos de ordem $(p+q)$ também podem ser interpretados como a resposta de um sistema de formação de imagem com uma função de transferência $x^p y^q$ e entrada $f(x,y)$. Os momentos de baixa ordem possuem relações intuitivas com os objetos; por exemplo, m_{00} refere-se à massa, m_{10} e m_{01} referem-se ao centro de massa e m_{11}, m_{20} e m_{02} referem-se às direções principais.

² Um momento é uma soma ponderada de um padrão $f(x)$ sobre todo o campo de entrada, com os pesos iguais a algum polinômio em x (Wood, 1996)

Figura 3.1 - Monômios de até terceira ordem que multiplicam a função $f(x,y)$ no cálculo dos momentos (Hall, 1979).

O momento bidimensional para uma imagem $f(x,y)$ discreta de tamanho $(N \times M)$ é:

$$m_{p,q} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q \cdot f(x,y) \quad (p, q = 0, 1, 2, \dots) \quad (3.2)$$

Esses momentos são chamados de momentos convencionais ou *raw moments*.

A fim de se trabalhar com imagens de qualquer tamanho, pode-se normalizar as coordenadas x e y entre $-1 \leq x, y \leq 1$.

Para normalizar os momentos e torná-los invariantes em relação à translação basta calculá-los em relação ao centro de massa da figura, por isso esses momentos são chamados de *Momentos Centrais*. Para normalizá-los em relação à escala é necessário normalizá-los em relação à área do objeto (da imagem).

As técnicas de normalização de tamanho, orientação, posição, inclinação ou espessura das linhas visam facilitar o manuseio dos dados e a extração de características, melhorar a taxa de reconhecimento em relação a invariância do sistema em relação a estes atributos.

Outros parâmetros como o ângulo do eixo principal do objeto, determinado pelos momentos, podem ser utilizados para normalizar em escala, orientação e posição os padrões contidos na imagem, para efeitos de comparações com imagens e parâmetros dos objetos previamente armazenados em bancos de dados.

Esses momentos não são invariantes à quaisquer distorções. Pode-se normalizar a equação (3.2) para tornar os momentos invariantes à mudanças de translação e escala. Para normalizar os momentos em relação à translação basta calculá-los em relação ao centro de massa da figura, por isso esses momentos são chamados de *Momentos Centrais*. Para torná-los invariantes à escala é necessário normalizá-los em relação à área da figura, conforme mostra a equação (3.3).

$$\mu_{pq} = \frac{1}{m_{0,0}^{1+(p+q)/2}} \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q \cdot f(x, y) \quad (p, q = 0, 1, 2, \dots) \quad (3.3)$$

onde $(p+q)$ é a ordem do momento e $m_{0,0}$ (momento de ordem 0) representa a área da região da imagem, como pode ser visto equação (3.4).

$$m_{0,0} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} f(x, y) \quad (3.4)$$

As equações (3.5) e (3.6) representam as coordenadas de centro de gravidade do objeto na imagem, onde $m_{1,0}$ e $m_{0,1}$ representam os momentos de primeira ordem:

$$\bar{x} = \frac{m_{1,0}}{m_{0,0}} \quad (3.5)$$

$$\bar{y} = \frac{m_{0,1}}{m_{0,0}} \quad (3.6)$$

O conjunto completo dos momentos de ordem n são todos os momentos $m_{p,q}$, com $(p+q) \leq n$ e $\frac{1}{2}(n+1)(n+2)$ elementos.

Os momentos Centrais são definidos em termos dos momentos ordinários, conforme pode ser visto nas equações (3.7):

$$\mu_{0,0} = m_{0,0} = \mu \quad (3.7.a)$$

$$\mu_{1,0} = \mu_{0,1} = 0 \quad (3.7.b)$$

$$\mu_{2,0} = m_{2,0} - \mu \bar{x}^2 \quad (3.7.c)$$

$$\mu_{1,1} = m_{1,1} - \mu \bar{x} \bar{y} \quad (3.7.d)$$

$$\mu_{0,2} = m_{0,2} - \mu \bar{y}^2 \quad (3.7.e)$$

$$\mu_{3,0} = m_{3,0} - 3m_{2,0}\bar{x} + 2\mu \bar{x}^3 \quad (3.7.f)$$

$$\mu_{2,1} = m_{2,1} - m_{2,0}\bar{y} - 2m_{1,1}\bar{x} + 2\mu \bar{x}^2 \bar{y} \quad (3.7.g)$$

$$\mu_{1,2} = m_{1,2} - m_{0,2}\bar{x} - 2m_{1,1}\bar{y} + 2\mu \bar{x} \bar{y}^2 \quad (3.7.h)$$

$$\mu_{0,3} = m_{0,3} - 3m_{0,2}\bar{y} + 2\mu \bar{y}^3 \quad (3.7.i)$$

ou genericamente como dado pela expressão (3.8):

$$\mu_{pq} = \frac{m_{pq}}{m_{00}^{\frac{(p+q)}{2}+1}} \quad (3.8)$$

Vários atributos podem ser determinados a partir destes momentos centrais. As informações sobre o centro de massa, área e orientação do eixo principal mais próximo ao eixo x (θ) podem servir tanto para um manipulador como um braço mecânico que irá manipular os objetos, ou para outro estágio mais elevado, onde as relações entre os diferentes objetos contidos na imagem proporcionam a interpretação da cena.

Um conjunto finito de momentos pode ser armazenada como uma matriz, como por exemplo mostra a matriz M abaixo, composta pelos momento não-centrais com $p, q \leq 3$.

$$M = \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \end{bmatrix} \quad (3.9)$$

Se uma ordem $n=p+1$ for selecionada e todos os momentos com n menor ou igual a um valor fixo forem selecionados, então é produzida uma matriz triangular superior. Matrizes similares podem ser produzidas para os momentos centrais e momentos normalizados.

Uma questão importante na implementação computacional dos momentos invariantes é a faixa dos momentos uma vez que isso determina os requisitos de armazenamento. É suficiente para implementação computacional usar limites superiores para o valor máximo. Para uma região retangular as bordas da região pode ser facilmente calculadas através da seguinte relação:

$$m_{pq} = \int_a^b \int_c^d x^p y^q \cdot f(x, y) dx dy \leq f \max \int_a^b \int_c^d x^p y^q dx dy \quad (3.10)$$

o que fornece

$$m_{pq} \leq f \max \left[\frac{b^{p+1} - a^{p+1}}{p+1} \right] \left[\frac{d^{q+1} - c^{q+1}}{q+1} \right] \quad (3.11)$$

Momentos centrais apresentam limites superiores similares. Limites inferiores podem ser calculados de maneira análoga. Os limites computados para os momentos podem ser

usados para escalonar as operações do computador, selecionar comprimento dos registros e normalizar valores das características computadas (Sadjadi, 1978).

Muitas aplicações necessitam que os descritores de forma sejam invariantes mas em outras aplicações a posição rotacional de algum objeto pode ser uma informação importante. O problema de se reconhecer objetos invariantes à rotação consiste no fato de que algumas formas apresentam simetrias similares em diversas posições rotacionais. Hu (1962) propôs dois métodos diferentes para se calcular momentos invariantes à rotação: Invariantes de Momento e Eixo Principal. Os momentos de segunda ordem definem eixos principais da mesma maneira que os momentos de inércia na Mecânica.

3.1.1.1 Os Momentos Invariantes de Hu

O conjunto dos momentos invariantes absolutos consiste em combinações não lineares aplicadas aos momentos, gerando funções invariantes à transformação de rotação. Através dos momentos de segunda e terceira ordem, Hu definiu sete funções denominadas Invariantes de Momento de Hu (MI) (Hu, 1962), que são invariantes à translação, rotação e escala. Essas funções são mostradas pela equação 3.12 a seguir.

$$\phi_1 = \mu_{20} + \mu_{02} \quad (3.12-a)$$

$$\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (3.12-b)$$

$$\phi_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad (3.12-c)$$

$$\phi_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} - \mu_{03})^2 \quad (3.12-d)$$

$$\begin{aligned} \phi_5 = & (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}) \cdot [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + \\ & + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03}) \cdot [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \quad (3.12-e)$$

$$\phi_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{03} + \mu_{21}) \quad (3.12-f)$$

$$\begin{aligned} \phi_7 = & (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - \\ & (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \quad (3.12-g)$$

Os seis primeiros momentos ϕ_1 a ϕ_6 são invariantes à rotação e à reflexão; entretanto, ϕ_7 é sensível à reflexão. Seu valor muda de sinal para a imagem refletida de um objeto, mas sua magnitude permanece inalterada. Portanto, caso se deseje invariância à reflexão, apenas os seis primeiros invariantes de Hu devem ser usados; caso contrário, seu sinal deve ser levado em consideração. O sétimo invariante é útil para distinguir imagens refletidas.

As variáveis de Hu tendem a ser muito pequenas. Masters (Masters, 1994) sugere um conjunto de transformações que atenua esses problemas e pode ser visto na equação 3.13 mostrada a seguir:

$$\begin{aligned}
 nh1 &= 10 \log(10\phi_1 - 1.2) \\
 nh2 &= 3 \log\left(\frac{\phi_2 + 0.0001}{0.0101}\right) \\
 nh3 &= 8 \log\left(\frac{\phi_3 + 0.0005}{0.0025}\right) \\
 nh4 &= 2 \log\left(\frac{\phi_4 + 0.0000001}{0.0000101}\right) \\
 nh5 &= 1000\sqrt[3]{\phi_5} \\
 nh6 &= 200\sqrt[3]{\phi_6} \\
 nh7 &= 500\sqrt[3]{\phi_7}
 \end{aligned} \tag{3.13}$$

Enquanto os MI possuem algumas propriedades que boas características de formato devem ter, nem sempre todos são utilizados na solução de um determinado problema em particular. Em alguns casos, poucos MI de baixa ordem podem refletir em características discriminantes de formatos suficientes (Castleman, 1979).

O método dos eixos principais descrito por Hu fornece as relações rotacionais entre os momentos de mesma ordem. Usando essas expressões, o conjunto de momentos de um objeto pode ser rotacionado sobre seu centro de gravidade. A vantagem dos momentos sobre outras técnicas é que esses descritores de formato têm uma implementação direta e também carregam uma interpretação “física” da forma.

Uma outra metodologia para normalização em relação à rotação consiste em rotacionar o objeto para uma posição padrão, método que foi definido por Hu (Hu, 1962) como Momentos Invariantes (MI). Nesse método, os momentos são calculados em referência

a um par de eixos principais determinados unicamente. A maneira mais fácil de se definir uma posição padrão para o objeto é achar seu eixo principal, ou seja, a direção onde ocorre o maior espalhamento, e rotacioná-lo até que fique horizontal. Existem duas orientações que satisfazem esse requisito, uma deslocada 180° da outra. Essa ambigüidade pode ser resolvida fazendo com que a parte mais pesada da Figura fique no lado esquerdo. Na linguagem estatística dizemos que o objeto está inclinado à direita. Como exemplo, a Figura 3.2 abaixo, extraída de (Masters, 1994) mostra a posição padrão.

Figura 3.2 – Posição padrão para Invariância à Rotação (Masters, 1994).

Os momentos invariantes discriminam bem classes nas quais a forma ou o contorno do padrão de entrada contém a maior parte da informação necessária para separar as classes. Às vezes é necessária uma característica diferente para lidar com classes que não são distintas apenas pela forma. Momentos de ordens superiores são uma opção, mas já foi visto que quanto maior a ordem maior a degradação, devido ao ruído.

Além de fornecer uma representação concisa das propriedades geométricas fundamentais da imagem, transformações geométricas básicas também podem ser realizadas sobre as representações dos momentos de uma imagem. Essas transformações são mais facilmente determinadas no domínio dos momentos do que no domínio dos pixels da imagem original. As transformações de escala, translação, rotação, reflexão, contraste e convolução resultam numa nova imagem e os momentos transformados podem ser calculados em função dos momentos da imagem original. A convolução de uma imagem $f(x,y)$ com um conteúdo ("kernel") discreto $w(i,j)$ de tamanho $N \times M$ pode ser considerada a soma de uma série de translações e escalas (Prokop, 1992).

Em particular, os Momentos Invariantes geométricos são altamente dependentes da simetria das formas. Os momentos envolvendo as potências ímpares de p (q) são zero para

formas simétricas sobre o eixo vertical (horizontal). Qualquer momento contendo uma potência ímpar de p e q é zero para formas que são simétricas em ambas direções, vertical e horizontal. Para tais formas, alguns momentos são iguais (zero) e não fornecem informações para separação das classes. Em relação ao método dos Eixos Principais, a grande dificuldade é a determinação do ângulo. Quando o objeto é assimétrico, é fácil determinar a direção do único eixo principal. Entretanto, às vezes os eixos principais não podem ser unicamente definidos, particularmente quando a imagem é circular ou apresenta simetria rotacional (n -fold)³. Por outro lado, se a imagem é simétrica, pode-se reduzir a quantidade de espaço de armazenamento para representação da forma, comparando apenas a porção repetitiva do objeto, ao invés de toda forma. Se certas porções de simetria forem perdidas, a mesma simetria pode fornecer pistas para recuperação da forma; além do mais, a ordem de simetria n representa uma boa característica de forma, uma vez invariante a escala, a translação e a rotação.

3.1.1.2 Interpretação geométrica dos Momentos

O uso de Momentos como descritores globais de características é conveniente do ponto de vista de processamento de imagem, uma vez que só requer poder computacional, o que é abundante atualmente. Entretanto, o grande problema refere-se à seleção do conjunto de características e também em como interpretar a informação fornecida pelos momentos.

Vários atributos de formato podem ser extraídos dos Momentos Cartesianos. O momento de ordem zero (m_{00}) fornece uma medida de área do objeto. Com os momentos de primeira ordem m_{10} e m_{01} , denominados “momentos de alavanca”, calcula-se o centro de massa do objeto, que é utilizado como uma localização padrão do objeto dentro do campo visual. As coordenadas do centro de massa são a intersecção das linhas $x = \bar{x}$ e $y = \bar{y}$. Essas linhas são

³ Uma forma é denominada simetria rotacional n -arestas (n -fold) se tornar igual a si mesma, após ser rotacionada sobre seu centro de gravidade através de um múltiplo qualquer de $2\pi/n$. Similarmente, os eixos principais são indefinidos para muitas formas, tais como as formas rotacionalmente simétricas de n -arestas com $n \geq 3$ ou figuras simétricas à reflexão.

paralelas aos eixos x e y e representam as linhas onde toda massa pode estar concentrada, sem alterar os primeiros momentos sobre os eixos x e y . O centro de massa define uma localização única do objeto e pode ser usado como ponto de referência para descrever sua posição dentro do campo de visão. Se a posição de um objeto é coincidente com a origem do campo visual, ou seja, $\bar{x} = 0$ e $\bar{y} = 0$, então os momentos computados para aquele objeto são denominados *Momentos Centrais* e são designados por λ_{pq} (note que $\lambda_{10} = \lambda_{01} = 0$). Os momentos m_{10} e m_{01} podem ser usados também para calcular o ângulo T do eixo de mínima inércia ou eixo principal da imagem, como mostra a equação 3.14 (Strachan, 1990):

$$T = \frac{1}{2} \arctan \frac{2(m_{00}m_{11} - m_{10}m_{01})}{(m_{00}m_{20} - m_{10}^2) - (m_{00}m_{02} - m_{01}^2)} \quad (3.14)$$

Os momentos de segunda ordem m_{02} , m_{11} e m_{20} são conhecidos como *Momentos de Inércia* e determinam várias características interessantes, tais como os eixos principais sobre os quais os segundos momentos são mínimo e máximo (eixo maior e eixo menor respectivamente), a elongação e orientação do eixo principal do objeto. Os momentos m_{20} e m_{02} são denominados “momentos de inércia” e o momento m_{11} é denominado produto de inércia. Os momentos de segunda e terceira ordem conduzem a uma série de interpretações intuitivas (Giuliano, 1961). As interpretações dos vários momentos de baixa ordem mostrados na tabela a seguir possuem uma analogia com a mecânica dos corpos e foram descritas em (Giuliano, 1961).

Tabela 3.1 – Interpretação geométrica dos Momentos Centrais de segunda e Terceira Ordem

<i>Momentos Centrais</i>	<i>Interpretação Geométrica</i>
μ_{20}	Centralidade horizontal
μ_{02}	Centralidade vertical
μ_{11}	Diagonalidade
μ_{12}	Divergência horizontal
μ_{21}	Divergência vertical
μ_{30}	Desequilíbrio horizontal
μ_{03}	Desequilíbrio vertical

A interpretação é feita com o ponto de vista centralizado sobre o centro de gravidade de cada objeto, com os eixos das coordenadas colocados sobre ele.

A *centralidade horizontal* μ_{20} é determinada pela variância da massa na direção x. Numa imagem do objeto, ela mede a variância da forma do objeto na direção x.

A *centralidade vertical* μ_{02} é determinada pela variância da massa na direção y. Numa imagem do objeto ela mede a variância da forma do objeto na direção y.

A *diagonalidade* μ_{11} é a distribuição da massa sobre os quatro quadrantes. Ela mede a proporcionalidade da distribuição da forma do objeto entre os 1º e 3º e os 2º e 4º quadrantes, vista do centro da massa.

Se $\mu_{11} < 0$ há mais massa no 2º e 4º quadrantes.

Se $\mu_{11} > 0$ há mais massa no 1º e 3º quadrantes.

Se $\mu_{11} = 0$ há uma distribuição de massa igual nos 1º e 3º em relação ao 2º e 4º quadrantes.

A *divergência horizontal* μ_{12} mede a dimensão da figura à esquerda comparada com sua direita, ou seja, a simetria do objeto na direção x.

Se $\mu_{12} < 0$ há mais massa no 2º e 3º quadrantes.

Se $\mu_{12} > 0$ há mais massa no 1º e 4º quadrantes.

Se $\mu_{12} = 0$ há uma distribuição de massa igual no 2º e 3º em relação ao 1º e 4º quadrantes.

A *divergência vertical* μ_{21} mede a dimensão da parte inferior da figura comparada com sua parte superior, ou seja, a simetria do objeto na direção y.

Se $\mu_{21} < 0$ há mais massa no 3º e 4º quadrantes.

Se $\mu_{21} > 0$ há mais massa no 1º e 2º quadrantes.

Se $\mu_{21} = 0$ há uma distribuição de massa igual no 3º e 4º em relação ao 1º e 2º quadrantes.

O *desequilíbrio horizontal* μ_{30} determina a localização do centro de gravidade CG com relação à metade da largura ($X_{1/2}$) do objeto.

Se $\mu_{30} < 0$ CG à direita de $X_{1/2}$

Se $\mu_{30} > 0$ CG à esquerda de $X_{1/2}$

Se $\mu_{30} = 0$ $\bar{x} = X_{1/2}$.

O *desequilíbrio vertical* μ_{03} determina a localização do centro de gravidade com relação a metade da altura ($Y_{h/2}$) do objeto.

Se $\mu_{03} < 0$ CG acima de $Y_{h/2}$.

Se $\mu_{03} > 0$ CG abaixo de $Y_{h/2}$

Se $\mu_{03} = 0$ $\bar{y} = Y_{h/2}$

As variâncias μ_{20} e μ_{02} são medidas de dispersão estatística que caracterizam a variabilidade de uma função densidade e medem as variações de amplitude da forma nas direções x e y.

Outra propriedade que pode ser determinada pelos momentos de segunda ordem é o Raio de Giração (RG) da imagem. O Raio de Giração⁴ sobre um eixo é a distância do eixo à linha onde toda massa está concentrada, sem alterar o segundo momento sobre o eixo. Os Raios de Giração sobre os eixos x e y são respectivamente dados pelas equações (3.15) e (3.16) abaixo:

$$RG_x = \sigma_x = \sqrt{\frac{\mu_{20}}{m_{00}}} \quad (3.15)$$

$$RG_y = \sigma_y = \sqrt{\frac{\mu_{02}}{m_{00}}} \quad (3.16)$$

Os Raios de Giração RG_x e RG_y medem, respectivamente, o desvio padrão σ_x e σ_y do objeto nas direções x e y.

O Raio de Giração sobre a origem é o raio de um círculo centrado na origem onde toda massa está centrada, sem alterar o segundo momento sobre a origem e é dado pela equação 3.17 a seguir:

$$RG_{xy} = \sqrt{\frac{\mu_{20} + \mu_{02}}{\mu_{00}}} \quad (3.17)$$

O Raio de Giração em relação à origem é inerentemente invariante à orientação da imagem e consequentemente tem sido utilizado como característica invariante à rotação para representação de objetos.

⁴ O raio de Giração pode ser interpretado também como o raio mínimo que inclui a maioria, senão todos os pontos em seu interior.

3.1.2 Outros atributos de forma

O perímetro e a área são características extraídas de uma cena ou objeto que muito podem contribuir para separação das classes. Medidas derivadas desses dois atributos também são relevantes.

Vários esquemas CAD usam essas características, pois em achados mamográficos, por exemplo, o formato, tamanho e distribuição dos objetos, como visto no capítulo 2, são importantíssimos na análise da imagem e laudo médico.

Medidas como irregularidade e compacidade são freqüentemente encontradas nos esquemas CAD vistos na literatura.

A área é dada pelo número total de pixels do objeto; essa medida pode ser calculada com algoritmos simples na maioria dos casos (Gonzáles, 1992; Mascarenhas, 1989; Masters, 1994; Schalkoff, 1989).

O perímetro é dado pelo número de pixels do contorno do objeto. Existem várias maneiras de se calcular essas medidas (Gonzáles, 1992; Mascarenhas, 1989).

A Compacidade é uma medida derivada do perímetro P e área A de um objeto. Em (Martì, 1998), é mostrada a extração de compacidade de imagens de microcalcificações, usando a equação 3.18.

$$C = P / 4\pi A \quad (3.18)$$

A irregularidade é outro atributo extraído da área e do perímetro. Tao, em 1998, usou a irregularidade como uma das medidas extraídas de achados mamográficos, usando a equação 3.19, pois em seu trabalho foi considerada irregularidade de fronteira, como sugestivo de malignidade (Tao, 1998).

$$I = A / P^2 \quad (3.19)$$

Muitas outras medidas de formato podem ser usadas nos esquemas CAD, para extrair informações de microcalcificações, regiões de interesse e nódulos. Essas medidas, quando combinadas, representam bem as classes em que se dividem os achados mamográficos.

3.1.3 Atributos de intensidade e descrição de textura

Os atributos mais básicos em medidas de amplitude da imagem são medidos em termos de luminância, valor espectral e outras unidades. Há muitos graus de liberdade na extração de atributos de amplitude. Alguns desses valores podem ser usados diretamente. As medidas de amplitude devem ser feitas em pontos específicos da imagem (Pratt, 2001).

O desvio padrão da imagem na literatura é também chamado de dispersão da imagem. A média e o desvio padrão podem ser calculados de forma indireta, através do histograma dos pixels da imagem dentro de uma vizinhança.

O histograma de uma imagem é um conjunto de números indicando a quantidade de pixels naquela imagem, que apresentam um determinado nível de cinza. Os valores são representados por um gráfico em barras que fornece para cada nível de cinza o número de pixels correspondentes. Através da visualização do histograma é possível se obter uma indicação de sua qualidade em relação ao nível de contraste e brilho médio (Ogê, 1999).

Através do histograma de uma imagem é possível se extrair vários atributos. Um histograma bimodal sugere que uma imagem contém um objeto com amplitude maior que seu fundo e pode sugerir o grau de contraste. Um histograma onde os pixels estão concentrados numa faixa pequena de valores indica uma imagem de baixo contraste (Pratt, 2001).

Alguns atributos de amplitude que podem ser extraídos de imagens médicas são:

- Média;
- Desvio padrão;
- Skewness;
- Kurtosis;
- Energia;
- Entropia;

Os atributos extraídos do histograma de segunda ordem são baseados em definições de distribuição de probabilidades de pares de pixels. Considerando 2 pixels $F(j,k)$ e $F(m,n)$ que estão localizados nas coordenadas (j,k) e (m,n) , respectivamente; eles são então separados por r unidades radiais num ângulo α em relação ao eixo horizontal (Pratt, 2001).

Algumas medidas extraídas através do histograma de segunda ordem:

- Autocorrelação;
- Covariância;
- Inércia;
- Valor absoluto;
- Diferença inversa;
- Energia;
- Entropia;

A textura é definida como uma combinação entre magnitude e frequência da variação de tons em uma imagem, sendo essa última produzida pelo efeito conjunto de todas pequenas feições que compõem uma área particular na mesma. Embora tonalidade e textura possam ser consideradas propriedades independentes, uma está intrinsecamente ligada à outra. Essa relação pode ser entendida da seguinte forma: quando uma área pequena de uma imagem tem pouca variação dos níveis de cinza, a propriedade dominante é a textura (Haralick, 1973; Haralick, 1979). A textura constitui uma característica importante em análises de diversos tipos de imagens.

Embora os métodos de análise de textura sejam muito utilizados, não há uma apresentação formal, nem uma definição precisa do termo. As técnicas de discriminação de textura encontradas na literatura são baseadas na intuição e percepção de cada autor (Haralick, 1979).

A textura pode ser representada por métodos estatísticos utilizando matrizes de co-ocorrência de níveis de cinza (SGLD – *Spatial Gray-Level Dependence*).

Os atributos de Haralick, propostos em 1973 como descritores de textura, são baseados em SGLD e formados por um conjunto de 14 medidas: variância, correlação, contraste, entropia, soma da entropia, diferença da entropia, momento da diferença inversa, média da soma, variância da soma, média da diferença, variância da diferença, energia, medida de informação de correlação 1 e medida de informação de correlação 2.

Conforme equacionamento de 3.20 a 3.33 dos descritores e seus significados apresentados a seguir, deve-se considerar n quantidade de níveis de cinza da imagem e $p(i,j)$ a matriz de co-ocorrência dos níveis de cinza.

1. *Uniformidade* ou *Energia*: indica uniformidade. Texturas homogêneas terão valor de energia mais alto, pois as texturas suaves possuem densidades mais concentradas. Texturas rugosas tem densidades com alta variância.

$$\text{UNIFORMIDADE} = \sum_i \sum_j P(i, j)_{\Delta x, \Delta y}^2 \quad (3.20)$$

2. *Contraste*: estimativa da variação de intensidade entre pares de pontos da imagem. Para texturas de baixo contraste, essa medida terá um baixo valor e, para texturas de alto contraste, o valor será alto.

$$\text{CONTRASTE} = \sum_i \sum_j (i - j)^2 P(i, j)_{\Delta x, \Delta y} \quad (3.21)$$

3. *Correlação*: indica a estrutura da textura.

$$\text{CORRELAÇÃO} = \frac{\sum_i \sum_j ij P(i, j)_{\Delta x, \Delta y} - \mu_i \mu_c}{\sigma_i \sigma_c} \quad (3.22)$$

4. *Variância*: indica a variação no fundo da imagem.

$$\text{VARIÂNCIA} = \sum_i \sum_j \frac{P(i, j)_{\Delta x, \Delta y}}{1 + (i - j)^2} \quad (3.23)$$

5. *Momento da Diferença Inversa*: medida de homogeneidade local. Enfatiza texturas sutis e pequenas mudanças.

$$\text{MOMENTO DIF. INVERSA} = \sum_k k P^D(k)_{\Delta x, \Delta y} \quad (3.24)$$

6. *Média da Soma*: média dos tons do fundo da imagem;

$$\text{MÉDIA DA SOMA} = \sum_k k P^S(k)_{\Delta x, \Delta y} \quad (3.25)$$

7. *Variância da Soma*: variação dos tons de fundo da imagem;

$$\text{VARIÂNCIA DA SOMA} = \sum_k (k - MS_{\Delta x, \Delta y})^2 P^S(k)_{\Delta x, \Delta y} \quad (3.26)$$

8. *Entropia da Soma*: indica quantidade de informação entre pares de pixels;

$$\text{ENTROPIA DA SOMA} = - \sum_k P^S(k)_{\Delta x, \Delta y} \log P^S(k)_{\Delta x, \Delta y} \quad (3.27)$$

9. *Entropia*: indica quantidade de informação resultante das interações entre os pixels da imagem;

$$\text{ENTROPIA} = - \sum_i \sum_j P(i, j)_{\Delta x, \Delta y} \log P(i, j)_{\Delta x, \Delta y} \quad (3.28)$$

10. *Variância da diferença*: variancia calculada com a média centrada em zero;

$$\text{VARIÂNCIA DA DIFERENÇA} = \sum_k (k - MD_{\Delta x, \Delta y})^2 P^D(k)_{\Delta x, \Delta y} \quad (3.29)$$

11. Entropia da diferença:

$$\text{ENTROPIA DA DIFERENÇA} = - \sum_k P^D(k)_{\Delta x, \Delta y} \log P^D(k)_{\Delta x, \Delta y} \quad (3.30)$$

12. *Medida de Informação de correlação 1*:

$$\text{MEDIDA INF. DE CORRELAÇÃO 1} = \sum_k P^D(k)_{\Delta x, \Delta y}^2 \quad (3.31)$$

13. *Medida de Informação de correlação 2*:

$$\text{MEDIDA INF. DE CORRELAÇÃO 2} = \sum_k P^S(k)_{\Delta x, \Delta y}^2 \quad (3.32)$$

14. *Máximo coeficiente de correlação*:

$$\text{MÁXIMO COEF. CORRELAÇÃO} = \sum_i \sum_j \frac{P(i, j)_{\Delta x, \Delta y}^2}{P(i, *)_{\Delta x, \Delta y} P(*, j)_{\Delta x, \Delta y}} \quad (3.33)$$

3.1.3.1 Atributos utilizados para descrição de imagens mamográficas

A análise de textura de imagens mamográficas propicia vários tipos de análises de lesões, o que auxilia na prática médica. Como já descrito anteriormente, o formato e o contorno são fatores cruciais na análise e caracterização de lesões mamárias (Kopans, 2000), para elaboração do diagnóstico do câncer de mama. Por isso, a preocupação em se tratar desse

problema em primeira instância. Porém, além dos descritores geométricos, uma análise de textura pode ser de fundamental importância no momento do diagnóstico de casos mais complexos. Essa preocupação em quantizar a densidade das lesões, ou mesmo da mama, deve-se também à análise prevista na padronização BI-RADS.

Alguns trabalhos para quantificar a variação de densidade ótica em mamogramas têm sido desenvolvido por Zheng, et al. (Zheng, 1997), eles elaboraram uma metodologia para quantificar o grau de dificuldade de identificação de lesões através de medidas que nomearam “índice de dificuldade”. Para isso, foram calculadas medidas para a análise de textura de áreas de interesse contendo nódulos mamários. A primeira etapa desse trabalho foi a definição do “índices de dificuldade”, baseada em características espaciais e de frequência da imagem. A segunda etapa foi dividir o conjunto de imagens em três subconjuntos, de acordo do os “índices de dificuldade” (Zheng, 1997).

Para classificar as imagens pelo grau de dificuldade, as seguintes características foram calculadas a partir do histograma de cada uma:

S_1 : a taxa entre o valor de pico no histograma e a área total de tecidos mais densos (parte mais clara) e menos densos da imagem (parte mais escura);

S_2 : a diferença entre o valor de pico e o valor médio do histograma segmentado;

S_3 : o desvio padrão da distribuição incluída no histograma segmentado;

S_4 : o desvio da distribuição do histograma; e

S_5 : o número de máximo local no histograma.

As seguintes características dependentes de frequência também foram calculadas:

S_6 : a média dos valores de cinco medidas de frequência na porção central da mama com cinco janelas parcialmente sobrepostas, cada janela de 128X128 pixels;

S_7 : o intervalo de valores de frequências nas cinco janelas (valores mínimos e máximos);

S_8 : o intervalo do número de pixels remanescentes depois de ser submetida a um filtro e à subtração de imagem na etapa de detecção de microcalcificações e *thresholding* para processamento da área total das imagens mamográficas; e

S_9 : o intervalo do número de pixels remanescentes depois de ser submetida a um filtro e à subtração de imagem na etapa de detecção de massa e *thresholding* para processamento da área total das imagens mamográficas.

Outros atributos foram calculados com base nos descritores de Haralick. As descrições dos atributos estão listadas a seguir:

1 - *Conspicuidade*: esse atributo correlaciona-se bem com a dificuldade na detecção da massa. Isso é definido como o "contraste da lesão" dividido pela "complexidade de borda". O contraste da lesão é definido pela diferença da média dos valores de pixels dentro da lesão e do fundo da imagem; enquanto a complexidade de borda é definida por $\frac{1}{N_M} \sum_{i=1}^{N_T} \|Max(I_i - I_F)\|$,

N_M e N_T são o número total de pixels dentro da região da massa e no fundo, respectivamente. $Max(I_i - I_F)$ é a diferença máxima entre os valores de pixels de fundo e sua vizinhança (para calcular usa-se uma janela de 7X7 pixels)

2 - *Circularidade*: calcula-se o tamanho da região da massa (em número de pixels N_M), e é calculado o centro da massa (centro de gravidade). Para um círculo para o mesmo tamanho da região de massa, é calculado o número de pixels localizados dentro do contorno da massa e o círculo N_C . A *circularidade* é definida por N_C / N_M , a taxa de pixels da "massa" convertida para o círculo detectado e o total de pixels dentro da região de massa;

3 - *Fator de Gross shape*: esse atributo fornece informação sobre a forma da massa e é definido como: N_b^2 / N_M , onde N_b é o comprimento em pixels do perímetro da borda da massa. Grandes valores representam maior complexidade de borda.

4 - *Desvio do comprimento radial normalizado*: o comprimento radial r_k é definido como a distância entre o centro de massa e um ponto k localizado na borda da massa. Essa característica indica mudanças microscópicas no formato da borda da massa. Ela é calculada

por $\sqrt{\frac{1}{N_b} \sum_{k=1}^{N_b} \left(\frac{r_k - m_r}{m_r} \right)^2}$, onde m é a média da distância radial.

5 - *Variação dos valores de pixel*: essa característica mede a suavização global da intensidade dos pixels na região da massa e é definida por $\sqrt{\frac{1}{N_M} \sum (I_i - \bar{I})^2}$, onde $\bar{I}$ é a média dos pixels na região da massa.

As características seguintes são propostas com o objetivo de analisar textura e ainda quantificar a densidade de regiões de interesse contendo lesões mamárias.

6 - *Fração de pixels de mínimo local*: essa característica representa a flutuação local dos pixels dentro da região de interesse;

7 - *Profundidade média do mínimo local*: essa característica mede a magnitude da flutuação local do valor de pixel dentro da região de interesse;

8 - *Gradiente médio na região de borda*: essa característica mede o contraste da massa na borda.

9 - *Densidade Integrada (DI)*: essa característica é definida multiplicando-se o tamanho da região pela média de contraste, na qual se simula o volume de uma massa.

10 - *Dimensão fractal e outros descritores de Fourier*: essas características são exploradas para caracterização de textura de lesões.

Uma descrição mais detalhada de todos esses 10 atributos pode ser vista em (Zheng, 2001).

3.2 Análise e seleção de características

A análise multivariada consiste na utilização de métodos da estatística multivariada para extração de informações dos dados (variáveis, características ou ainda descritores), que podem estar representados na forma de matrizes ou vetores. A média e a variância são descritores da estatística descritiva que resumem a variabilidade expressa nos dados contidos na matriz. A média é uma estatística descritiva que fornece uma medida de localização, um valor central para um conjunto de números. Fisicamente, a média representa o centro de gravidade. A partir da média, tida como um referencial, tem-se uma noção de distância que é importante para a interpretação da variância como medida de dispersão. A variância então pode ser entendida como a medida de dispersão entre as observações, ou seja, cada objeto de uma amostra, e a média do conjunto de observações. A Figura 3.3 representa um conjunto de 3 observações, cada qual representada por duas variáveis, e o centro de gravidade do conjunto (Monteiro, 1999).

Figura 3.3 - Gráfico representando uma matriz X com $n = 3$ observações e $p = 2$ variáveis. O vetor $\bar{X}$ denota as médias.

O vetor $\bar{X}$ contém a média para as duas variáveis, e tanto a matriz de dados originais quanto o vetor de médias podem ser representadas espacialmente como mostra a Figura 3.3. Assim tem-se a noção de dispersão dos pontos em torno da média construída a partir das observações, conduzindo à segunda estatística descritiva, a variância.

Ao extrair a raiz quadrada da variância, obtém-se o desvio padrão da amostra. Uma outra medida é a covariância entre duas variáveis em uma amostra, que é uma medida de associação linear entre duas variáveis j e k ; caso ambas sejam iguais ($j = k$), a covariância reduz-se à variância. A expressão para a covariância é um produto médio dos desvios de cada variável em relação a sua média. O valor de covariância entre duas variáveis depende da unidade de medida. Uma outra forma de expressar a associação entre variáveis é o coeficiente de correlação, que não depende da unidade de medida.

Alguns aspectos são específicos da análise de dados multivariados ou multidimensionais. Dois aspectos, em particular, são bastantes relevantes (Box, 1992):

- a redução de dimensionalidade, que envolve a determinação das variáveis mais importantes de um problema;
- a explicação da associação entre variáveis observáveis (medidas) em termos de um conjunto de variáveis não observáveis (latentes).

Tais fatores estão intimamente relacionados e compreendem princípios fundamentais no estudo estatístico de dados multivariados (Krzanowski, 1988).

Para se lidar com os dois aspectos citados anteriormente também existem vários métodos diferentes. O método adequado é escolhido de acordo com a natureza do problema.

A representação gráfica de um conjunto de dados é uma forma bastante simples para obtenção de informações sobre o relacionamentos entre os objetos de uma amostra, porém a representação gráfica de uma amostra multidimensional é limitada.

Um método bastante usado para a redução de dimensionalidade de variáveis e explicação da associação entre variáveis é a Análise dos Componentes Principais (PCA). A análise dos componentes principais é motivada como um procedimento para visualizar um conjunto de dados multivariados, como um sistema de projeção. Para a análise de um conjunto de p variáveis é necessário transformar o conjunto de variáveis originais em um novo conjunto de variáveis não correlacionadas, denominadas componentes principais. Essas novas variáveis são combinações lineares das variáveis originais, deduzidas em ordem decrescente de importância, de maneira que o primeiro componente tenha o máximo possível da variação dos dados principais. PCA é a técnica para encontrar essa transformação. O objetivo é verificar se alguns componentes podem representar a maior parte da variação dos dados originais. Se isso realmente é conseguido, então a dimensionalidade do problema é menor que p , o número original de variáveis. Com isso é possível compreender a estruturação e variação dos dados, permitindo a visualização do problema com um número menor de variáveis (Monteiro - 1999).

3.3 Métodos para análise e processamento de imagens médicas

Quando a imagem é processada, os objetos de interesse são identificados e caracterizados (geralmente por métodos de extração de características), e um algoritmo de classificação deve ser usado para analisar ou classificar cada formato. Esse procedimento pode ser feito por classificadores supervisionados ou não supervisionados, ou ainda pelo uso de técnicas híbridas que utilizam um pouco de cada uma das técnicas para uma classificação final

e mais refinada. A escolha da técnica de classificação é talvez a etapa mais complicada de um sistema de reconhecimento de padrões aplicado em medicina, bem como, em muitos casos, a definição do número de classes nas quais estarão inseridos os objetos de interesse.

O processamento de imagem médica tem como principal objetivo auxiliar o diagnóstico na prática médica. A capacidade de observação, extração, quantificação e interpretação de informações para identificar uma doença é algo subjetivo e varia de acordo com a experiência do especialista.

Na prática médica são geralmente seguidos alguns passos para análise de imagens (Toumoulin, 2002):

- aquisição;
- transmissão;
- visualização;
- análise;
- decisão; e
- armazenamento de informação e diagnóstico.

Tais passos podem incluir ainda uma comparação com outras imagens ou consulta a outros diagnósticos. O processamento de imagem pode ajudar na identificação de estruturas que levam o médico a decisões em casos pré-operatórios ou ainda pode auxiliar em terapias.

Para a Engenharia, os principais passos no processo do diagnóstico incluem a caracterização dos sensores, a formação da imagem, a extração de características, quantificação e interpretação. Esse processo sequencial não pode ser considerado uma solução ótima, mas o estudo de cada um desses passos ajuda na identificação das tarefas a serem cumpridas em cada um deles (Toumoulin, 2002).

Um trabalho desenvolvido em 2001 por Kononenko mostra uma aplicação de aprendizagem de máquinas no diagnóstico médico (Kononenko, 2001). Naquele trabalho foi feito um estudo comparativo de alguns métodos de aprendizagem de máquina, incluindo redes neurais artificiais (RNA), classificadores bayesianos simples e algumas variações e extensões de métodos de aprendizado simbólico.

Segundo Kononenko, para que um sistema de aprendizado de máquinas possa ser usado para resolver problemas em diagnóstico médico, deve apresentar as seguintes características:

- 1- Boa Performance: o algoritmo tem que ser capaz de extrair informações significativas para avaliação dos dados;
- 2- Tratamento de dados errados: muitos dados médicos apresentam erros e ruídos e algoritmos de aprendizado de máquinas aplicados à área médica devem apresentar medidas efetivas para trabalhar ruídos;
- 3- Transparência do conhecimento do diagnóstico: generalização do conhecimento e explicação da decisão devem ser transparentes para os médicos, de modo que eles possam analisar o conhecimento generalizado e observar os pontos de vista dados no problema de maneira explícita pelo algoritmo, que muitas vezes podem não ter sido observados anteriormente;
- 4- Capacidade de explicação: o sistema deve ser capaz de explicar a decisão quando do diagnóstico de uma nova paciente. Quando o médico se depara com um resultado inesperado, poderá requerer uma explicação de maneira que lhe sirva de sugestão na tomada de decisão do diagnóstico final;
- 5- Redução do número de exames: na prática médica, a coleta de dados é cara, consome muito tempo, espaço e é prejudicial às pacientes. Entretanto, é interessante um classificador que ajude a predizer o diagnóstico com um número menor de informações sobre a paciente. Assim, a seleção de subconjunto de atributos para redução de tempo de processamento e verificação dos mais significativos é um ponto essencial em algoritmos de aprendizagem de máquinas;

Para avaliar essas características, Kononenko e seus colegas testaram sete algoritmos utilizando aprendizagem simbólica, métodos estatísticos e RNA, técnicas que serão apresentadas no próximo capítulo.

3.4 Conclusão

A extração de atributos de imagens é uma tarefa que envolve não só conhecimentos de amplitude da imagem mas também da distribuição espacial. Isso implica afirmar-se que a caracterização dos atributos está relacionada às resoluções espacial e de contraste.

A relevância das informações contidas nos atributos das imagens depende do atributo escolhido para representar a imagem em análise da identificação prévia do que se pretende caracterizar daquela imagem. Então, pode-se dizer que os vários tipos de atributos devem ser relacionados de acordo com o tipo de problema a ser tratado.

Muitas vezes um conjunto de atributos pode trazer informações redundantes, o que prejudica a etapa de classificação de um padrão, quando baseada em tais atributos. Para isso, a análise e seleção de atributos são indicadas, a fim de se estabelecer um sub-conjunto de atributos antes de serem utilizados em um processo de classificação.

A extração de atributos constitui um procedimento importantíssimo num processo de classificação ou reconhecimento de padrões, pois, através de tal procedimento, é que serão geradas informações da imagem.

4

CLASSIFICADORES AUTOMÁTICOS

A tarefa de classificação ocorre numa vasta gama de atividades humanas. Nessa vastidão, o termo poderia cobrir qualquer contexto em que alguma decisão ou previsão é feita com base na disponibilidade de informações atualizadas. Dessa maneira, o processo de classificação é, então, algum método formal para fazer repetidamente tais julgamentos em novas situações.

O problema da classificação concentra-se na construção de um processo que será aplicado numa seqüência de casos contínuos, devendo cada novo caso ser associado a um conjunto de classes pré-definidas, com base em atributos e características observadas. A construção de um procedimento de classificação para um conjunto de dados em classes tem sido também designado como: reconhecimento de padrões, discriminação e aprendizagem supervisionada.

Para construção de um procedimento de classificação para um conjunto de dados em classes, nas quais as verdadeiras não são bem conhecidas, pode-se definir o chamado de aprendizagem não supervisionada e "clusterização" (agrupamento de dados por semelhança, de atributos ou características) (Trosset, 2001).

Muitas técnicas de classificação vêm sendo usadas em diversas aplicações, cada qual se adequando melhor para solucionar diferentes tipos de problemas. A questão enfatizada neste

trabalho é a classificação de achados mamográficos para auxiliar o diagnóstico de doenças mamárias, principalmente no diagnóstico precoce do câncer de mama.

As Redes Neurais Artificiais (RNA) estão sendo bastante usadas em esquemas CADs para classificação de achados mamográficos, não apenas os modelos supervisionados, mas também modelos não supervisionados e até mesmo técnicas que combinam características dos dois tipos, chamadas técnicas híbridas.

Não apenas os classificadores neurais estão sendo combinados na tentativa de se conseguir uma técnica híbrida capaz de reconhecer padrões, mas também classificadores estatísticos e não paramétricos. Os classificadores que utilizam a teoria Bayesiana vêm-se destacando, pois permitem uma descrição detalhada tanto dos atributos como da classe escolhida para uma determinada amostra.

Por isso, algumas técnicas de classificação aplicadas em esquemas CAD serão brevemente comentadas neste capítulo, para suporte bibliográfico e conceitual deste trabalho.

A combinação de classificadores envolve quatro categorias: múltiplos classificadores, usando uma representação idêntica do padrão; múltiplos classificadores, usando representações distintas do padrão; estratégias de combinações dependentes dos dados e classificadores de múltiplos estágios.

Existem também várias técnicas com diferentes critérios para análise comparativa de classificadores de padrões. Elas incluem habilidade de generalização, complexidade computacional e compreensão do espaço de características. Em muitas aplicações, como sistemas de diagnóstico, essas informações são muito importantes, pois ajudam a diminuir o risco de diagnóstico errado.

A escolha do classificador, ou de seu *design*, é um dos pontos chave para o desenvolvimento de uma aplicação. O tipo de classificador e sua arquitetura é que podem distinguir características normais e anormais de um padrão. Isso tem mostrado que a performance de um classificador para casos não conhecidos previamente depende também do conjunto de exemplos usados para seu treinamento e teste (Fukunaga, 1990).

4.1 Análise discriminante

A Análise Discriminante Linear – LDA (do inglês, *Linear Discriminant Analysis*) - é um classificador linear que pode ser derivado da média e da matriz de covariância da distribuição das classes. Para distribuição normal multivariada, o discriminante linear é uma boa ferramenta. A performance do classificador varia de acordo a configuração do conjunto de exemplos (Lanchenbruch, 1975).

Através do LDA é possível encontrar um ou mais dos seguintes itens:

- uma função discriminante para avaliar a qual classe uma amostra pertence, baseando-se somente nos valores das suas variáveis quantitativas;
- um conjunto de combinações lineares de variáveis que mostre as diferenças entre as classes;
- um subconjunto das variáveis quantitativas que represente as diferenças entre as classes (Oliveira, 1997).

Todas variedades do LDA requerem conhecimento *a priori* sobre as classes.

O LDA pode ser usado em certos casos como um classificador de padrões ou separador de aglomerados de pontos. Essa separação (ou classificação) só é possível se as duas classes estiverem no espaço euclidiano v^n ($n=2$), tal que seja possível passar um hiperplano que separe esses dois aglomerados como mostra a Figura 4.1. Se os aglomerados forem tais que torne isso possível, então são chamados de linearmente separáveis (Hertz, 1991).

Figura 4.1 – LDA separa o espaço em duas regiões

O LDA denomina uma análise que se baseia em encontrar funções discriminantes canônicas, sendo as amostras associadas às classes de acordo com a menor distância de Mahalanobis (Oliveira, 1997).

Em geral, existem m amostras para diferentes classes, de tamanhos $n_1, n_2, \dots, n_m$, e valores que estão disponíveis para p variáveis $X_1, X_2, \dots, X_p$, para cada membro da amostra (Oliveira, 1997).

O discriminante QDA está inserido no contexto de análise discriminante como um melhoramento do LDA, onde pode ser considerado um aproximador linear de funções que procura a melhor aproximação linear possível da função desejada, no sentido de mínimo erro quadrático.

4.2 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA) foram criadas com o intuito inicial de explorar e reproduzir as tarefas essenciais do processamento de informação, realizadas pelo cérebro humano, assim como a fala, audição, visão e raciocínio lógico. Mais aplicações para as redes neurais foram surgindo, na medida em que elas foram sendo investigadas e, então, passaram a ser empregadas na solução de outros tipos de problemas, como compressão de dados, reconhecimento de padrões, otimização combinatória e classificação.

A modelagem de uma rede envolve neurônio, arquitetura e algoritmo de aprendizagem. Os neurônios podem ser chamados de unidades de processamento, nós, ou ainda processadores; eles estabelecem a unidade computacional básica da rede e podem ser unidades de entrada, que recebem as entradas de fontes externas, unidades escondidas, que estão nas camadas intermediárias da rede e recebem as entradas vindas de outras unidades, enviando saídas para as próximas camadas do sistema, e ainda existem as unidades de saída, que enviam o sinal para fora do sistema. A arquitetura ou topologia da rede, como também é conhecida, determina a estrutura das conexões entre os neurônios, que pode ser esquematizada intra-camadas (colaterais, ocorrem entre neurônios da mesma camada), inter-camadas (ocorre entre neurônios de camadas diferentes) e recorrente (a conexão parte de um neurônio e chega ao mesmo neurônio). O Algoritmo de aprendizagem estabelece o método pelo qual o estado

individual de cada neurônio e suas conexões serão alterados, para que a rede atinja um comportamento global (Masters, 1994).

O primeiro passo para o desenvolvimento de uma rede é a coleta de dados relativos ao problema e a separação deles num conjunto de treinamento e um conjunto de teste. Os dados coletados devem ser significativos e cobrir de maneira ampla o domínio do problema (Patterson, 1995).

As RNA diferem, entre outros aspectos, também quanto à aprendizagem, que pode ser supervisionada e não supervisionada. No aprendizado não supervisionado, os vetores de entrada não são associados com a saída desejada no processo de treinamento; o próprio modelo extrai propriedades estatísticas do conjunto de treinamento, possibilitando agrupar vetores similares em classes, como no modelo ART (*Adaptive Resonance Theory*) e no Kohonen (Hertz, 1991). Em redes neurais de aprendizado supervisionado, o conhecimento é adquirido através de relacionamentos supervisionados, ou dos chamados "professores". Nesse caso, as entradas apresentadas à rede para o treinamento são chamadas de pares de entrada, pois os vetores de entradas contêm os valores de saída desejados, como acontece no modelo *Backpropagation* (Masters, 1994; Hertz, 1991; Patterson, 1995).

Além da divisão dos dados citada acima, pode ser usada ainda uma subdivisão do conjunto de treinamento, através da criação de um conjunto de validação, que verifica a eficiência da rede quanto a sua capacidade de generalização durante o treinamento.

Os dados podem passar por pré-processamento, tais como escalonamentos, normalizações e conversões de formato, a fim de torná-los mais apropriados a sua utilização na rede.

Uma RNA genérica deve ter :

- 1- Um conjunto de unidades de processamentos (os neurônios);
- 2- Definição de um estado de ativação;
- 3- Definição de uma função de saída para cada unidade;
- 4- Um padrão de conectividade entre as unidades;
- 5- Uma regra de propagação;
- 6- Uma regra de ativação;
- 7- Uma regra de aprendizagem;

8 - O meio ambiente.

4.2.1. Algoritmo *Backpropagation*

O algoritmo *Backpropagation* (BP) é baseado em um método estatístico proposto em 1951 por Hebert Robbins e Sutton Monro (Robbins,1951), os quais o denominaram como o método de aproximação estocástica. Várias aplicações dessa técnica têm obtido êxito em reconhecimento de padrões e sistemas de identificação. O Algoritmo BP tornou-se popular com os trabalhos de Rumelhart, Hilton e Williams em 1986, apresentados no livro PDP (*Parallel Distributed Processing*) (Rumelhart,1986).

A regra de aprendizagem mais usada no algoritmo BP é a Regra Delta Generalizada ou *Backpropagation Rule*, ou seja, uma variação da Regra Delta proposta para redes com múltiplas camadas de neurônios, as chamadas *Multilayer-perceptrons* (MLP).

O objetivo da regra de aprendizagem é aproximar as saídas desejada e obtida. Calcula-se o erro para camada de saída e propaga-se o mesmo no sentido *saída-entrada*. Os pesos vão sendo ajustados como funções destes erros. Os erros para uma unidade em particular devem ser ajustados na proporção direta ao erro nas unidades com as quais ela está conectada. Por isso, retropropagar esses erros através da rede permite que os pesos entre todas camadas sejam ajustados corretamente; assim, mudanças na entrada de uma camada qualquer da rede acarretam mudanças também nas saídas da mesma (Hertz,1991).

Seguindo o algoritmo *backpropagation* tradicional e considerando-se as variáveis seguintes, tem-se (Hertz, 1991):

w_{ij} = pesos das conexões da camada intermediária para camada de saída;

w_{jk} = pesos das conexões da camada intermediária j para camada de entrada k .

ξ_k^μ = entradas;

ζ_i^μ = saídas desejadas;

O_i = saídas obtidas pela rede;

V_j = valores das saídas dos neurônios da camada intermediária j .

As entradas são introduzidas à rede com um valor particular, sendo referenciadas por μ . O índice i sempre se refere a uma unidade de saída, o índice j a uma unidade intermediária (ou da camada escondida) e k a uma unidade de entrada. Usa-se N para o número de unidades de entrada e p para número de padrões de entradas ($\mu = 1, 2, \dots, p$).

Dado um padrão de entrada μ , as unidades j da camada escondida recebem uma entrada da rede:

$$h_j^\mu = \sum_k w_{jk} \xi_k^\mu \quad (4.1)$$

As saídas são processadas:

$$V_j^\mu = g(h_j^\mu) = g(\sum_k w_{jk} \xi_k^\mu) \quad (4.2)$$

onde V_j^μ é a saída dos neurônios da camada intermediária e g é a função de ativação da rede.

As saídas da unidade i recebem:

$$h_i^\mu = \sum_j W_{ij} V_j^\mu = \sum_j W_{ij} g(\sum_k w_{jk} \xi_k^\mu) \quad (4.3)$$

onde W_{ij} são os pesos das conexões dos neurônios da camada intermediária para os neurônios da camada.

As saídas da rede são processadas (O_i), sendo então as chamadas saídas obtidas pela rede:

$$O_i^\mu = g(h_i^\mu) = g(\sum_j W_{ij} V_j^\mu) = g(\sum_j W_{ij} g(\sum_k w_{jk} \xi_k^\mu)) \quad (4.4)$$

A função custo ou erro dada por:

$$E[w] = 1/2 \sum_{\mu i} [\zeta_i^\mu - O_i^\mu]^2 \quad (4.5)$$

torna-se

$$E[w] = 1/2 [\zeta_i^\mu - g(\sum_j W_{ij} g(\sum_k w_{jk} \xi_k^\mu))]^2. \quad (4.6)$$

A regra de atualização é local.

Considerando:

η = taxa de aprendizagem;

δ = erro;

Para calcular as saídas das camadas escondidas tem-se:

$$\begin{aligned}\Delta W_{ij} &= -\eta(\partial E / \partial W_{ij}) = \eta \sum_{\mu} [\zeta_i^{\mu} - O_i^{\mu}] g'(h_i^{\mu}) V_j^{\mu} = \\ &= \eta \sum_{\mu} \delta_i^{\mu} V_j^{\mu}\end{aligned}\quad (4.7)$$

onde δ_i^{μ} é definido por:

$$\delta_i^{\mu} = g'(h_i^{\mu}) [\zeta_i^{\mu} - O_i^{\mu}] \quad (4.8)$$

$$\begin{aligned}\Delta w_{jk} &= -\eta(\partial E / \partial w_{jk}) = -\eta \sum_{\mu} (\partial E / \partial V_j^{\mu}) (\partial V_j^{\mu} / \partial w_{jk}) \\ &= \eta \sum_{\mu} [\zeta_i^{\mu} - O_i^{\mu}] g'(h_i^{\mu}) W_{ij} g'(h_j^{\mu}) \xi_k^{\mu} \\ &= \eta \sum_{\mu} \delta_i^{\mu} W_{ij} g'(h_j^{\mu}) \xi_k^{\mu} \\ &= \eta \sum_{\mu} \delta_j^{\mu} \xi_k^{\mu}\end{aligned}\quad (4.9)$$

$$\text{com} \quad \delta_j^{\mu} = g'(h_j^{\mu}) \sum_i W_{ij} \delta_i^{\mu}. \quad (4.10)$$

Esse é o método de atualização dos pesos para redes de múltiplas camadas, conhecido como *Gradiente Descendente*.

O método do gradiente descendente é uma técnica de otimização usada para minimizar a dimensão dos erros. Ela pode ser generalizada da seguinte forma:

$$\Delta W_{pq} = \eta \sum_{\text{padrões}} \delta_{\text{saídas}} \times V_{\text{entradas}} \quad (4.11)$$

onde p e q são as conexões entre duas camadas e V as entradas e/ou saídas processadas nas camadas intermediárias (ou escondidas).

O método requer funções de ativação que variem suavemente, pois discontinuidades abruptas produzem resultados indesejáveis no espaço de estado da rede.

A função de ativação $g(h)$ é uma função sigmóide, que tem que ser diferenciável e normalmente o desejo é que ela sature para uma das extremidades. Ela pode ser dada por:

$$g(h) = \frac{1}{1 + \exp(-2\beta h)} \quad (4.12)$$

$$g'(h) = 2\beta g(1-g) \quad (4.13)$$

Desta forma a equação (4.8) é rescrita como:

$$\delta_i^\mu = O_i^\mu (1 - O_i^\mu) (\zeta_i^\mu - O_i^\mu). \quad (4.14)$$

O parâmetro β nas equações acima é um parâmetro de declividade e geralmente tem valor 1 ou $\frac{1}{2}$.

Na figura 4.2 está representada uma rede BP, onde as setas cheias indicam a fase de propagação das informações e as setas pontilhadas indicam a fase de retropropagação.

Figura 4.2 - Rede *Backpropagation*

Em geral, o funcionamento de uma rede BP pode ser entendido através de duas fases distintas: fase de aprendizagem ou treinamento e fase de uso ou teste.

Os passos de mais alto nível do algoritmo BP são sintetizados da seguinte maneira:

- 1- Inicializar os pesos com pequenos valores aleatórios (entre 0 e 1);
- 2- Escolher um par de entrada ξ_k^μ e aplicar na camada de entrada ($m = 0$) tal que

$$V_k^0 = \xi_k^\mu \text{ para todos } k;$$

- 3- Propagar o sinal através da rede diretamente usando

$$V_i^m = g(h_i^m) = g(\sum_j w_{ij} V_j^{m-1}) \quad (4.15)$$

para cada i e m até que todas saídas V_i^m tenham sido calculadas;

- 4- Computar os deltas para a camada de saída

$$\delta_i^m = g'(h_i^m) [\zeta_i^\mu - V_i^m] \quad (4.16)$$

para comparar a saída atual V_i^m com a desejada ζ_i^μ para o padrão μ inicialmente considerado.

5- Calcular os deltas das camadas precedentes para retropropagar os erros

$$\delta_i^{m-1} = g'(h_i^{m-1}) \sum_j W_{ij}^m \delta_j^m \quad (4.17)$$

para $m = M, M-1, \dots, 2$ até que o delta tenha sido calculado para todas unidades;

6- Usar

$$\Delta w_{ij} = \eta \delta_i^m V_j^{m-1} \quad (4.18)$$

para atualizar as conexões de acordo com $w_{ij}^{novo} = w_{ij}^{antigo} + \Delta w_{ij}$;

7- Voltar ao passo 2 e repetir para todos padrões.

Conjuntamente, os passos 2 e 3 são conhecidos como passos de propagação para frente (*feedforward*). Um vetor de entrada é aplicado e a saída resultante é computada. As computações são realizadas camada por camada. Os passos 4 e 5 são conhecidos conjuntamente como passo reverso (retropropagação). No passo 4 o vetor de saída produzido é subtraído do vetor de saídas desejadas do conjunto de treinamento, produzindo um erro. Esse erro é então utilizado pelo passo 5 para ajustar os pesos da rede.

A rede só é considerada treinada quando, após várias repetições desses 4 passos, o erro entre as saídas obtidas e as saídas desejadas estiver reduzido a um patamar aceitável, ou seja, um valor bastante baixo, que varia de acordo com cada aplicação.

O treinamento deve ser interrompido quando a rede apresentar uma boa capacidade de generalização e um erro global aceitavelmente baixo. Assim, o ponto ótimo de parada deve ocorrer com erro mínimo e capacidade de generalização máxima.

A fase de uso da rede BP corresponde aos passos 2 e 3 de seu processo de treinamento. Um vetor de entrada é aplicado à rede e vai progressivamente sendo propagado, através das ativações dos neurônios, desde a camada de entrada, passando pelas camadas escondidas, até atingir a camada de saída. Ao final do processo, a camada de saída manterá os valores correspondentes às entradas aplicadas.

4.2.2 ART

A técnica ART (pode ser traduzida como *Teoria da Ressonância Adaptativa*). Essa rede neural foi produzida por Stephen Grossberg e seus colegas durante os anos 70 e 80 (Patterson, 1995). As redes ART mapeam padrões de entradas n-dimensionais em categorias ou classes com base nas características dos padrões de entrada. Padrões similares são agrupados dentro de uma mesma classe e padrões não similares são separados em classes distintas. O grau de similaridade requerido para agrupar padrões em intra-classes é ajustado, de forma que muitas classes de padrões altamente similares são criadas quando o valor *threshold* (limiar) de similaridade é definido num nível alto. Por outro lado, poucas classes são criadas quando o *threshold* é definido com um valor baixo. O ajuste do *threshold* pode ser feito manual ou dinamicamente durante a operação da rede, dependendo da aplicação. Isso faz com que a rede se torne mais ou menos seletiva quanto a agrupar padrões e vai se adaptando.

O processo de aprendizagem da rede se dá em tempo real durante a operação da mesma, ou seja de uma maneira contínua. Essa técnica de aprendizagem é chamada aprendizado não supervisionado adaptativo, onde uma nova categoria é automaticamente formada toda vez que um novo padrão de entrada é apresentado à rede. Novas categorias continuarão a ser formadas de novas entradas até que a rede esgote sua capacidade de “memorizar” novas categorias a partir dos neurônios de saída, os quais rejeitarão em tempo qualquer entrada adicional de novos padrões à rede.

As redes ART são capazes de reconhecer padrões de entrada com valores reais (ou binários). Um detalhamento da técnica pode ser encontrado em (Patterson,1995; Hertz,1991), entre outros.

A arquitetura proposta por Grossberg trata a estabilidade versus a plasticidade num meio mutável. A rede é composta de duas camadas. O fluxo de dados através da rede é controlado por partes externas às camadas: controle 1 e controle 2. As entradas são comparadas pelo *threshold* de similaridade para determinar a criação de uma nova classe. A rede funciona basicamente da seguinte maneira:

- O sinal de Controle 1 determina o curso dos dados para a camada de entrada;

- O sinal de Controle 2 habilita a operação na camada de reconhecimento;
- As conexões *Feedforward* (W), onde W são os pesos, são inicializadas por um valor real $w_{ij} = 1 / n + 1$ onde n é o numero de nós de entrada;
- As conexões de *Feedback* (T) são configuradas com um valor binário representando as conexas entre todos os nós de saída para todas unidades na camada de entrada;
- O *threshold* de similaridade (θ) é tal que $0 < \theta < 1$.

4.2.3 Modelo de “Kohonen”

O método de Kohonen, um dos mais conhecidos métodos de redes competitivas, é conhecido por SOFM – *Self-Organizing Feature Map* - mapas de características auto-organizáveis. As redes SOFM se auto-adaptam aos estímulos dos padrões de entrada x que são descritos por algumas distribuições de probabilidades ($p(x)$) desconhecidas. A adaptação resultante produz uma rede que mapea padrões de entrada em padrões de saída, num caminho topologicamente coerente. O mapeamento é contínuo, preservando ordem, refletindo a distribuição de probabilidade da população de entrada. Essa propriedade da topologia preservada é guardada, uma vez que o SOFM desempenha uma redução na dimensão do espaço de atributo (Patterson, 1995; Kohonen, 1984).

O vetor de entrada é totalmente conectado com todas unidades da rede através de vetores de pesos adaptáveis. A unidade com um vetor de pesos que é o mais próximo do padrão de entrada ganha a competição. A resposta vencedora é maximizada e as outras unidades associadas a saídas iguais a zero. A ação competitiva é implementada através de conexões laterais entre as unidades vizinhas, onde são geradas conexões excitatórias e inibitórias.

O algoritmo SOFM pode ser resumido da seguinte maneira:

- 1- Inicializar os pesos w_r (números aleatórios pequenos), e configurar a taxa de aprendizagem e os parâmetros de vizinhança topológica. Repita os passos de 2 a 4 até que os pesos da rede sejam estabilizados;
- 2- Selecionar o vetor x da distribuição dos padrões de entrada para a entrada da rede;
- 3- Determinar a unidade da matriz r com o vetor de pesos para x através da equação:

$$||w_r - x|| = \min_r ||w_r - x||$$
- 4- Atualizar o vetor de pesos na próxima iteração de acordo com:

$$w_r(t+1) = w_r(t) + \alpha(t) h_{rr}(x - w_r(t)) \quad \text{para unidades } r \in N_r$$

$$w_r(t+1) = w_r(t) \quad \text{para unidades } r \notin N_r$$

onde, t é o número a iteração e N_r é a vizinhança da unidade r .
- 5- Reduzir a vizinhança e a taxa de aprendizagem;

Uma vez tenha sido processado o aprendizado suficientemente, tal que os pesos estejam estabilizados, o SOFM treinado desempenha um mapeamento para um determinado X de padrões de entrada para um centro excitatório c na rede, na qual assemelha-se a um mapa contínuo de entradas formadas sobre a rede como determinado por valores relativos dos pesos da rede. A localização do centro de excitação máxima depende da direção do vetor de entradas x , os quais são mapeados para uma posição r numa matriz bi-dimensional.

4.3 Redes Bayesianas

Um classificador Bayesiano aprende a partir das probabilidades condicionais de cada atributo (A_i) dos dados de treinamento dada uma classe (C). A classificação é feita aplicando-se a regra de Bayes para computar a probabilidade de C dado um exemplo particular de $A_1, \dots, A_n$, e então prever a classe com maior probabilidade *a posteriori*. Essa computação é interpretada fazendo-se uma forte suposição de independência: todos atributos A_i são condicionalmente independentes, dado o valor da classe C . Pela independência, entende-se independência probabilística, ou seja, A é independente de B dado C , quando $P(A|B,C) =$

$= P(A|C)$ para todos os possíveis valores de A , B e C , sempre que $P(C) > 0$ (Duda,1973).

A performance do classificador Bayesiano é um tanto surpreendente, visto que a suposição é claramente não realística. Através das Redes Bayesianas é possível manipular e representar melhor as asserções de independência dos atributos, pois utilizam linguagens apropriadas para isso.

As redes Bayesianas são grafos diretos acíclicos que permitem representação efetiva e eficiente do conjunto de distribuição de probabilidades sobre um conjunto de variáveis aleatórias. Cada nó (vértice) do grafo representa uma variável aleatória e os arcos (bordas) representam a correlação direta entre as variáveis. Mais precisamente, a rede representa as seguintes declarações de independência condicional: cada variável é independente de seus nós não descendentes num grafo, dado o estado de seus nós pais. Essas independências são então exploradas para reduzir o número de parâmetros necessários, caracterizar uma distribuição de probabilidades e computar eficientemente as probabilidades *a posteriori* dada a evidência. Parâmetros probabilísticos são representados num conjunto de tabelas, um para cada variável, na forma de distribuições condicionais locais de uma variável dados seus pais. Usando as independências declaradas representadas na rede, o conjunto de distribuição é unicamente determinado por essas distribuições locais (Duda,1973; Langley,1992).

Quando representado como uma rede Bayesiana, o classificador Bayesiano “puro” tem uma estrutura simples. Essa rede captura a idéia principal do classificador Bayesiano (“naive bayes”), sendo que todos atributos são independentes entre si, dado o estado de uma variável classe (a rota na rede).

A forma de aprendizado é não supervisionada, no sentido de que o aprendizado não distingue a variável classe das variáveis de atributos nos dados. O objetivo é induzir uma rede ou um conjunto de redes que melhor descrevam a distribuição de probabilidades sobre um conjunto de treinamento. Esse processo de otimização é implementado na prática, usando técnicas de busca heurística para encontrar o melhor candidato sobre o espaço das possíveis redes (Bacchus, 1994; Heckerman, 1995).

Considere-se um conjunto de dados finitos $U = \{X_1, \dots, X_n\}$ de variáveis aleatórias discretas, onde cada variável X_i deve conter um valor de um conjunto finito, denotado por $Val(X_i)$. Usam-se letras maiúsculas tais como X , Y e Z para nomes de variáveis, e letras minúsculas tais como x , y e z para denotar valores específicos contidos por aquelas variáveis. Um conjunto de variáveis maiúsculas é representado em negrito ficando **X**, **Y** e **Z** e os valores

associados para esses conjuntos também são representados em negrito ($\mathbf{x}$, $\mathbf{y}$ e $\mathbf{z}$). Finalmente tem-se P como sendo o conjunto da distribuição de probabilidades das variáveis em U , e $\mathbf{X}$, $\mathbf{Y}$ e $\mathbf{Z}$ sendo subconjuntos de U . Pode-se dizer que $\mathbf{X}$ e $\mathbf{Y}$ são condicionalmente independentes dado $\mathbf{Z}$ se para todos $\mathbf{x} \in Val(\mathbf{X})$, $\mathbf{y} \in Val(\mathbf{Y})$ e $\mathbf{z} \in Val(\mathbf{Z})$ $P(\mathbf{x} | \mathbf{z}; \mathbf{y}) = P(\mathbf{x} | \mathbf{z})$ quando $P(\mathbf{y}; \mathbf{z}) > 0$.

Formalmente, uma rede Bayesiana para U é um par $\mathbf{B} = \langle G; \Theta \rangle$. O primeiro componente G é um grafo acíclico dirigido, cujos nós correspondem às variáveis aleatórias $X_1, \dots, X_n$ e os arcos representam dependências diretas entre as variáveis. O grafo G representa suposições independentes: cada variável X_i é independente de seus não descendentes dado seus pais em G . O segundo componente do par, Θ , representa o conjunto de parâmetros que quantifica a rede. Esse termo contém um parâmetro $\theta_{xi | \Pi_{xi}} = P_B(X_i | \dots, X_n)$ para cada possível valor x_i de X_i ; e Π_{xi} de Π_{Xi} onde Π_{Xi} denota o conjunto de parâmetros de X_i em G . Uma rede Bayesiana $\mathbf{B}$ define um único conjunto de distribuição de probabilidades sobre U dado pela equação 4.19 (Langley, 1992).

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^n P_B(X_i | \Pi_{Xi}) = \prod_{i=1}^n \theta_{xi | \Pi_{xi}} \quad (4.19)$$

O problema de aprendizado de uma rede Bayesiana pode ser informalmente expresso como: dado um conjunto de treinamento $D = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ de exemplos de U , encontra-se uma rede $\mathbf{B}$ que melhor combinam D . Uma aproximação comum para esse problema é introduzir uma função score que avalie cada rede em razão dos dados de treinamento, e então pesquisar a melhor rede de acordo com essa função. Em geral o problema de otimização é difícil. Já para certas classes restritas das redes, existem eficientes algoritmos polinomiais.

As duas principais funções score usadas em aprendizado de redes Bayesianas são: a função score Bayesiana (Cooper, 1992; Heckerman, 1995) e a função baseada no princípio da descrição de comprimento mínimo (do inglês *Minimal Descriptor Length* - MDL) (Lam, 1994).

O princípio MDL (Rissanen, 1978) projeta o aprendizado em termos de compressão de dados, sendo o objetivo do aprendizado encontrar um modelo que facilite uma descrição mais curta dos dados originais. No contexto de aprendizagem de redes Bayesianas o modelo é uma rede, uma vez que uma rede $\mathbf{B}$ descreve uma distribuição de probabilidade P_B dos exemplos apresentados nos dados. Usando essa distribuição, pode-se construir um esquema de

representação que associe menores códigos a exemplos mais prováveis. De acordo com o princípio MDL deve-se escolher uma rede B , desde que o comprimento descrito na rede, combinado com a representação dos dados (em razão de P_B), seja minimizado.

Um outro método de treinamento de redes Bayesianas se baseia na teoria da informação com técnicas de análise de dependência, onde o relacionamento de independência condicional é uma regra muito importante, da qual com a condição de Markov, obtém-se uma coleção de estados de independências condicionais de uma rede Bayesiana (Cheng, 1998).

Seguindo o método de treinamento baseado na teoria da informação com técnicas de análise de dependência, pode-se ver uma rede como canais de sistema de informação, onde cada nó é uma válvula que está ativa ou inativa e as válvulas são conectadas pelos canais de informações (os arcos). O fluxo de informação só pode passar por uma válvula ativa. Quando todas as válvulas em um caminho adjacente entre dois nós estão ativas, pode-se dizer que tal caminho está aberto; do contrário, está fechado.

No algoritmo proposto por Cheng e Liu, mede-se o volume do fluxo de informação entre dois nós para ver se um grupo de válvulas correspondentes a um conjunto de condições pode reduzir e eventualmente bloquear o fluxo de informação. Esses resultados direcionam a construção de uma estrutura de rede Bayesiana correta para um determinado conjunto de dados. Para medir o volume do fluxo de informação, eles usaram informação mútua e medidas de informação condicional mútua. Em teoria da informação, informação mútua é usada para representar o ganho de informação esperado no envio de algum símbolo e recebimento de um outro. Em redes Bayesianas, se dois nós são dependentes, sabendo-se o valor de um nó, obtém-se informações sobre os valores de outros nós. Essa informação ganha pode ser medida através de informação mútua. Assim, a informação mútua entre dois nós pode dizer se são independentes e como é o relacionamento entre eles.

No aprendizado, o algoritmo tem como entrada um conjunto de dados dispostos em uma tabela. Cada atributo (ou seja, cada campo da tabela) é tratado como uma variável aleatória, a qual é representada por um nó na rede Bayesiana. A probabilidade marginal e a probabilidade condicional são estimadas usando-se cálculo das frequências relativas dos dados da tabela.

Usando-se o método de treinamento MDL discutido anteriormente, pode-se induzir uma rede Bayesiana B que represente uma distribuição $P_B(A_1, \dots, A_n; C)$ a partir de um dado conjunto de treinamento. Então, usando o modelo resultante, tal que dado um conjunto de

atributos $a_1, \dots, a_n$, o classificador baseado em B retorna o rótulo c , que maximiza a probabilidade *a posteriori* $P_B(c | a_1, \dots, a_n)$.

4.4 "Clusterização" - Classificador *K-Means*

Os algoritmos de "clusterização" são métodos que dividem um conjunto de n observações em g grupos (*clusters*), tal que membros do mesmo grupo sejam mais parecidos que membros de um grupo diferente (Ripley, 2000).

O número de grupos pode ser pré-associado ou decidido pelo algoritmo. Formalmente, um algoritmo de "clusterização" deve mapear $c: \{1, \dots, n\} \rightarrow \{1, \dots, g\}$, associando um grupo com todos exemplos.

Alguns algoritmos de "clusterização" trabalham com a representação de cada grupo por um ponto representativo (informação extraída do vetor de atributos) e isso tem ligações próximas com a quantização de vetores utilizada nas redes SOFM (Ripley, 2000).

Na prática, parte do conhecimento ou hipóteses razoáveis sobre os dados ou processos através dos quais o ponto representativo é produzido é altamente valiosa e deveria ser cuidadosamente considerada, enquanto se decide por um critério de "clusterização" específico. Entretanto, a falta de informação sobre a estrutura natural dos *clusters* faz com que se considerem muitos critérios e se julgue o resultado num caminho subjetivo. Um dos critérios de "clusterização" mais adotados neste caso é a somatização (Costa, 2001).

4.4.1 Matrizes de espalhamento e medidas de dispersão

Uma maneira de formalizar o critério de similaridade da "clusterização" pode ser através da matriz de espalhamento. Conceitualmente, para qualificar como candidato a *cluster*, cada conjunto de pontos deve incluir elementos que são pouco dispersos. Ao mesmo tempo, deve-se esperar uma dispersão relativamente alta entre os pontos ao longo de diferentes *clusters*.

A matriz de espalhamento total, S indica a dispersão dos vetores de atributos. A matriz de espalhamento para a classe C_i , a partir de S_i , expressa a dispersão do vetor de atributos dentro de cada classe. A matriz de espalhamento intraclases, S_{intra} indica a dispersão combinada em cada classe. E a matriz de espalhamento interclasses, S_{inter} expressa a dispersão das classes (em termos de seus centróides). Através da soma das matrizes de espalhamento intra e interclasses tem-se, então, a determinação da classe (Costa, 2001).

4.4.2 “Clusterização” Não-Hierárquica

Através da "clusterização" não-hierárquica (também chamada método particional), os *clusters* são obtidos como uma partição definida do espaço de atributos a respeito de um número fixo de *clusters*. Um algoritmo simples de "clusterização", utilizando método particional, pode ser obtido diretamente em termos de sinais de medidas de dispersão, que podem ser usados para implementar a similaridade do critério de "clusterização". Assim, uma boa "clusterização" deve exibir uma baixa dispersão intraclases e alta dispersão interclasses. Entretanto, a dispersão global é preservada.

O algoritmo pode ser mostrado através dos seguintes passos:

- 1- atribuir classes aleatórias para cada objeto;
- 2- selecionar um objeto e mudar aleatoriamente sua classe, evitando descartar qualquer classe;
- 3- verificar o sinal de dispersão interclasse medido através da matriz de espalhamento; se o valor de dispersão interclasses aumentar, associar novamente o exemplo à classe original.

Os passos 2 e 3 devem ser repetidos até que estejam estabilizados os valores de dispersão interclasses (Costa, 2001).

A condição de parada do algoritmo envolve identificar quando os *clusters* foram estabilizados, sendo então arquivados; ou seja, isso ocorre quando o número de classificações inalteradas sucessivas excede um limiar pré especificado. Nesse algoritmo, um ponto importante é que o número de classes é pré-definido. Isso é uma consequência do fato de que

a dispersão intraclassas tende a diminuir com um grande número de *clusters*, ocorrendo o oposto, se o número de *clusters* for pequeno.

4.4.3 *K-means*

K-means é uma técnica de classificação baseada no critério de mínima dispersão intraclassas. *K-means* começa com N objetos, caracterizados por seus respectivos vetores de atributos, e tenta classificá-los em K classes. Entretanto, o número de classes tem que ser previamente conhecido. Esse método requer K protótipos iniciais P_i , os quais podem ser providos por uma caracterização supervisionada ou estimados automaticamente, como, por exemplo, por uma seleção aleatória de K pontos (se possível, uniformemente distribuída ao longo do espaço de atributos da região considerada). O método *K-means* é formalizado seguindo o seguinte procedimento:

- 1- obter os K pontos dos protótipos iniciais e armazená-los numa lista W ;
- 2- calcular todas distâncias entre cada protótipo P_i (ou média) e cada vetor de atributos, produzindo-se uma matriz de distâncias D de $K \times N$;
- 3- Usar a Matriz D para identificar os atributos que são próximos de cada P_i (isso pode ser feito encontrando os mínimos valores em cada coluna de D);
- 4- Armazenar os pontos dentro da respectiva lista L_i ;
- 5- Obter como novos protótipos os centróides das características armazenadas dentro de cada L_i respectiva;

Os passos de 2 a 5 devem ser repetidos até que se estabilize o número de centróides de classes. Um possível critério de estabilidade corresponde à situação na qual o máximo deslocamento de cada centróide é menor que o *threshold* definido previamente (Costa, 2001).

Um outro método de implementação do *K-means* é encontrado em (Ripley, 2000). Como um método de particionamento, o algoritmo divide os exemplos dentro de um número pré associado de *clusters*. Para dados em um espaço Euclidiano, pode-se associar o centro de um *cluster* m para cada grupo. Então, escolhem-se os centros dos *clusters* e dos grupos, de tal forma que seja mínima a distância da soma dos quadrados de cada exemplo para o centro do *cluster*.

$$\text{Min}_{c, m_j} \sum ||\mathbf{x}_i - \mathbf{m}_{c(i)}||^2 \quad (4.20)$$

A minimização sobre os centros dos *clusters* é simples: basta escolher o centro do *cluster* j para ser a média dos exemplos associados ao *cluster* j ; assim, conhecem-se os *clusters* c , o que é suficiente para se definir os centros dos *clusters*. Todos exemplos são associados simultaneamente a seus centros de *clusters* mais próximos, cada centro de *cluster* movido para a média do grupo, e esse processo é repetido. Um grupo pode ser descartado pelo algoritmo, e então pode ser escolhido menos que K *clusters*. Uma solução apresentada é recalcular os centros dos *clusters* a cada nova entrada de exemplo.

Muitas variações do *K-means* podem ser implementadas, conforme vários algoritmos apresentados em (Gose, 1996).

4.5 Análise de resultados da classificação

Existem vários métodos de avaliação dos resultados provenientes de um processo de classificação. Uma análise incorreta pode gerar interpretações conflitantes e conclusões errôneas (Tosset, 2001). Especialmente na área médica, vem sendo utilizada uma técnica de análise de resultados desenvolvida por Metz (Metz, 1986), chamada Curvas ROC (do inglês, *Response Operating Characteristic*).

O método das curvas ROC consiste em mostrar para o observador imagens cujo conteúdo pode - ou não - apresentar anormalidades e verificar se o observador encontrou ou não essas anormalidades. Existem apenas duas situações possíveis: a imagem contém o sinal indicador de uma doença, por exemplo, ou não. O observador, por sua vez, pode responder afirmativa ou negativamente em relação a esse aspecto. Logo, o diagnóstico verdadeiro-positivo (VP), que acontece quando o observador percebe o sinal realmente existente na imagem, será representado por uma função de probabilidade $p(S/s)$ e o verdadeiro-negativo (VN), quando o observador percebe que não há sinal (inexistente realmente) na imagem, será representado por $p(N/n)$ (Evans, 1981).

Porém, pode acontecer do observador interpretar um ruído como sinal e o diagnóstico se torna falso-positivo (FP), representado por $p(S/n)$. Ou então, ele interpreta um sinal como ruído e o diagnóstico passa ser falso-negativo (FN), representado por $p(N/s)$.

Se o sistema de imagem e o observador fossem ideais, transferiria a informação perfeitamente e as probabilidades de acerto e erro seriam um e zero, respectivamente. Ou seja:

$$\begin{aligned} p(S/s) &= 1 & p(N/n) &= 1 \\ p(S/n) &= 0 & p(N/s) &= 0; \end{aligned}$$

Mas, num sistema real, essas probabilidades passam a ser fracionais, dados os índices de erro do observador.

Uma curva ROC é uma descrição empírica da capacidade do sistema de diagnóstico discriminar dois estados num universo. Cada ponto da curva representa um compromisso diferente entre a fração VP e a fração FP, que pode ser registrado pela adoção de um valor diferente de corte quanto ao que é ou não anormalidade ou nível crítico de confiança no processo diagnóstico. Sob o ponto de vista da teoria de testes de hipóteses estatísticas, uma curva ROC é conceitualmente equivalente a uma curva que mostra a relação entre a potência de teste e a probabilidade de cometer um erro.

A curva ROC é gerada a partir de mudanças sucessivas nos critérios do observador, a partir do processo comentado acima. Cada ponto da curva descreve o critério do observador para distinguir o que ele interpreta como sinal e o que ele interpreta como ruído. Esses pontos são os pontos de operação da curva ROC (Evans, 1981).

Para que a curva indique qual critério apresenta maior número de acertos e o menor número de erros, é necessário definir qual o ponto ideal de operação das curvas ROC; ou seja, qual o critério que trará a menor probabilidade de erro, qual trará maior quantidade de informação e qual resultará em um menor custo para o paciente.

O falso-positivo e o falso-negativo são os dois erros que podem acontecer em procedimentos de diagnósticos por imagens. Então, a probabilidade dos erros existirem é dada pela seguinte equação:

$$p(\text{erro}) = p(S/n) \cdot p(n) + p(N/s) \cdot p(s) \quad (4.21)$$

Os erros são minimizados quando o declive da curva ROC é igual à razão entre probabilidade da doença estar presente e a probabilidade da doença estar ausente. Ou seja, pela equação:

$$\frac{\delta p(\text{erro})}{\delta p(S/n)} = 0 \quad (4.22)$$

ou ainda quando:

$$\frac{\delta p(S/s)}{\delta p(S/n)} = \frac{p(n)}{p(s)} \quad (4.23)$$

usando $p(N/s) + p(S/s) = 1$

Utilizando esse parâmetro para seleção de um ponto ótimo de operação nas curvas ROC, é possível encontrar o critério que proporciona uma menor probabilidade de erro por parte do observador.

A precisão (P) é um índice que indica a fração dos casos em que o observador esteja correto. As precisão é calculada em termos da sensibilidade e da especificidade.

$$\text{Sensibilidade: } \frac{[\text{número de decisões verdadeira-positivas(VP)}]}{[\text{número de casos realmente positivos}]}$$

$$\text{Especificidade : } \frac{[\text{número de decisões verdadeira-negativas(VN)}]}{[\text{número de casos realmente negativos}]}$$

$$P = \text{sensibilidade} \times \left[\begin{array}{c} \text{Fração da população,} \\ \text{em estudo, que é} \\ \text{realmente positiva} \end{array} \right] + \text{especificidade} \times \left[\begin{array}{c} \text{Fração da população,} \\ \text{que é realmente} \\ \text{negativa} \end{array} \right]$$

A sensibilidade é um parâmetro que indica acertos em casos positivos e a especificidade indica acertos em casos negativos. Esses parâmetros representam uma das maneiras de quantizar os diagnósticos VP e VN respectivamente. E a medida de precisão dá uma informação de confiabilidade do sistema.

Numa descrição mais detalhada das curvas ROC, é possível verificar que a máxima informação transferida por seu ponto de operação pode ser obtida graficamente pela

sobreposição de isoinformações das curvas ROC plotadas. Para isso, é necessário o conhecimento da probabilidade do sinal estar presente (Evans, 1981).

As funções densidade de probabilidade normais de igual variância foram as primeiras a ser consideradas. Nessa situação, o índice de discriminação mais utilizado é:

$$d = \frac{(\mu_A - \mu_N)}{\sigma_N} \quad (4.24)$$

Isso expressa a diferença entre as médias das duas funções: densidade de probabilidade em termos do desvio padrão, onde μ_A representa média de casos positivos, sob uma determinada observação, e μ_N a média de casos negativos. A figura 4.3 mostra um exemplo de distribuição de duas classes analisadas.

Figura 4.3 – Distribuições de duas classes sob a análise de um parâmetro.

O desvio padrão na curva ROC representa a variação que pode ser esperada em qualquer ponto traçado da curva.

Segundo Metz (Metz, 1986), uma curva ROC representa o compromisso que pode ser obtido entre sensibilidade e especificidade de um sistema de diagnóstico, quando o *threshold* (limiar) é variado. Uma comparação entre os sistemas pode ser feita a partir das respectivas áreas das curvas de cada sistema em questão. A Figura 4.4 mostra um exemplo de curva ROC.

Figura 4.4 – Exemplo de curva ROC

A área sob a curva ROC é um dos índices mais utilizados para sumariar o significado da curva e, assim, do sistema que está sendo analisado. Representa o índice de acerto do sistema (classificador), ou seja, quanto maior a área, maior é o número de acertos. Isso significa que num sistema bastante calibrado e com uma alta precisão, a curva deve estar o mais próxima possível do canto superior esquerdo do eixo cartesiano, aumentando a área sob a curva. Pela análise, quando o verdadeiro-positivo é alto, o falso-negativo deve ser baixo, pois o sistema ideal deve ter probabilidades de falso-positivo e falso-negativo iguais a zero; por isso, no sistema real, obviamente deseja-se que esses índices sejam os menores possíveis.

A área pode ser calculada de três maneiras diferentes:

1. Regra trapezoidal;
2. Método de avaliação de máxima probabilidade (Dorfman, 1969);
3. Rampa de interceptação dos dados originais quando traçados num gráfico binomial (Hanley, 1983).

A principal vantagem desse índice é que sua utilização não depende da distribuição associada aos dados.

4.6 Classificadores aplicados em esquemas CAD

Num estudo feito por Chan *et al.*, em 1999 (Chan, 1999), foi investigada a performance de 3 diferentes classificadores utilizados em esquemas CAD: análise discriminante linear - LDA (do inglês *Linear Discriminant Analysis*), discriminante quadrático - QDA (do inglês *Quadratic Discriminant Analysis*) e uma Rede Neural *Backpropagation*.

No trabalho de Chan e seus colegas, quatro tipos de distribuições de classes que têm propriedades específicas foram utilizados para comparar o desempenho dos classificadores: distribuição normal, com matrizes de covariâncias iguais e médias desiguais; com matrizes de covariâncias desiguais e médias desiguais, com matrizes de covariância desiguais e médias iguais, e um espaço de características, onde as duas classes foram uniformemente distribuídas em regiões desconexas.

Foram avaliadas as performances dos classificadores em espaços de características de dimensões variando de 3 a 15, e o tamanho do conjunto de exemplos de 20 a 800 por classe. A dependência da substituição e permanência da configuração do tamanho do conjunto de exemplos no treinamento foram investigadas.

Para os autores, a concepção de generalização pode ser de vários aspectos técnicos, quando avaliada a performance de um classificador: um em relação a medir a performance do classificador; outro face à variância da performance do classificador. Para avaliação em sistemas de diagnóstico médico, o índice mais comumente utilizado é a área sob a curva ROC (A_z) (Metz, 1986). Eles observaram que para, o classificador LDA, a performance do sistema em termos de A_z pode ser aproximada por uma função linear em $1/N_i$ (onde N_i é o tamanho do conjunto de exemplos).

Para classe de distribuição normal multivariada, com matrizes de covariância iguais, os LDA é o classificador ótimo. Em espaços de características com matrizes de covariância desiguais onde o QDA é ótimo, o LDA é inferior ao QDA e ao *Backpropagation* quando o tamanho do conjunto de exemplos é grande. Entretanto, quando o conjunto de exemplos é pequeno, um classificador relativamente simples, tal como o LDA ou uma rede neural com poucos neurônios intermediários, pode ser preferido porque a performance parcial aumenta com a complexidade do classificador. Num regime onde a performance do classificador é aproximada por $1/N_i$, a performance no limite de conjunto de exemplos de tamanho infinito

pode ser estimada como a retenção ($1/N_i = 0$) de uma regressão linear de A_z versus $1/N_i$. A compreensão da performance de um classificador sob o constrangimento de conjunto de exemplos de tamanho finito é esperada, para facilitar a seleção de um classificador apropriado para uma dada tarefa de classificação e desenvolvimento (configuração) de um esquema eficiente (Chan, 1999).

Um outro trabalho do mesmo grupo de pesquisadores, também em 1999, mostra um classificador híbrido utilizando a combinação de uma técnica de classificação não supervisionada, o ART2 (do inglês *Adaptive Resonance Network*), e uma técnica supervisionada, o LDA, para classificação de massas encontradas em mamogramas, baseada na similaridade das entradas (dos vetores de características) (Hadjiiski, 1999).

No trabalho de Hadjiiski, as classes de massas resultantes são divididas em dois grupos: classes contendo somente massas malignas; e classes contendo ambas, maligna e benigna ou somente massas benignas. Todas massas pertencentes ao segundo grupo foram usadas para formular um modelo LDA simples, a fim de classificá-las como benignas e malignas. Nesta aproximação, a rede ART2 identifica os casos altamente suspeitos de malignidade e os remove do conjunto de treinamento, facilitando, assim, a formulação do modelo LDA. Para se examinar a utilidade dessa aproximação, uma base de dados de 348 ROI contendo massas comprovadas por biopsia (149 benignas e 179 malignas) foi usada. Dez diferentes grupos de treinamento e testes foram gerados aleatoriamente, onde 73% das ROI foram usadas para treinamento e os 27% restantes para os testes.

O classificador projetado inclui seleção de características e a otimização dos pesos foi feita a partir dos grupos de treinamento. Os grupos de testes foram mantidos independentes dos grupos usados no treinamento.

A performance do classificador híbrido foi comparada à de um classificador LDA simples. O desempenho do classificador é avaliado com análises de curvas ROC (Metz, 1986), sendo a área sob a curva A_z igual 0,81 para o classificador híbrido e A_z igual a 0,78 para o classificador LDA. Com esses resultados, os autores concluíram que a classificação híbrida promete melhorar a exatidão da classificação nas aplicações em esquemas CAD.

Um trabalho de Grohman e Dhawan (Grohman, 2001) descreve uma rede neuro-Fuzzy baseada em conjunto de dados convexos para classificar exemplos de câncer de mama.

Eles introduziram um novo método de reconhecimento de padrões, o classificador de padrões neuro-Fuzzy – NFPC (do inglês, *neuro-fuzzy pattern classifier*), que incorpora

informações *fuzzy* dentro da superfície de decisão para melhorar o desempenho da classificação. Eles compararam a performance do NFPC com a performance de uma rede “Back-propagation”.

O método de treinamento do NFPC primeiramente identifica agrupamentos dentro dos dados de treinamento, e então constrói a arquitetura da rede, com isso aumenta a robustez do sistema e também o perigo da função de erro mínimo cair no mínimo local é minimizado; a classificação é baseada num agrupamento de dados real.

No espaço de características é separado um certo numero de pontos para treinamento de todas categorias. Isso é arquivado para dividi-los em somente pontos homogêneos (contendo somente pontos de uma categoria); os não sobrepostos formam um subconjunto convexo, e, então, ambos são separados por hiperplanos que passam entre subconjuntos vizinhos de diferentes categorias. Isso completa o desenho de uma rede de camadas de hiperplanos.

Existem vários procedimentos que podem ser usados para criar camadas. Em qualquer caso, desde que o hiperplano de separação dos subconjuntos obtidos resulte na criação de um conjunto homogêneo de regiões convexas, a camada consecutiva da rede serve para determinar a qual região uma dada entrada pertence. Uma função de pertinência fuzzy M_f é imaginada para cada subconjunto convexo criado ($f = 1, 2, \dots, K$). A classificação de decisão é feita pela camada de saída, baseando-se no princípio “winner- take-all”. A categoria resultante (C) é a categoria do conjunto convexo com maior valor da função de pertinência.

A construção do modelo NFPC envolve quatro estágios:

- Criação de conjunto convexo – eles devem ser homogêneos e não sobrepostos. Para satisfazer a condição de homogeneidade, é preciso encontrar um método de achar pontos de uma categoria dentro de outra.
- Seleção de um subconjunto de pontos iniciais – o algoritmo requer um ponto de partida do subconjunto convexo para cada categoria. Existem muitas maneiras de encontrar esses pontos de partida. Eles podem simplesmente ser escolhidos aleatoriamente pela média das coordenadas dos pontos de todas categorias. Nesse trabalho, os pontos de partida foram selecionados de maneira um pouco grosseira, onde o espaço de pontos foi dividido em áreas e o ponto de partida escolhido foi na área de maior concentração de pontos.

- Colocação de hiperplanos – criação de camada hiperplano - uma vez encontrados os subconjuntos convexos, supõe-se que eles não são sobrepostos, de tal forma que somente um hiperplano é necessário para separar dois subconjuntos vizinhos. O programa verifica subconjuntos de todas categorias e coloca um hiperplano entre dois conjuntos de diferentes categorias que ainda não tenham sido separados por hiperplanos já existentes. Assim, o número de hiperplanos pode variar, dependendo do conjunto de treinamento.

- Construção da função de pertinência *fuzzy* – os hiperplanos colocados definem as regiões convexas treinadas para uma amostra apresentada. Essas regiões criadas são as bases para construção das funções de pertinência *fuzzy*, que representam a relação entre os pontos em um subconjunto convexo em cada categoria.

As entradas dos neurônios são uma camada hiperplana, criadas como no estágio previamente discutido. A saída do neurônio é uma função de pertinência fuzzy M_j para o subconjunto convexo f .

A saída do classificador é a saída C da função de pertinência fuzzy do conjunto convexo que atinge o maior valor para um padrão de entrada específico. Uma superfície de decisão para cada categoria pode ser determinada pela união fuzzy de todas funções de pertinência dos subconjuntos convexos pertencentes a essa categoria.

Para avaliar o desempenho do método, foram utilizados 191 casos mamográficos com microcalcificações. Deles 128 são benignos e 63 malignos. As imagens selecionadas foram digitalizadas com 256 níveis de cinza.

As características foram extraídas das imagens através de estatísticas do histograma de segunda ordem (10 características) (Gonzales, 1992). Depois, técnicas *Wavelets* foram aplicadas nas regiões contendo microcalcificações e 20 características com entropia e energia foram calculadas. Em seguida as imagens foram segmentadas a fim de se obter informações para representar os *clusters* de microcalcificações, através de mais 10 características.

Os autores afirmaram que os resultados, utilizando-se o método NFPC, foram melhores do que utilizando-se uma rede *Backpropagation* na maioria dos testes.

No ano de 2001, Allan e Kinsner mostraram um trabalho com imagens microscópicas de tecidos de mama, usando uma rede neural competitiva para associar atributos quantitativos

a quatro diferentes tipos de doenças de mama, sendo duas benignas e duas malignas (Allan, 2001).

Os atributos, de modo geral, abrangem limiarização e detecção de bordas. Essas características são extraídas de maneira supervisionada ou não supervisionada. Para inspeção visual, não há distinção nítida entre doenças malignas e benignas. Através de técnicas de processamento (Gonzales, 1992) e análise de imagens (Haralik, 1993), foi encontrada uma correlação das características extraídas com doenças malignas e benignas, uma vez que tais características não são visíveis ao olho humano.

Nesse trabalho, as imagens microfotográficas de tecidos de mama apresentando quatro diferentes doenças (carcinoma lobular infiltrado, carcinoma ductal infiltrado, arranjos fibrocísticos e fibroadenomas) foram adquiridas utilizando-se uma câmara CCD com resolução de contraste de 256 níveis de cinza e resolução espacial de 1260X960 pixel e um microscópio *Zeiss Axioshop*, com magnificação de 400x. O formato dos arquivos das imagens é TIFF.

Cada imagem foi segmentada em blocos de pixels. Os blocos são varridos da esquerda para direita e de cima para baixo. Foram estudados blocos de diferentes tamanhos. Os tamanhos variaram de 4X4 a 60X60. Dois tipos de blocos diferentes foram analisados: bloco do tamanho inteiro (não sobreposto) e bloco com a metade do tamanho (sobreposto). Cada pixel do bloco foi então contido numa normalização, num vetor unidimensional, cujos valores ficaram entre 0 e 1. Cada vetor é uma coluna do conjunto de treinamento da rede neural.

O conjunto de treinamento obtido de diferentes imagens é apresentado para diferentes tipos de redes competitivas: 1) uma rede neural Kohonen convencional com uma consciência; 2 a 4 redes neurais com mapas de características auto-organizáveis (SOFM) de uma, duas e três dimensões; e 5) uma rede neural LVQ – “learned vector quantization”. No caso da rede LVQ, a rede é instruída, já que existem quatro diferentes tipos de doenças de mama para ela descobrir.

Os pesos no treinamento da rede neural competitiva, bem como os pesos da primeira camada da rede LVQ, formam as características gerais do bloco de pixels que são extraídas das imagens pelas redes. O número de características extraídas de cada rede neural variou entre 24 e 75.

Cada rede foi treinada com um certo número de épocas, então parado e testado. O número de épocas variou de 1000 a 2 milhões. Para verificar se a rede estava correta, o conjunto de treinamento foi apresentado como conjunto de testes. No caso da rede Kohonen

clássica, o segundo classificador não foi usado no experimento. A habilidade dessa rede neural para identificar características de bloco de pixels foi avaliada.

O autor principal (que é patologista) julgou útil para o diagnóstico essas características de bloco de pixels em cada imagem. A habilidade das redes competitivas melhorou o desempenho com mais treinamento e também a avaliação qualitativa foi melhorada.

Os autores concluíram que as redes competitivas foram capazes de extrair características gerais das imagens que são importantes e têm um reconhecimento significativo para os olhos. O processo imita como a inspeção visual poderia fazer para analisar uma cena ou região de interesse, olhando para características simples e características mais gerais. Com esse trabalho, o objetivo foi desenvolver uma nova ferramenta capaz de identificar características visuais em imagens de difícil análise e a inteligência artificial tem contribuído para essas aplicações (Allan, 2001).

Um outro trabalho utilizando RNA em esquemas CAD bastante interessante foi desenvolvido por Zheng *et al.* em 2001 (Zheng, 2001), onde os autores investigaram um novo método de otimizar RNA com filtros adaptativos usados em esquemas de auxílio a detecção em mamogramas digitalizados e, para mudanças em performances nas taxas médias de acertos de classificação para três diferentes conjuntos de esquemas.

Eles utilizaram duas bases diferentes de imagens no estudo: uma, contendo um total de 978 mamogramas digitalizados, dos quais, 545 imagens foram adquiridas de pacientes submetidas a exames mamográficos no Centro Médico da Universidade de Pittsburgh; as outras 433 imagens foram providas de um grupo de imagens da Escola Médica da Universidade de Washington. A segunda base de imagens (usada para teste), continha 830 imagens. Todos mamogramas foram digitalizados em digitalizador de filmes laser *Lumisys*, com tamanho de pixel de 100 micros e 12 bits de resolução de contraste. Para detecção de massas, nas imagens foram recortadas regiões de 600X450 pixels, onde todas massas verdadeiro-positivas foram comprovadas por exames anato-patológicos e a localização das massas foi marcada por radiologistas.

Cada mamograma foi processado seguindo os seguintes procedimentos:

- uso de um filtro dual-Kernel;
- subtração;
- métodos de *thresholding*;
- identificação de grande número de massas suspeitas;

- extração de conjunto de atributos da imagem;
- classificação em regiões positivas ou negativas.

O esquema tem três estágios distintos de identificação de massas. O primeiro é através do filtro dual-kernel, seguido de subtração e rotulação de grande número de regiões suspeitas. O segundo, baseado na medida de contraste local, usou um algoritmo de crescimento de região adaptativo para definir três camadas topográficas para cada região suspeita. Para cada camada, um simples conjunto de condições de borda na taxa de crescimento de região e um fator de forma foram aplicados para eliminar o grande número inicial de regiões suspeitas. Depois do segundo estágio, o número de regiões suspeitas (incluindo positivas e negativa) diminuiu nos conjuntos de treinamentos e testes. Para cada região suspeita foram extraídos atributos computados automaticamente pelo esquema, que incluem principalmente características geométricas como tamanho da massa, circularidade, desvio padrão do comprimento radial e atributos de intensidade, tais como contraste, desvio padrão, distribuição dos pixels e conspicuidade. Para reduzir a redundância contida nos atributos foi utilizado o Algoritmo Genético (GA), para selecionar um subconjunto ótimo de atributos. Usando esses atributos, o terceiro estágio, usou RNA *feedforward* de três camadas, de forma a classificar as regiões como massas positivas e negativas.

Eles nomearam RNA adaptativas quando subdividiram o conjunto de treinamento em 3 subconjuntos de características de entradas similares, utilizando GA. Nas redes nomeadas não adaptativas o conjunto de treinamento não foi dividido. A média de acertos de três RNAs adaptativas para observações independentes atingiu A_z de curva ROC de 0,95.

Um outro trabalho desse mesmo grupo, em 2003, mostra um estudo das mudanças de performance de esquemas CAD para um mesmo conjunto de casos, utilizando imagens recentes e anteriores processadas previamente (Zheng, 2003).

Os autores utilizaram uma base de 134 casos de mamogramas, totalizando 260 pares de imagens (anterior e atual), cada imagem com uma ou duas ocorrências de massas. Todas 520 imagens (260 anteriores e 260 atuais) foram processadas pelo esquema CAD, desenvolvido em 2001 (Zheng, 2001). Como resultado obtiveram boa performance de classificação, depois da rede ter sido re-treinada para as imagens atuais, e dessa forma, ao testar o classificador com as imagens anteriores, o desempenho do classificador foi inferior. A conclusão dos autores foi que, no re-treinamento da rede, devem ser incluídos tanto exemplos das imagens atuais quanto das anteriores.

As Redes Bayesianas, também conhecidas como Redes de Crença Bayesina (BBN – *Belief Bayesian Network*), vêm sendo investigadas em aplicações de sistemas de suporte a decisões médicas, tornando-se uma importante ferramenta em esquemas CAD, pois é uma técnica que usa teoria de probabilidade para lidar com incertezas.

Foi realizado um trabalho, focando as duas primeiras categorias de combinação de classificadores: múltiplos classificadores, usando uma representação idêntica do padrão; e múltiplos classificadores usando representações distintas do padrão; conforme citado anteriormente, utiliza-se múltiplas decisões inteligentes e endereça o problema de combinação de classificadores, onde a representação usada não se encaixa nem em casos compartilhados nem nos casos distintos. Nesse trabalho Kittler *et al.* consideram casos onde as características empregadas em cada classificador são parcialmente partilhadas com outro classificador e parcialmente distintas (Kittler, 1997).

O esquema classificador de Kittler foi desenvolvido para aplicação em sistemas de detecção automática de microcalcificações em imagens mamográficas, onde foram implementadas opiniões de quatro diferentes sistemas inteligentes para melhorar a performance do sistema de detecção.

Considerando-se um problema de reconhecimento de padrões, o padrão Z deve ser associado com uma das m possíveis classes $\{\omega_1, \dots, \omega_m\}$. Deve-se supor que para R atributos cada padrão é representado por um vetor de medidas x_i $i = 1, \dots, R$. Os componentes do vetor x_i de cada padrão podem ser divididos em dois grupos, formando vetores y e ξ_i i.e. $x_i = [y^T, \xi_i^T]^T$, onde o vetor de medidas y é partilhado por todos os R classificadores, enquanto ξ_i é específico para o i -ésimo classificador. Foi assumido que, dada a identificação de uma classe, para o classificador a representação específica ξ_i do padrão é condicionalmente independente de $\xi_j, j \neq i$.

No espaço de medidas, cada classe ω_k é modelada pela função densidade probabilidade $p(x_i | \omega_k)$ e a probabilidade (*a priori*) de ocorrência é denotada por $p(\omega_k)$. Foi considerado que os modelos são mutuamente exclusivos, o que significa que somente um modelo pode ser associado a cada padrão.

De acordo com a teoria de Bayes (Mitchell, 1998), dadas as medidas $x_i, i = 1, \dots, R$, o padrão, Z , pode ser associado à classe ω_k , i.e., seu rótulo θ deve assumir o valor $\theta = \omega_k$.

considerada uma probabilidade *a posteriori* que a interpretação é máxima, associando-se que $\theta \rightarrow \omega_j$ se $P(\theta = \omega_j | x_1, \dots, x_R) = \max_k P(\theta = \omega_k | x_1, \dots, x_R)$.

Dessa maneira a probabilidade *a posteriori* $P(\theta = \omega_k | x_1, \dots, x_R)$ pode ser rescrita, utilizando-se o teorema de Bayes, da seguinte maneira :

$$P(\theta = \omega_k | x_1, \dots, x_R) = (p(x_1, \dots, x_R | \theta = \omega_k)P(\omega_k)) / (p(x_1, \dots, x_R))$$

onde $p(x_1, \dots, x_R | \theta = \omega_k)$ e $p(x_1, \dots, x_R)$ são densidades probabilidades de conjunto de medidas condicionais e incondicionais. O último termo pode ser expresso em termos de distribuições de medidas condicionais como $p(x_1, \dots, x_R) = \sum_{j=1}^m p(x_1, \dots, x_R | \theta = \omega_j)P(\omega_j)$.

Para entender o termo $p(x_1, \dots, x_R | \theta = \omega_k)$ é preciso esclarecer, antes de tudo, que $P(\theta = \omega_k | x)$ é a probabilidade *a posteriori* da k-ésima classe calculada para cada um dos R classificadores, enquanto $P(\theta = \omega_k | y)$ é a probabilidade da k-ésima classe baseada nas características compartilhadas.

A combinação de regras expressas em termos de probabilidade *a posteriori* de uma classe, é retornada por classificadores individuais, os quais podem usar representações mistas e a probabilidade *a posteriori*, baseada em características compartilhadas.

O esquema de detecção de microcalcificações de Kittler *et al.*, baseado nessa teoria, usou combinações de quatro diferentes classificadores: funções de bases radiais (RBF), *Multi-layer-perceptron* (MLP), vizinhos mais próximos (K-NN) e classificador gaussiano. Eles utilizaram um conjunto de 277 mamogramas, incluindo 22 imagens com microcalcificações e 205 mamogramas normais (sem achados mamográficos). Num processamento prévio, foram extraídas 39 características que serviram como base para os classificadores. Os conjuntos de características usados pelos classificadores não são idênticos, mas usam quatro características compartilhadas.

Os classificadores foram treinados com 320 microcalcificações, as quais foram extraídas de três mamogramas anormais, em 960 regiões com manchas extraídas de cinco mamogramas normais. As oito imagens usadas para o treinamento foram excluídas da base de dados e as imagens restantes foram divididas em dois conjuntos, A e B. O conjunto A foi usado para identificar probabilidades *a priori* que garantiriam a detecção de todas microcalcificações do conjunto. As medidas de desempenho são ERRO1, que mede a taxa de *cluster* falso-positivo por imagem, e ERRO2, que dá a porcentagem de imagens normais

classificadas erroneamente como imagens anormais. Os resultados são obtidos no conjunto B e então a regra dos dois conjuntos é permutada. Duas estratégias de combinações foram desenvolvidas e comparadas experimentalmente em dados mamográficos para mostrar suas eficiência. Ambas combinações melhoram a performance do classificador.

Em 1997, Kahn *et al.* apresentaram uma Rede Bayesiana para auxiliar no diagnóstico do câncer de mama. A motivação de desenvolver esse trabalho deu-se à complexidade da detecção precoce do câncer de mama, que está entre as doenças que mais matam mulheres no mundo todo; com a detecção precoce, as chances de cura e a sobrevida da paciente aumentam bastante (Kahn, 1997).

Figura 4.5 – Rede Bayesiana para auxílio ao diagnóstico do câncer de mama (Kahn, 1997).

A operação primária executada com uma Rede Bayesiana é o cálculo de probabilidades *a posteriori*.

As Redes Bayesianas podem expressar relações entre diagnósticos, achados físicos, resultados de testes laboratoriais e achados em imagens. Médicos podem determinar as probabilidades *a priori* e então incorporar resultados de imagens e laboratoriais para calcular as probabilidades *a posteriori* (Andreassen, 1991).

No trabalho de Kahn et. al. foi implementado um modelo de Rede Bayesiana para diagnóstico de câncer de mama, mostrado na Figura 4.5, chamada MammoNet, o qual incorpora cinco atributos do histórico da paciente, dois achados físicos e quinze atributos de achados mamográficos. O modelo assume todas as evidências pertencentes a um caso particular identificado pelo mamograma. O MammoNet infere a probabilidade *a posteriori* de câncer de mama de um determinado caso, baseado na avaliação das evidências.

Os nós da rede e seus estados são enumerados em uma tabela. O MammoNet usa a terminologia BI-RADS, Muitos achados mamográficos, tais como margem de massa, são completamente dependentes da presença de uma massa; esses nós atingem um estado “não aplicável” se o nó precedente tem um valor “nulo”. “Tamanho de Calcificação” tem muitos estados para indicar uma faixa de valores de uma calcificação observada; muitas dessas faixas de sobrepõem. Na prática, o médico seleciona uma faixa que melhor descreve os tamanhos das microcalcificações vistas.

Dos atributos referentes ao histórico da paciente, quatro influenciam na presença do câncer de mama, que influencia na presença de achados físicos e mamográficos. Os achados mamográficos são divididos em manifestações diretas de malignidade, tais como massa e calcificação, sinais indiretos (deformação da mama). Um dos fatores do histórico da paciente, que é uma biópsia prévia, serve como causa complementar dos achados mamográficos e deformação da mama.

As massas e calcificações detectadas mamograficamente são modeladas como manifestações de independência condicional. Os nós “Massa” e “Calcificação” têm três estados: “maligno”, “benigno” e “nulo”. Se não há evidência de massa, por exemplo, o nó “presença de massa” é configurado com “não”, o que força o nó “massa” para o estado “nulo” e nós tais como “margem de massa” para o estado “não aplicável (NA)”. O nó “presença de massa” permite expressar incerteza, considerando-se a presença de massa, independentemente das características descritivas. As características mamográficas de massa - através da independência condicional do câncer de mama, dada a massa - afetam o diagnóstico, por suas influencias nos estados “maligno” e “benigno” do nó “Massa”. O modelo trata as calcificações de maneira similar às massas.

Os autores usaram o sistema BNG – *Bayesian Network Generation* – para gerar uma rede Bayesiana específica para o problema do diagnóstico do câncer de mama, para cada caso de uma base de conhecimentos de regras representada por relações pai-filho de uma Rede

Bayesiana. Dada uma base de conhecimento na forma de regras probabilísticas, um conjunto de evidências E e uma questão (*query*) Q , o sistema BNG gera a rede de mínima estrutura para computar a probabilidade condicional de Q , dada a evidência E , que é $P(Q|E)$. Uma rede é de estrutura mínima, se contém somente nós relevantes para a questão, onde a relevância é determinada pela topologia da rede. A rede gerada é passada para o sistema de domínio público IDEAL para desenvolver a computação probabilística. A inferência (ou seja, o cálculo das probabilidades *a posteriori*) foi desenvolvida numa SPARC station 20 (Sun Microsystems Computer Co., Mountain View, CA). O BNG somente gera aqueles nós necessários para um problema de inferência; então, o tempo de computação para grandes modelos pode ser reduzido significativamente.

Dada uma questão Q e um conjunto de nós evidência, E , o BNG constrói uma rede para computar $P(Q|E)$, de tal forma que a probabilidade computada com a rede seja igual para aquela definida pela base de conhecimentos semântica.

Como resultado, o MammoNet foi avaliado com curvas ROC e teve $A_z = 0,881$. O sistema acertou 23 dos 25 casos positivos de câncer de mama; a sensibilidade foi de 92%, num intervalo de confiança de 95%.

Esses resultados foram comparados com aqueles obtidos com discriminantes lineares e RNA. A performance das redes Bayesianas apresentou ganhos.

Uma das vantagens das redes Bayesianas é que elas podem prover de explicações de suas decisões, pois possuem uma capacidade flexível de especificar dependência e independência de variáveis, num caminho natural através da topologia da rede. Sua estrutura tende a refletir a estrutura lógica inerente à uma tarefa de decisão.

No trabalho de Wang et. al. em 1999, eles também investigaram uma rede Bayesiana para diagnóstico do câncer de mama (Wang, 1999a). Segundo eles, as Redes Bayesianas têm sido atrativas em sistemas de diagnóstico médico, porque podem ser aplicadas para fazer inferências em casos onde a entrada de dados é incompleta. E essa é a situação em muitos conjuntos clínicos, onde a decisões dos diagnósticos tem que ser feita com dados limitados.

Como parte de esforços para aplicar técnicas de sistemas inteligentes em diagnóstico de câncer, eles começaram a investigar as questões que têm surgido para desenvolver uma rede Bayesiana para o mecanismo de decisão para o câncer de mama. Eles tem tentado endereçar suas investigações para duas questões principais: associação do impacto de uma decomposição aparentemente natural de uma Rede Bayesiana para o diagnóstico do câncer de mama, e

determinar a possibilidade de uso de aprendizagem de máquinas para configurar redes, pela análise direta de grandes bases de dados.

Especialmente naquele estudo eles aplicaram uma Rede Bayesiana simples para diagnóstico do câncer de mama, baseada em atributos de achados mamográficos, colhidos em exames físicos e outros dados importantes do histórico da paciente. As redes usadas foram automaticamente construídas por aplicações de métodos de aprendizagem de máquinas para um conjunto de dados de treinamento. Com essas redes, eles investigaram uma possível decomposição da tarefa de decisão em sub-tarefas relacionadas para componentes de imagem e componentes que não são de imagens do conjunto de atributos, além da questão sobre qual a melhor combinação de dados para se usar em diagnóstico.

Os casos selecionados para esse estudo são da biblioteca de filmes do *Magee Women Hospital's Breast Care Center* em Pittsburgh, PA, e correspondem a exames mamográficos feitos entre 1987 e 1991. Somente foram usados os casos que davam seguimento à documentação completa. De um total de 419 casos selecionados, 92 eram positivos (malignos), comprovados por biópsia ou cirurgia, enquanto um caso negativo que foi selecionado se apresentava com documentação de pelo menos dois anos de acompanhamento. As características de histórico da paciente foram extraídas dos arquivos das pacientes.

Para obter as características mamográficas, eles usaram sistemas computacionais de formulação escore, dentro dos quais especialistas relataram seus achados e como eles leram os filmes.

Quando aplicado um algoritmo de aprendizagem de máquina de dados guiados para diagnóstico médico, o tamanho do conjunto de características de entrada deve ser limitado pelo tamanho das amostras de treinamento, se o robusto desempenho é para ser guardado. Nesse trabalho, somente um número relativamente pequeno de casos foi considerado. Eles se limitaram a treze atributos para a base de dados, os quais foram usados para treinar e testar a rede. Os atributos selecionados e seus possíveis estados na rede são demonstrados na Tabela 4.1.

Tabela 4.1 – Atributos usados na rede Bayesiana (Wang, 1999a)

Categoria	Descrição do nó	Estados
Diagnóstico	Câncer de mama	Presente, ausente
Histórico Clínico	Hábito de beber e/ou fumante	Sim, não
	Tratamento com hormônios femininos	Sim, não
	Menopausa	Sim, não
	Teve alguma gravidez	Sim, não
	Caso de câncer de mama na família	Sim, não
Achados físicos	Deformação do mamilo	Sim, não
	Enrugamento da pele	Sim, não
	Dor na mama	Sim, não
	Tem um ou mais caroço (nódulo)	Sim, não
Achados mamográficos	Distorção no formato da mama	Presente, ausente
	Massa	Escore de 1 a 3, escore de 4 e 5, ausente
	Clusters de microcalcificações	Escore de 1 a 3, escore de 4 e 5, ausente
	Assimetria	Presente, ausente

Baseados nas condições de dependência e independência das características selecionadas, eles adotaram uma estrutura de conexão um a um para a topologia da rede. A rede foi construída usando um software comercial, Hungin Demo. Conforme a definição de redes de crença Bayesiana (Mitchell, 1997), BBN, que incluem as propriedades de conexão acíclica e *d*-separation, não há “loops” de “feedback” entre os nós. A ausência de ligação entre dois nós indica que as variáveis não são necessariamente independentes estatisticamente; entretanto a existência de dependência não é considerada importante, no processo de decisão modelado.

Para construção da rede foi considerada, primeiramente, uma série de probabilidades *a priori* e condicionais, de acordo com a topologia da rede. Uma prática comum em aplicações com redes Bayesianas é representar essas probabilidades numa tabela de probabilidades condicionais (Jensen, 1997). Basicamente, a rede da Figura 4.6 consiste em três camadas. Na primeira, há cinco características do histórico da paciente, cada uma delas com dois possíveis estados, como mostra a Tabela 4.2. A tabela de probabilidade para essa camada contém dez valores, dos quais somente cinco são necessários para se determinar completamente esta camada. Na segunda camada, existe somente um nó, que é o “Câncer de Mama”, ele tem cinco nós pais ($Y_i, i = 1, \dots, 5$) que são os nós da primeira camada. Nessa camada as probabilidades condicionais, $P(\text{Câncer} \mid Y_1, Y_2, Y_3, Y_4, Y_5)$ devem ser calculadas. Para cada um dos cinco pais têm dois estados, como mostrado na tabela 4.2. A tabela de probabilidades contém 64 valores,

mas como nem todos são independentes, as 32 combinações de probabilidades condicionais seguintes são suficientes para se especificar essa tabela de probabilidades:

$$P_1(\text{Câncer} = \text{sim} \mid Y_1 = \text{sim}, Y_2 = \text{sim}, Y_3 = \text{sim}, Y_4 = \text{sim}, Y_5 = \text{sim}),$$

$$P_2(\text{Câncer} = \text{sim} \mid Y_1 = \text{sim}, Y_2 = \text{sim}, Y_3 = \text{sim}, Y_4 = \text{sim}, Y_5 = \text{não}),$$

$$P_3(\text{Câncer} = \text{sim} \mid Y_1 = \text{sim}, Y_2 = \text{sim}, Y_3 = \text{sim}, Y_4 = \text{não}, Y_5 = \text{não}),$$

$$P_4(\text{Câncer} = \text{sim} \mid Y_1 = \text{sim}, Y_2 = \text{sim}, Y_3 = \text{não}, Y_4 = \text{não}, Y_5 = \text{não}),$$

$$P_5(\text{Câncer} = \text{sim} \mid Y_1 = \text{sim}, Y_2 = \text{não}, Y_3 = \text{não}, Y_4 = \text{não}, Y_5 = \text{não}),$$

...

$$P_{32}(\text{Câncer} = \text{sim} \mid Y_1 = \text{não}, Y_2 = \text{não}, Y_3 = \text{não}, Y_4 = \text{não}, Y_5 = \text{não}).$$

O nó “Câncer de Mama” também tem oito nós filhos, ($X_i, i = 1, \dots, 8$), representados na terceira camada; preliminarmente, esses nós são supostamente independentes uns dos outros, seis nós têm dois estados e os outros dois restantes têm três possíveis estados, como apresentado na tabela 4.2. A tabela de probabilidades condicionais para essa camada contém 32 valores, dos quais 20 são independentes e suficientes para especificar completamente a camada. Para especificar completamente a rede Bayesiana da Figura 4.6, 110 valores de probabilidades condicionais devem ser determinados, mas somente 57 dos mesmos são independentes. Nesse trabalho, todas as probabilidades necessárias foram calculadas automaticamente dos casos selecionados para o treinamento da rede. No evento em que uma probabilidade condicional necessitada na segunda camada não podia ser calculada da base de dados, pelo fato da limitação de tamanho, os valores de probabilidade *default* $P(\text{Câncer} = \text{sim} \mid Y_1, \dots, Y_5) = 0,5$ e $P(\text{Câncer} = \text{não} \mid Y_1, \dots, Y_5) = 0,5$ foram usados.

Figura 4.6 – Rede Bayesiana (Wang, 1999a)

A base de dados original de 419 casos foi dividida aleatoriamente em cinco partições mutuamente exclusivas, cada uma com 66 casos negativos e 19 casos positivos, exceto uma partição que continha 63 casos negativos e 16 positivos. Uma série de cinco ciclos experimentais foi desenvolvida. Em cada ciclo, um grupo diferente das quatro partições foi usado para cada treinamento da rede (derivando todos 110 valores de probabilidades), e o grupo de casos positivos e negativos da partição restante foi usado para teste. Nos cinco ciclos experimentais, cada partição foi usada para treinamento em quatro ciclos e para teste em um ciclo. As curvas ROC foram construídas para as combinações dos resultados dos testes dos cinco ciclos, e as áreas sob as curvas foram calculadas usando o programa ROCFITT (Metz, 1985).

Os autores também investigaram as contribuições relativas das características das imagens e das características que não são extraídas das imagens no processo de decisão. Essa comparação envolveu mudanças na performance, quando aplicados diferentes métodos para integrar todas características dentro de uma única decisão resultante. Dois métodos de integração de características foram comparados. Primeiro, eles avaliaram a performance de uma rede na qual foram incorporadas características mamográficas e não-mamográficas. No segundo método, eles produziram um sistema híbrido de decisão, que separa redes Bayesianas por tipos de características, uma baseada em características mamográficas e outra em características não mamográficas, e a saída das duas redes foi combinada. Ambas, uma média simples das saídas e uma técnica baseada em regressão logística para combinação das saídas das redes separadas, foram testadas.

Eles dividiram a rede Bayesiana original (mostrada na Figura 4.6) em duas sub-redes. A primeira sub-rede usou somente atributos não mamográficos e o nó “Câncer de mama” excluía quatro atributos mamográficos da rede. Ao contrário, a segunda sub-rede continha somente os atributos mamográficos e o nó “Câncer de Mama”. O método de treinamento *cross-validation* foi usado para treinamento e teste das duas sub-redes. O classificador híbrido, que combinou resultados das duas sub-redes, foi também testado. As áreas sob as curvas ROC do classificador híbrido foram comparadas, para que a rede Bayesiana utilizasse todas características.

As áreas sob as curvas ROC foram calculadas para os resultados de uma rede Bayesiana, incorporando todos os 14 nós e para as duas sub-redes, incorporando somente os

nós dos atributos mamográficos e somente os nós dos atributos não-mamográficos. Os valores de A_Z para as curvas das três redes foram, respectivamente, 0,89, 0,81 e 0,71.

Com esse trabalho os autores concluíram que os atributos de maior contribuição para o processo de decisão são os atributos mamográficos, como mostram os valores de A_Z , quando comparados.

Um outro trabalho de Wang et. al., também divulgado em 1999, investiga maneiras de melhorar a performance de uma rede Bayesiana (bastante semelhante com a descrita no trabalho anterior), para sistemas de diagnóstico do câncer de mama, através da otimização do conjunto de atributos de entrada da rede utilizando Algoritmo Genético (Wang, 1999b).

O Algoritmo Genético foi utilizado para selecionar um sub-conjunto otimizado de atributos, partindo de 21 características inicialmente consideradas como base para o classificador BBN.

A base de dados utilizada nesse trabalho contou com 421 casos, sendo 92 deles positivos (tinham comprovação histológica de malignidade). De cada caso foram extraídas características mamográficas e não-mamográficas; todas características somaram 21.

Para uma avaliação da BBN global (empregando-se os 21 atributos), usaram o método *leave-one-out* para treinamento e teste; em seguida foram calculadas as áreas sob as curvas ROC dos resultados gerados.

Depois de 93 generalizações pesquisadas, o Algoritmo Genético encontrou um conjunto ótimo de características com quatro atributos mamográficos e quatro atributos não-mamográficos.

Numa BBN de três camadas, onde a primeira era composta de nós dos atributos do histórico da paciente, a segunda era composta por um único nó “Câncer de mama” e a terceira camadas continha os nós dos atributos mamográficos. Os nós da primeira camada tinham dois possíveis estados (sim ou não) e os nós da terceira camada tinham cinco possíveis estados, onde um escore 1 indicava “mais provavelmente benigno” e o grau de suspeita vai aumentando até o escore 5 que indica malignidade (esse escore foi criado de acordo com a padronização BI-RADS).

Os casos foram usados para calcular as distribuições de densidades (probabilidades *a priori*), bem como as probabilidades de câncer, através das suposições de independência condicional.

Em cada generalização, o médio e o melhor do desempenho da BBN foram calculados para os cromossomos que apareciam na generalização. Quando o Algoritmo Genético pareceu ter convergido para um subconjunto ótimo, a BBN correspondente para este tal conjunto foi avaliada.

Os valores de A_z tanto para a BBN com as 21 características como para a BBN, com as oito características, foram altos, 0,851 e 0,927 respectivamente. Com esses resultados, os autores concluíram que o desempenho da BBN depende do conjunto de atributos que servirão de base para o classificador.

O trabalho de Burnside et. al. é mais uma contribuição em Redes Bayesiana. Eles desenvolveram uma BBN para integrar achados mamográficos baseados na padronização BI-RADS (Burnside, 2000).

Eles construíram uma rede e compararam com uma outra previamente desenvolvida por Kahn (Kahn, 1995). Os atributos extraídos foram baseados nas descrições de doenças de mama do BI-RADS. O BI-RADS consiste em 43 descrições organizadas numa hierárquicas.

Para construir a rede, eles usaram o software GeNIE, desenvolvido pelo *Decision System Laboratory of the University of Pittsburg* (<http://www.sis.pitt.edu/~dsl>). Eles iniciaram a construção de uma rede global assumindo todas características descritas no BI-RADS como nós filhos de um nó “doença”, exceto a característica densidade.

As descrições das calcificações foram modeladas como condicionalmente independentes da doença. Casos especiais de achados associados foram modelados também como expressões condicionalmente independentes da doença.

O nó determinístico na rede tem três possíveis estados: “benigno”, “maligno” e “pré-maligno”. A decisão final para a biópsia é baseada no valor desse nó.

O modelo hierárquico do BI-RADS incorporado na rede é de dependência condicional entre as diferentes massas achadas.

A patologia da doença da mama apresentou um desafio na construção da rede. A transformação de células benignas em células atípicas para células malignas desafia suposições mutuamente exclusivas no modelo. Eles obtiveram as probabilidades *a priori*, através da especificação de idade, fatores de risco da doença, censos de dados da doença e uma grande triagem aleatória; derivaram muitos dos conjuntos de probabilidades dos estudos de radiologistas e patologistas da correlação individual das doenças de mama.

Para os testes, eles usaram como base de dados 105 casos retirados de um atlas mamográfico. Os testes serviram para avaliar a performance da rede, através de curvas ROC, cujos cálculos das áreas sob as curvas foram executados usando o programa ROCKIT da Universidade de Chicago (<http://www.radiology.uchicago.edu/krl/toppage11.htm>).

O resultado foi superior ao do modelo de Kahn, $A_z = 0,95$ contra $A_z = 0,881$. Os autores atribuíram tal ganho ao diferencial de saída da rede cujo nó doença contava com três possíveis estados e não apenas dois, como o do modelo de Kahn.

Um trabalho desenvolvido em 2001 por Kononenko mostra uma aplicação de aprendizagem de máquinas no diagnóstico médico (Kononenko, 2001). Naquele trabalho, foi feito um estudo comparativo de alguns métodos de aprendizagem de máquina incluindo redes neurais artificiais (RNA), classificadores bayesianos simples, algumas variações, extensões e métodos de aprendizado simbólico.

Depois de escolhidos os algoritmos, foi feita uma comparação de seus desempenhos quando aplicados a problemas médicos. Kononenko e seus colegas utilizaram oito diferentes bases de dados médicos para testes dos algoritmos e construíram uma tabela de comparação como mostrada a seguir.

Tabela 4.2 - avaliação dos classificadores testados para problemas médicos

Classificador	Performance	Transparência	Explicação	Redução no. De testes	Dados perdidos
Assistant-R	Bom	Ótimo	Bom	Bom	Aceitável
Assistant-I	Bom	Ótimo	Bom	Bom	Aceitável
LFC	Bom	Bom	Bom	Bom	Aceitável
Bayes Simples	Ótimo	Bom	Ótimo	Não	Ótimo
Bayes Semi-Simples	Ótimo	Bom	Ótimo	Não	Ótimo
Backpropagation	Ótimo	Pobre	Pobre	Não	Aceitável
K-NN	Ótimo	Pobre	Aceitável	Não	Aceitável

Ao avaliar a performance dos classificadores, todos eles foram satisfatórios e similares, apresentando bons desempenhos, embora tenham-se destacado os classificadores bayesianos. Em conjuntos de dados médicos, em geral, dada uma certa classe de atributos, esses são independentes condicionalmente, dada a classe. Os médicos tentam definir atributos condicionalmente independentes. Os humanos tendem a pensar linearmente e atributos independentes tornam o processo do diagnóstico mais fácil. Por isso, não é surpreendente que os classificadores bayesianos mostrem uma clara vantagem, quando aplicados a conjuntos de dados médicos. Assim, Classificador K-NN é bom nesse domínio.

Na avaliação de transparência do algoritmo, ou seja, a clareza do procedimento que está sendo usado para atingir os resultados e a habilidade de explicação, que é o detalhamento do resultado, há muitas diferenças entre os algoritmos.

Através desse estudo os autores procuraram definir uma técnica de classificação adequada para ser utilizada em conjunto com dados médicos, em busca de auxiliar o diagnóstico médico (Kononenko, 2001).

Muitos outros trabalhos usando BBN estão sendo divulgados, e nossas atenções estarão voltadas para as inovações nessa área durante todo o desenvolver deste projeto. Para uma revisão bibliográfica prévia do assunto, os trabalhos citados aqui nos encoraja a seguir pesquisando e desenvolvendo modelos que sirvam de comparação com outras técnicas classificação.

4.7 Conclusão

A tarefa de classificação ocorre numa vasta gama de atividades humanas. Nessa vastidão, o termo poderia cobrir qualquer contexto em que alguma decisão ou previsão é feita com base na disponibilidade de informações atualizadas. Dessa maneira, o processo de classificação é, então, algum método formal para fazer repetidamente tais julgamentos em novas situações.

Para solução de um problema de classificação, devem-se ter claros alguns passos a ser seguidos. A classificação, na verdade, deve ser encarada como um processo que será aplicado numa seqüência de casos contínuos, devendo cada novo caso ser associado a um conjunto de classes pré definidas, com base em atributos e características observadas. A construção de um procedimento de classificação para um conjunto de dados em classes tem sido bastante investigada e muitas técnicas propostas, cada qual apresentado melhores desempenhos, de acordo com o tipo de dados a ser classificados, a distribuição de amostra e complexidade do problema. No entanto, todas técnicas têm como objetivo auxiliar a resolver problemas, não apenas de classificação propriamente, mas também de separação de classes, reconhecimento de

padrões, otimização na aprendizagem, discriminação de dados e até mesmo em tomadas de decisões.

Diversas técnicas de classificação vêm sendo empregadas, na tentativa de auxiliar o diagnóstico precoce do câncer de mama, incluindo as RNA, classificadores estatísticos, enfim, classificadores paramétricos e não paramétricos. Alguns trabalhos mostram ainda a fusão dos dois tipos de técnicas, nomeando o processo de classificação híbrida.

Em esquemas CAD, deve ser considerada a forma de como apresentar os resultados fornecidos pelo classificador, cuidado esse que não tem sido tomado por muitos pesquisadores, pois, além de um resultado com o menor erro possível, o classificador num esquema CAD tem como objetivo auxiliar na tomada de decisão; por isso, um bom grau de detalhamento das informações produzidas através dos resultados do processo de classificação é necessário.

5

MATERIAIS E MÉTODOS

*"Nós geralmente descobrimos o que fazer percebendo aquilo que não devemos fazer. E provavelmente aquele que nunca cometeu um erro nunca fez uma descoberta."
(Samuel Smiles)*

O presente trabalho foi desenvolvido em várias etapas, e trata de uma particularidade das imagens mamográficas que são as mamas densas, ponto que foi alvo de estudo durante este trabalho e a quantização da densidade tornou-se uma característica que foi considerada nos resultados apresentados pelo esquema de classificação.

Como já discutido anteriormente, esquemas de classificação automatizado dependem de vários outros procedimentos prévios à classificação propriamente dita. Com isso, podemos enumerar as etapas deste trabalho da seguinte forma:

- Investigação de técnicas de extração de atributos para caracterização de lesões, segundo a observação e informações de interesse médico;
- Investigação de técnicas de classificação existentes, bem como o desempenho de cada uma na separação em diferentes classes das lesões detectadas após processamento digital de imagens mamográficas⁵;
- Padronização de um conjunto de imagens⁶;
- Classificação segundo padronização BI-RADS;

⁵ As imagens são processadas com esquemas computadorizados previamente desenvolvidos pelo grupo.

-Desenvolvimento de um esquema de classificação híbrida automática de lesões mamográficas;

- Apresentação dos resultados obtidos dos classificadores, detalhadamente, através de uma interface, com o objetivo de fornecer informações não só quantitativas, mas também qualitativas, visando à boa interpretação de usuários não apenas técnicos, mas de médicos especialistas.

Os principais achados mamográficos considerados são nódulos e calcificações, ou microcalcificações, que são de importante investigação quando se encontram agrupadas. Esses agrupamentos, os *clusters*, despertam a atenção dos especialistas por estarem muitas vezes associados a tumores. Dessa maneira, foram testadas diferentes técnicas para esses dois principais achados, separadamente.

5.1. Base de Dados

Um dos trabalhos que vem sendo desenvolvido no LAPIMO há alguns anos é a elaboração de uma base de imagens mamográficas que já conta com 4800 imagens referentes a aproximadamente 1200 exames. Essas imagens provieram de diferentes instituições, principalmente do Hospital das Clínicas de Ribeirão Preto - FMRP/USP, Hospital São Paulo - UNIFESP, Santa Casa de Misericórdia de São Carlos e Hospital das Clínicas de Botucatu - UNESP. As imagens foram digitalizadas em 2 equipamentos diferentes, Lumiscan50 e Lumiscan75 (*Lumisys*, Inc) com resoluções espaciais de 85 e 150 μm de tamanho de pixel, respectivamente, todas com 12 bits de resolução de contraste.

O processo de aquisição da imagem mamográfica inclui fatores que influenciam na densidade ótica na imagem e no contraste, sendo a densidade ótica diretamente relacionada a alguns fatores físicos, dentre eles o regime da técnica de aquisição, tensão e corrente (kV e mA) utilizados. Essas variáveis devem ser empregadas de acordo com normas de calibração de aparelhos radiográficos e mamográficos, conforme já discutido no capítulo 2. Através de medição e determinação dos parâmetros utilizados para o controle de qualidade do aparelho

⁶ As imagens padronizadas estão sendo organizadas em uma base de dados com informações inclusive de laudos

mamográfico e da imagem, é possível considerar essas medidas no processamento da imagem digital, com a finalidade de amenizar as possíveis interferências na detecção de estruturas de interesse e melhorar a performance do pré-processamento e processamento das imagens, conforme trabalhos prévios já têm demonstrado (Schiabel, 2004; Vieira, 2004; Martinez, 2004).

O fator seguinte é a questão da digitalização. A sensibilidade do digitalizador influencia na escala de níveis de cinza da imagem digitalizada. Então, imagens digitalizadas seguindo os mesmos padrões de resolução de contraste e espacial podem variar, ainda assim, se digitalizadas em diferentes equipamentos, em função da conversão da densidade ótica do filme em intensidade de pixel da imagem não ser linear.

Na base de imagens desenvolvida por nosso grupo, as imagens são digitalizadas em 12 bits de resolução de contraste, o que, teoricamente, levaria a uma escala de 4096 níveis de cinza. Mas a escala real de níveis de cinza varia entre 3600 a 3800 níveis de cinza como característica intrínseca dos digitalizadores (Escarpinati, 2003).

Em paralelo, durante a elaboração do esquema classificador, foram também colhidas informações para complementar a base de imagens, através de pesquisas de buscas de laudos radiográficos, informações de anamnese e eventuais laudos histopatológicos relativos às imagens já pertencentes a base.

Atualmente, está disponível uma interface que possibilita a busca de imagens para que se possam fazer testes dos sistemas desenvolvidos pelo grupo. A Figura 5.1 mostra essa interface.

(a)

(b)

Figura 5.1 - (a) e (b) interfaces do banco de imagens (Benatti, 2003)

Além das imagens, informações vêm sendo coletadas e estão permanentemente sendo inseridas na base. Um exemplo dos dados que estão sendo disponibilizados e foram utilizados para elaboração deste trabalho está apresentado na Tabela 5.1 (a) e (b).

Tabela 5.1 (a) Resumo dos dados coletados relativos as imagens da base.

Imagem (RI)	Localização	Data exame	Câncer na Família	Reposição Hormonal	Micro-calcificação	calcificação
013654_8_las_722r1	S	22/09/1989	Sim	Não	Sim	sim
013965_8_las_533r1	S	24/09/1997	Não	Sim	Não	sim
015754_8_las_358r1	S	15/08/1999	Sim	Não	Sim	sim
015754_8_las_359r1	S	11/11/1988	Não	Não	Sim	sim
017874_08_las_216r1	S	12/12/2001	Não	Sim	Sim	não
019338_8_las_756r1	S	30/10/1999	Não	Sim	Não	não
019338_8_las_758r1	S	22/02/1999	Não	Sim	Não	não

Tabela 5.1 (b) – Continuação das informações coletadas respectivas às imagens da base.

Imagem (RI)	Nódulo palpável	Densidade	Simetria	BI-RADS
013654_8_las_722r1	Sim	Parcialmente liposubstituída	Simétricas	
013965_8_las_533r1	Sim	Densa	Simétricas	IV
015754_8_las_358r1	Não	Densa	Assimétricas	III
015754_8_las_359r1	Sim	Densa	Simétricas	
017874_08_las_216r1	Não	Parcialmente liposubstituída	Simétricas	III
019338_8_las_756r1	Não	Densa	Simétricas	I
019338_8_las_758r1	Sim	Parcialmente liposubstituída	Simétricas	II

As informações apresentadas nas tabelas acima estão atualmente sendo inseridas no banco de imagens, pois os dados são coletados manualmente, devido à disponibilidade dos laudos. Na maioria dos hospitais colaboradores não se encontram, ainda, esses laudos no formato digital, especialmente informações colhidas da anamnese que são encontradas nos prontuários e estão divididas entre as diversas clínicas pela quais a paciente possa ter passado.

5.2 Imagens geradas a partir de pré-processamentos e processamentos de nível médio

Para elaboração do processo de classificação híbrida, as lesões foram previamente detectadas por esquemas já desenvolvidos pelo grupo, específicos para cada um dos tipos de lesões que foram considerados pelo classificador. Foram usados um esquema de detecção de *clusters* de microcalcificações (Nunes, 2001) e outro de detecção de nódulos (Santos, 2002).

A partir desses esquemas, as imagens binarizadas resultantes foram utilizadas para o processo de extração de atributos das lesões, especialmente os descritores geométricos. Uma descrição dos esquemas utilizados está nos itens 5.2.1 e 5.2.2.

5.2.1 Sistema para detecção de *clusters* de microcalcificações

O esquema computacional previamente desenvolvido por Nunes (Nunes, 2001) – representado no diagrama da Figura 5.2 – consiste na detecção de *clusters* de microcalcificações em RI através de técnicas de pré-processamento e processamento de imagens.

Figura 5.2 – Diagrama do sistema de processamento para detecção de microcalcificações (Nunes, 2001).

O primeiro passo é submeter o conjunto de RIs a técnicas de realce. O realce tem por objetivo destacar bordas e detalhes procurados na imagem. A técnica de Realce pela Curva Característica (RCC) realiza uma transformação sobre a imagem digitalizada, a fim de considerar sua influência no contraste da imagem. Ela consiste em tornar linear a região final do histograma que corresponde à região da curva característica, que se refere à área com baixas densidades óticas, e conseqüentemente, às áreas mais claras da imagem. Definidos os parâmetros, o realce é realizado, gerando assim novas imagens.

As imagens obtidas pela RCC são submetidas às técnicas de Realce pelo Coeficiente de Atenuação (RCA) e Realce pela Transformação de Histograma (RTH). A primeira aumenta o contraste entre as possíveis microcalcificações existentes e os demais tecidos que compõem o restante da imagem; a segunda realiza um conjunto de transformações nos níveis de cinza da imagem, com a finalidade de aumentar seu contraste.

Em seguida, as imagens resultantes das técnicas RCA e RTH são segmentadas e agregadas por meio do procedimento de Junção de Imagens Segmentadas (JIS). Esse procedimento soma os sinais resultantes da segmentação das imagens anteriores, gerando uma nova imagem. As imagens resultantes desse conjunto de processamento foram, então,

utilizadas para a extração de atributos geométricos. Exemplos dessas imagens, cujas microcalcificações já estão detectadas, estão na Figura 5.3.

Figura 5.3 Exemplos de RIs binarizadas com microcalcificações detectadas (Nunes, 2001)

5.2.2 Sistema para detecção de nódulos

O esquema computacional de detecção e segmentação de nódulos desenvolvido por Santos (Santos, 2002) consiste na detecção de massas nodulares em RIs, através de técnicas de pré-processamento e processamento de imagens. Primeiramente, foi implementada uma técnica de pré-processamento para equalização de histograma, a fim de aumentar o contraste da imagem. Em seguida, foi aplicada a segmentação, com base na técnica da Transformada *Watershed*, que analisa o histograma dos níveis de cinza da imagem, baseada no princípio de que as estruturas de interesse correspondem a uma mesma frequência desses níveis, sendo que gradiente e bordas correspondem a altas frequências.

O esquema computacional de detecção de nódulos segue os seguintes passos:

- Cálculo do gradiente interno (objeto) e externo (fundo) para toda imagem;
- Cálculo e armazenamento dos pixels mais brilhantes, (pontos de mínimo), que serão os marcadores da imagem.
- Inversão do gradiente e cálculo de distâncias (Hit-and-Miss: elimina o objeto para trabalhar com o fundo; Sequência de *thinning* no fundo; SKIZ:

tickening no 1º plano, para montar o esqueleto através da análise da vizinhança, eliminando pontos e gerando as zonas de influência. Utilizando-se o valor absoluto do gradiente, constrói-se uma espécie de "topografia". Regiões homogêneas compõem as regiões planas da "topografia";

- "Inundação das regiões baixas", a partir dos pontos de mínimo agindo como pontos sementes para o crescimento de regiões, onde a parada é o gradiente da imagem;
- As áreas marcadas com diferentes "níveis de água" formam os segmentos separados pelas linhas da transformada *Watershed* (Figura 5.4 (a)).
- A imagem é então binarizada e o nóculo destacado do restante da imagem, conforme explica a parte (b) da Figura 5.4.

(a)

Figura 5.4 (a)Imagem com as linhas de *Watershed* e junção com a imagem original;
(b)imagem segmentada e seu contorno (Santos, 2003).

Para corrigir os defeitos residuais, na etapa denominada de pós-processamento, utilizou-se a técnica de rotulação, que atribui uma única cor para pixels pertencentes a uma mesma região conectada.

Após a rotulação, a imagem resultante é alinhada à imagem original em níveis de cinza, com o intuito de comparar as duas imagens, destacar o nódulo e chamar a atenção do radiologista para uma área de interesse. O diagrama do esquema completo é apresentado na Figura 5.5.

Figura 5.5 – Diagrama esquemático do sistema para detecção de nódulos (Santos, 2002).

As imagens resultantes do conjunto de procedimentos apresentados no esquema da Figura 5.5 foram então utilizadas para a extração de atributos geométricos. Exemplos dessas imagens, cujas massas já estão detectadas, estão na Figura 5.6.

Figura 5.6 - Exemplos de RIs binarizadas com massas detectadas (Santos, 2003)

5.3 Etapas de extração de atributos

O procedimento de extração de atributos está dividido em módulos distintos, de acordo com o tipo de lesão a ser analisada e com o tipo de descrição da imagem: caracterização de formas e caracterização de variação de intensidade de pixels na área que compreende a lesão.

RIs contendo nódulos foram submetidas a duas etapas distintas de análise: primeiramente, as RIs foram submetidas a um procedimento de extração de atributos em níveis de cinza, para a quantificação da intensidade dos *pixels* nas áreas de interesse; as características extraídas foram analisadas e resultaram numa modificação dos valores de intensidades do recorte original que, na maioria dos casos, proporcionou uma melhora na nitidez da imagem apresentada na interface do programa. Em seguida, as RIs (originais) foram submetidas ao esquema de processamento de imagem para realce do nódulo e segmentação, resultando daí uma imagem binária com a lesão destacada (Santos, 2002). Dessas imagens binarizadas com nódulos detectados foram, então, extraídos atributos descritores geométricos para caracterização de formas.

Recortes contendo *clusters* de microcalcificações foram submetidos também ao esquema de realce, segmentação e detecção de microcalcificações (Nunes, 2001) e das imagens binarizadas resultantes com microcalcificações detectadas foram, então, extraídos os descritores geométricos para a caracterização não apenas de formas dos *clusters*, mas também da distribuição dos sinais na RI.

Considerando que a variação de intensidade das microcalcificações é pouco perceptível, em função da própria composição das partículas que formam essas lesões, os atributos de intensidade não foram extraídos das RIs contendo *clusters*.

Os procedimentos de extração dos atributos foram implementados em *Delphi 7.0* juntamente com uma interface, onde estão disponíveis opções a ser selecionadas, indicando qual tipo de lesão será investigada (*clusters* ou nódulos) e também qual tipo de caracterização da lesão se pretende (atributos de intensidade ou descritores de forma). A partir daí pode-se abrir a RI binária no formato BMP, ou a RI original no formato TIFF de 12 bits. Para manipulação das imagens em 12 bits utilizou-se uma rotina desenvolvida pelo grupo do LAPIMO, que

possibilita que a imagem seja visualizada em 8 bits, mas processada em 12 bits (Escarpinati, 2001). A Figura 5.7 mostra a interface para extração de atributos.

5.3.1 Extração de atributos de intensidade

Na imagem digitalizada, a variação de densidade está relacionada com a intensidade de pixels; logo, a escala de níveis de cinza é um fator relevante na análise das imagens. Diante disso, foi feita uma normalização das imagens antes que fossem submetidas aos procedimentos de extração de atributos de intensidade, para serem classificadas a seguir.

As imagens digitalizadas com 12 bits de resolução de contraste possuem escalas reais de 3600 e 3800 níveis de cinza como citado anteriormente. Para quantizar a intensidade dos pixels de uma região, foi feita uma normalização, que consiste em alterar a escala de cinza das regiões para 4096 níveis, que seria a escala efetiva, teoricamente, de imagens digitalizadas em 12 bits.

Nas imagens digitalizadas é feita uma interpolação em seguida, de modo que a média dos maiores valores de resolução real de contraste (3600 e 3800 níveis) serve de referência para a interpolação. Dessa maneira, o que se pretende é, de fato, trabalhar com todas imagens em uma escala de 4096 níveis, onde o branco é representado pelo valor 4095 e o preto absoluto deveria ser representado pelo valor 0. A partir disso, chega-se ao que nomeamos fator de correção das imagens, dado pela equação 5.1.

$$Fator_de_Correção = \frac{Quantidade_Níveis_12bits}{Média_MAX_Valores_Efetivos} = \frac{4096}{[(3600 + 3800)/2]} = 1,107027 \quad (5.1)$$

A normalização das intensidades de pixels, considerando-se a média dos maiores valores reais dos digitalizadores, produz uma imagem com um erro médio de até 2,58% da escala de níveis de cinza do valor original das imagens, ou seja, para as imagens digitalizadas no Lumiscan 50, os valores de intensidades de pixels corrigidos podem estar em média 2,58% acima do valor real, caso a imagem fosse digitalizada por um equipamento com a relação linear entre densidade ótica e intensidade de pixel. O contrário acontece para as imagens digitalizadas no Lumiscan 75, pois os valores de intensidades de pixels podem estar em média 2,58% abaixo do valor real caso a imagem fosse digitalizada por um equipamento com a relação linear entre densidade ótica e intensidade de pixel. Portanto, o desvio médio das intensidades dos pixels é de 1,98% e o desvio padrão é de 2,28%.

Com a imagem corrigida, ao se multiplicar cada ponto por esse fator de correção, pode-se admitir que as diferenças de escalas efetivas de resolução de contraste não vão influenciar na análise dos atributos extraídos das imagens em níveis de cinza, para a quantização de densidade, pois os valores de intensidade de pixel são relativos à escala, teoricamente, de 12 bits. A Figura 5.8 (a) e (b) mostra os histogramas de uma imagem antes e depois da correção.

Figura 5.8 – a) Histograma de imagem antes de ser normalizada – escala efetiva de níveis de cinza; b) histograma de imagem normalizada pelo fator de correção.

Nas imagens originais provenientes do equipamento com escala real de 3800 níveis de cinza, quando a intensidade do pixel era maior que 3700, o valor corrigido ultrapassou o

máximo da escala de 12 bits. Nesse caso, os valores foram aproximados para o máximo de 4095.

Embora com a correção das intensidades, o preto seja alterado, isso pode ser ignorado, pois o objetivo principal da normalização era a normalização das intensidades máximas afim de quantizar densidade, portanto os pontos de intensidades mínimas são praticamente desprezados. Mesmo assim uma análise da variação de fundo da imagem tem sido uma das preocupações e já está em andamento uma quantização dessas variações.

Com as imagens corrigidas, segue-se então a extração dos atributos de intensidade com base na literatura revisada descrita no item 3.1.3 do capítulo 3.

Para o cálculo da variação de níveis de cinza numa região de interesse, foi implementada uma rotina de varredura pixel a pixel da imagem. Assim, é possível encontrar os pixels de menor e maior intensidade da região. Os pixels de menor intensidade geralmente estão associados a tecidos gordurosos (Wang, 2003), e os pixels de maior intensidade a tecidos mais densos (fibroglandulares) ou a lesões.

Os atributos extraídos dessas imagens foram:

- Média de intensidade do histograma;
- Valor de intensidade do maior pico do histograma;
- Menor intensidade do histograma;
- Maior intensidade do histograma;
- Porcentagem do maior valor de intensidade do histograma em relação à escala de 4096 níveis;
- Diferença da média para o menor valor de intensidade;
- Diferença da média para o maior valor de intensidade;
- Variação de intensidade global da RI, que consiste na média da variação de intensidade de todos os pixels da região;
- Índice de variação de intensidade local (DI - Densidade Integrada), que integra as diferenças das variações locais de intensidades e a variação média, dando uma medida que quantifica contraste entre as partes mais claras da imagem e o restante;
- Quantidade de pixels com intensidades maiores que a intensidade do pico do histograma;

- Porcentagem da área da RI com intensidade maior que a intensidade do pico do histograma;
- Quantidade de níveis de cinza da RI;

Os valores dos atributos foram apresentados em porcentagens, para que a análise dos valores se tornasse mais objetiva. Além disso, é uma maneira de normalização de dados que facilita no processo de classificação, que ocorre posteriormente. A Figura 5.9 mostra parte da interface onde são exibidos os atributos de intensidade.

Figura 5.9 – Interface de extração de atributos de intensidade.

As imagens são então modificadas seguindo alguns critérios, com o objetivo de melhorar a nitidez da RI.

Considerando-se o valor de pico do histograma e a média do histograma, são feitos os seguintes ajustes:

- Para imagens cuja média é maior que o pico, os pixels com valores abaixo do pico são diminuídos do valor do desvio médio da correção da imagem, ou seja, 1,98% do valor de suas intensidades. Os valores entre o pico e a média são acrescidos de 1,98% dos seus valores; e para os pixels maiores que a média deve-se acrescentar o valor de desvio padrão da correção inicial, que é de 2,58%;
- Para imagens cuja média é menor que o pico, os pixels com valores abaixo da média são diminuídos da variação média do histograma, os valores entre a média e o pico são aumentados da variação média do histograma e os valores entre o pico e o máximo são aumentados 2,58% de seus valores, correspondentes ao desvio padrão da correção inicial.

Essas duas regras de modificação das imagens foram estabelecidas experimentalmente. A Figura 5.10 mostra as imagens antes e depois da modificação.

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

Figura 5.10 – RIs antes de serem modificadas pela análise dos atributos de intensidade em (a), (c), (e), (g), (i), (l) e (n). RIs depois da modificação em (b), (d), (f), (h), (j), (m) e (o).

5.3.2 Extração dos descritores geométricos

As técnicas baseadas em momento descrevem as propriedades globais de uma imagem, isto é, a maneira com que as características de pixels brancos e pretos representam o padrão.

A forma é frequentemente representada por uma imagem binária e seus momentos podem ser obtidos através de algoritmos rápidos (Andersen, 1985).

Em geral, funções de várias variáveis podem ser caracterizadas por uma sequência de números. Momentos representam uma técnica descritiva e intuitiva baseada no estudo da Mecânica dos Corpos. Os Momentos Invariantes (MI) de Hu denotam uma quantidade (ou um valor) que permanece inalterada sob certas condições (rotação, translação e escala). Eles

representam um conjunto completo de descritores de imagem. Os momentos de baixa ordem possuem relações intuitivas com os objetos; por exemplo, m_{00} refere-se à massa, m_{10} e m_{01} referem-se ao centro de massa e m_{11} , m_{20} e m_{02} referem-se às direções principais.

Com base na literatura descrita no item 3.1.1 do capítulo 3, foi implementado um procedimento para o cálculo dos momentos das RIs, contendo nódulos detectados e nas RIs com *clusters* detectados, além dos outros atributos de forma - compacidade; irregularidade, perímetro e área.

A partir das imagens binarizadas contendo nódulos foram calculados os momentos cartesianos, os MI de Hu, o Raio de Giração (RG), perímetro, área, compacidade e irregularidade, tendo como objeto-base a própria lesão detectada. Os objetos nas RIs são representados pelos pontos brancos.

Para as imagens contendo *clusters*, foram calculados os mesmos atributos das imagens de nódulos. Portanto, nessas imagens o objeto-base foi todo o *cluster* e não apenas as microcalcificações isoladamente, porque o que se pretende é a caracterização do *cluster* como um todo. Dessa maneira, os *clusters* foram demarcados nas RIs através da varredura na imagem pixel a pixel, com o intuito de delinear o início e o final da área efetiva ocupada por cada um. A Figura 5.11 mostra um *clusters* delineado dentro de uma RI.

Figura 5.11 - *Cluster* delineado dentro de RI binarizada

Os atributos "descritores geométricos" extraídos tanto para caracterização de formas de *clusters* quanto para formas de nódulos são apresentados na tabela 5.2.

Tabela 5.2 – Atributos geométricos extraídos das RIs

Atributos	Objetos
Área (A)	Nódulo, microcalcificação
Perímetro da microcalcificação (P)	Microcalcificação
ÁreaCluster	<i>cluster</i>
PerímetroCluster	<i>cluster</i>
Hu1(ϕ_1)	<i>cluster</i> , nódulo
Hu2(ϕ_2)	<i>cluster</i> , nódulo
Hu3(ϕ_3)	<i>cluster</i> , nódulo
Hu4(ϕ_4)	<i>cluster</i> , nódulo
Hu5(ϕ_5)	<i>cluster</i> , nódulo
Hu6(ϕ_6)	<i>cluster</i> , nódulo
Hu7(ϕ_7)	<i>cluster</i> , nódulo
Momento ₂₀	<i>cluster</i> , nódulo
Momento ₀₂	<i>cluster</i> , nódulo
Momento ₃₀	<i>cluster</i> , nódulo
Momento ₀₃	<i>cluster</i> , nódulo
Momento ₁₁	<i>cluster</i> , nódulo
Momento ₁₂	<i>cluster</i> , nódulo
Momento ₂₁	<i>cluster</i> , nódulo
Irregularidade	Microcalcificação, <i>cluster</i> , nódulo
Compacidade	Microcalcificação, <i>cluster</i> , nódulo
RG	<i>cluster</i> , nódulo

Tanto nos *clusters* quanto nos nódulos, área é representada pelo Momento₀₀, utilizado para calcular os Momentos Invariantes de Hu, Raio de Giração e Irregularidade, cujos equacionamentos foram apresentados no capítulo 3.

A Figura 5.12 (a) e (b) ilustra parte da interface onde são exibidos os atributos de descritores geométricos extraídos para RI contendo nódulos e para RI contendo *cluster*.

(a)

(b)

Figura 5.12 – (a) Interface com imagem de nódulo processada e (b) interface com imagem de *cluster* processada

A mesma imagem contendo nódulo pode ser aberta em sua versão original e então calculados os atributos de intensidade. A Figura 5.13 mostra a interface com o processo completo de extração de atributos para nódulos.

Figura 5.13 – Interface mostrando o processo completo de extração de atributos para nódulos.

5.4 Análise de atributos

Uma análise das informações utilizando estatística descritiva foi feita para a avaliação da dispersão dos padrões.

O gráfico da figura 5.14 mostra a correlação dos atributos entre as classes suspeitos e não suspeitos.

Figura 5.14 Gráfico dos coeficientes de correlação dos atributos das classes suspeitos e não suspeitos de *clusters*.

A tabela 5.3 apresenta as medidas de dispersão entre os *clusters* das classes suspeitos e não suspeitos.

A medida de correlação é importante pois normaliza a covariância num intervalo $[-1,1]$, onde os valores próximos de zero são os mais dispersos.

Tabela 5.3 - Medidas de dispersão entre classes

Medidas	Hu1	Hu2	Hu3	Hu4	RG	Área	ÁreaCluster	Perímetro
Variância	0,582	13,258	150,346	9,351	15,297	833,903	1323226,482	2848,849
Covariância	0,037	-2,647	-34,439	-1,883	-4,283	22,307	-156418,411	-362,516
Correlação	0,165	-0,210	-0,242	-0,214	-0,321	0,034	-0,143	-0,148

Medidas	Compacidade	Irregularidade	Mom02	Mom20	Mom11	Mom12	Mom21
Variância	0,304	0,366	51,205	23,093	71,698	14918,935	4069,207
Covariância	-0,025	0,004	-5,273	-4,201	-5,918	-1117,400	-607,527
Correlação	-0,122	0,015	-0,102	-0,188	-0,144	-0,074	-0,150

Medidas	Mom03	Mom30	Hu5	Hu6	Hu7
Variância	4320,334	20218,761	5538017034455,650	319566639,015	1645000,040
Covariância	-546,292	-870,175	-554682024681,858	-43662546,313	-225352,889
Correlação	-0,159	-0,112	-0,115	-0,154	-0,155

Os valores apresentados na Tabela 5.4 referem-se às medidas de dispersão dos atributos intraclassas.

Tabela 5.4 - Medidas de dispersão dos atributos intra-classes

Medidas	Hu1	Hu2	Hu3	Hu4	RG	Área	AreaCluster
Médias_Suspeitos	0,558	8,279	42,440	14,842	10,245	51,759	2038,483
Médias_N.suspeitos	0,618	9,247	45,545	15,717	10,798	34,581	1826,839
Desvio_padraão_Suspeitos	0,543	4,096	13,559	3,432	4,217	35,384	1275,146
Desvio_padraão_N.suspeitos	0,446	3,224	11,149	2,706	3,720	18,758	1051,896
Variância_N.suspeitos	0,199	10,392	124,289	7,325	13,841	351,852	1106484,406
Variância_Suspeitos	0,295	16,777	183,845	11,779	17,785	1252,047	1625996,616
desvio medio_Suspeitos	0,437	3,323	10,978	2,761	3,200	28,715	1061,567
desvio medio_N.Suspeitos	0,358	2,555	8,653	2,129	2,774	15,088	798,974

Medidas	Perímetro	Irregularidade	Compacidade	Mom02	Mom20	Mom11
Médias_Suspeitos	176,069	5,753	0,947	3,483	3,312	5,317
Médias_N.suspeitos	166,516	5,894	0,967	3,643	3,140	3,206
Desvio_padraão_Suspeitos	57,280	0,691	0,517	8,000	5,516	11,672
Desvio_padraão_N.suspeitos	50,891	0,526	0,599	6,531	4,216	3,744
Variância_Suspeitos	3280,995	0,478	0,267	64,005	30,426	136,234
Variância_N.suspeitos	2589,858	0,277	0,358	42,659	17,774	14,020
desvio medio_Suspeitos	46,273	0,487	0,398	4,430	3,915	7,005
desvio medio_N.Suspeitos	38,144	0,378	0,371	3,342	2,726	2,944

Medidas	Mom12	Mom21	Mom03	Mom30	Hu5
Médias_Suspeitos	45,136	31,337	39,839	72,893	1225235,194
Médias_N.suspeitos	42,268	33,040	29,758	31,712	657317,415
Desvio_padraão_Suspeitos	137,907	64,636	83,937	201,052	2944977,318
Desvio_padraão_N.suspeitos	109,924	65,097	44,879	43,167	1679359,791
Variância_Suspeitos	19018,264	4177,873	7045,499	40421,906	8672891403158,720
Variância_N.suspeitos	12083,383	4237,618	2014,108	1863,384	2820249307689,290
desvio medio_Suspeitos	65,944	41,672	52,774325	107,04142	1874620,458
desvio medio_N.Suspeitos	47,487	34,821	30,057011	33,226909	826012,3

Medidas	Hu6	Hu7
Médias_Suspeitos	11244,607	770,558
Médias_N.suspeitos	8051,479	530,508
Desvio_padrao_Suspeitos	22131,195	1598,993
Desvio_padrao_N.suspeitos	13300,517	935,357
Variância_Suspeitos	489789796,389	2556778,780
Variância_N.suspeitos	176903765,476	874893,205
desvio medio_Suspeitos	15045,0794	1067,7273
desvio medio_N.Suspeitos	7731,1554	560,9503

As células hachuradas na Tabela 5.4 representam a grande dispersão de valores intra classes de mesmo atributo, onde o desvio médio é maior que a média.

Uma análise utilizando estatística descritiva também foi feita para os atributos descritores geométricos extraídos de nódulos, em busca de verificar a dispersão intra classes das categorias BI-RADS associadas às lesões e também entre as classes, que são BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4 e BI-RADS 5. A Figura 5.15 apresenta as correlações por atributos entre as classes, onde a sigla BR é usada para representar o termo BI-RADS.

Figura 5.15 - Correlações entre as classes por atributo.

Tabela 5.5 - Correlações entre as classes por atributos descritores geométricos extraídos de nódulos

Medidas	Hu5	Hu6	Hu7	Área	Área Nódulo	Perímetro	Hu1	Hu2	Hu3
correlação(BR4-BR5)	0,2629	0,2556	0,4478	-0,4237	-0,0704	-0,0507	0,1007	0,2250	0,1130
correlação(BR2-BR5)	-0,0598	-0,0679	-0,5943	0,3386	-0,0830	-0,0886	0,1021	0,3266	-0,0713
correlação(BR2-BR3)	0,1515	0,2230	0,2991	0,2796	0,0949	0,0004	0,5653	0,2543	0,3974
correlação(BR2-BR4)	-0,3375	-0,2753	-0,3966	-0,1687	-0,3828	-0,2837	-0,1449	-0,2384	-0,2240
correlação(BR3-BR4)	0,2894	0,2488	0,3263	-0,1931	0,3071	0,2308	0,0667	0,2633	0,1340
correlação(BR3-BR5)	-0,3541	-0,3025	-0,1870	0,7057	-0,7550	-0,6662	-0,1301	-0,3186	-0,1867
correlação(BR1-BR2)	0,4919	0,5603	0,3859	-0,2792	0,8143	0,8401	0,6650	0,3876	0,7752
correlação(BR1-BR3)	-0,3524	-0,2627	-0,1256	-0,4480	-0,0727	-0,0285	0,0670	-0,3721	0,0934
correlação(BR1-BR4)	-0,4822	-0,5663	-0,6192	-0,4609	-0,1906	0,0013	-0,7116	-0,5379	-0,6245
correlação(BR1-BR5)	0,1162	0,1806	-0,5468	-0,6024	0,3012	0,2033	0,2319	0,1344	0,1815

Medidas	Hu4	RG	Irregularidade	Compacidade	Mom02
correlação(BR4-BR5)	0,1240	0,1332	-0,4776	-0,1426	0,2973
correlação(BR2-BR5)	-0,1563	-0,0721	0,2955	-0,1252	-0,5988
correlação(BR2-BR3)	0,2058	0,0666	-0,2480	-0,1058	-0,1931
correlação(BR2-BR4)	-0,1682	-0,2227	0,0027	-0,1844	-0,1966
correlação(BR3-BR4)	0,1074	0,2587	0,5569	0,1162	0,0280
correlação(BR3-BR5)	-0,0844	-0,5756	-0,5878	-0,5287	0,0725
correlação(BR1-BR2)	0,8072	0,8022	-0,0451	0,8672	0,1931
correlação(BR1-BR3)	0,1772	-0,0058	-0,2881	-0,1187	-0,0558
correlação(BR1-BR4)	-0,3385	-0,2823	-0,8974	0,2687	0,1878
correlação(BR1-BR5)	0,2882	0,3036	0,3151	0,0233	0,6288

Medidas	Mom20	Mom11	Mom12	Mom21	Mom03	Mom30
correlação(BR4-BR5)	0,1116	-0,3192	0,3921	0,3293	-0,0214	-0,3059
correlação(BR2-BR5)	0,0721	0,1932	-0,6620	-0,3332	0,0480	0,3479
correlação(BR2-BR3)	0,4027	0,5688	-0,1861	0,0670	0,3237	0,3422
Correlação(BR2-BR4)	-0,3283	-0,1763	-0,2782	-0,3832	-0,3868	-0,2286
Correlação(BR3-BR4)	0,2369	0,1096	0,0091	0,2348	0,3577	0,3456
Correlação(BR3-BR5)	-0,3071	-0,5609	0,0823	-0,2349	-0,5208	-0,6863
Correlação(BR1-BR2)	0,6120	-0,1987	0,0799	0,6736	0,3631	-0,1887
Correlação(BR1-BR3)	-0,2552	-0,2987	-0,1352	-0,2814	-0,3247	-0,3049
Correlação(BR1-BR4)	-0,6799	-0,8052	0,2164	-0,3501	-0,6538	-0,6884
Correlação(BR1-BR5)	0,2501	0,2199	0,6547	0,3054	0,1510	0,1354

As células coloridas na Tabela 5.5, de acordo com as linhas do gráfico da Figura 5.15, mostram os atributos com correlações próximas de zero entre as classes.

Através da tabela acima, é possível observar que a medida de correlação entre as classes por atributo é dispersa, não sendo possível definir um mesmo conjunto de atributos com correlação próxima de zero para todas as classes. Por isso, a necessidade do uso de todos os atributos.

As Figuras 5.16, 5.17 e 5.18 mostram, respectivamente, as correlações por atributos entre classes que na prática médica podem causar a maioria das confusões, por terem diferenças bastante tênues entre as lesões e que são importantes na definição da categoria BI-RADS do caso.

Figura 5.16 Correlações entre as categorias BI-RADS 2 e BI-RADS 3.

Figura 5.17 - Correlações entre as categorias BI-RADS 3 e BI-RADS 4.

Figura 5.18 - Correlações entre as categorias BI-RADS 4 e BI-RADS 5.

A Tabela 5.6 mostra medidas de dispersão dos atributos intraclassas.

Tabela 5.6 - Medidas de dispersão intraclassas.

Medidas	Hu1	Hu2	Hu3	Hu4	RG	Irregularidade	Compacidade
Variância (BR1)	0,3528	4,8171	104,1797	7,2204	0,2316	0,9397	1,0699
Variância (BR2)	0,0856	0,0020	140,3637	23,6431	0,3908	0,0387	0,0069
variância (BR3)	1,0270	21,6884	395,7827	25,7455	9,5147	0,0077	0,0172
Variância (BR4)	0,1717	0,4203	57,3334	3,5581	0,3595	0,0226	0,1059
Variância (BR5)	0,0286	0,0017	17,9269	1,3563	0,3150	0,0156	0,0025
Media(BR1)	-2,4945	-5,2122	36,6591	13,4067	3,2293	5,7461	0,5643
Media(BR2)	-2,7357	-5,9443	30,3833	11,1685	3,0468	6,0454	0,2393
Media(BR3)	-2,6929	-5,9332	32,8009	12,3910	3,2564	6,1960	0,2222
Media(BR4)	-2,7184	-5,7443	31,1350	12,0547	2,6137	6,1197	0,3845
Media(BR5)	-2,7028	-5,9495	32,9777	12,4850	3,2488	6,1368	0,2071
Desviopadrão(BR1)	0,6261	2,3135	10,7590	2,8324	0,5073	1,0218	1,0903
Desviopadrão(BR2)	0,1612	0,0541	4,9982	1,2830	1,0703	0,2842	0,1788
Desviopadrão(BR3)	1,0682	4,9090	20,9704	5,3485	3,2515	0,0926	0,1382
Desviopadrão(BR4)	0,4395	0,6876	8,0312	2,0007	0,6360	0,1595	0,3451
Desviopadrão(BR5)	0,1825	0,0439	4,5733	1,2579	0,6063	0,1349	0,0536

Medidas	Mom02	Mom20	Mom11	Mom12	Mom21	Mom603	Mom30
Variância (BR1)	0,0000	0,0001	0,0024	1,2169	13,4105	717,1338	32571,4686
Variância (BR2)	0,0000	0,0000	0,0000	1,8972	1,3551	1,9261	5,9104
variância (BR3)	0,6562	0,3116	0,1479	78051608,3263	37070362,2967	17605800,5768	8361012,0123
Variância (BR4)	0,0000	0,0000	0,0000	16,5541	19,5237	23,3980	27,9142
Variância (BR5)	0,0000	0,0000	0,0000	2,3097	1,4272	1,3508	1,5358
Media(BR1)	0,0013	0,0035	0,0177	2,1436	3,3624	11,3827	63,1552
Media(BR2)	0,0009	0,0010	0,0013	1,8352	1,9797	2,3883	3,2038
Media(BR3)	0,0012	0,0011	0,0012	2,7653	2,6527	2,6013	2,6131
Media(BR4)	0,0018	0,0018	0,0018	3,2414	3,0654	3,0416	3,1246
Media(BR5)	0,0011	0,0010	0,0010	2,7312	2,3724	2,1800	2,1052
Desviopadrão(BR1)	0,0010	0,0076	0,0521	1,1628	3,8601	28,2279	190,2381
Desviopadrão(BR2)	0,0000	0,0003	0,0009	0,2291	0,7349	1,7107	3,5370
Desviopadrão(BR3)	0,8539	0,5884	0,4054	9312,5726	6417,8884	4422,8951	3047,9523
Desviopadrão(BR4)	0,0028	0,0031	0,0034	4,3155	4,6866	5,1306	5,6039
Desviopadrão(BR5)	0,0005	0,0004	0,0004	1,6415	1,2904	1,2554	1,3386

Medidas	Hu5	Hu6	Hu7
Variância (BR1)	2561563109678,1900	5767387,4454	1160628,6431
Variância (BR2)	82350518,8645	2297,7116	375,9394
variância (BR3)	352378863886527000,0000	17895495590,3601	13500182479,0836
Variância (BR4)	2303782530,7712	39761,3376	2533,4178
Variância (BR5)	68910434,4692	1596,2299	546,7131
Media(BR1)	547760,3570	895,5968	398,1616
Media(BR2)	13818,9491	91,8342	30,5473
Media(BR3)	16172,2376	101,0485	40,9115
Media(BR4)	24779,8416	131,2880	40,6741
Media(BR5)	13270,2339	91,0107	31,5571
Desviopadrão(BR1)	1687062,9013	2531,4439	1135,6000
Desviopadrão(BR2)	11730,9321	55,9498	7,2338
Desviopadrão(BR3)	625725236,0142	141010,2265	122475,3149
Desviopadrão(BR4)	50909,2855	211,4982	53,3863
Desviopadrão(BR5)	8966,3542	43,1540	25,2553

Medidas	Area	Area Nódulo	Perimetro
Variância (BR1)	2135634,6100	81074560,8000	32355,2400
Variância (BR2)	10837634,0165	80443818,2645	17784,9917
variância (BR3)	4508082,2100	190660639,2500	42507,3600
Variância (BR4)	3491250,9877	98022633,3580	33779,3580
Variância (BR5)	295634,7755	88036596,6939	14611,2653
Media(BR1)	4328,3000	20939,0000	558,6000
Media(BR2)	5402,2727	22537,9091	594,9091
Media(BR3)	4636,7000	25964,5000	613,2000
Media(BR4)	4394,1111	15626,4444	471,5556
Media(BR5)	5167,7143	27295,8571	654,8571
Desviopadrão(BR1)	1540,4309	9491,1983	189,6056
Desviopadrão(BR2)	1274,2064	12201,8346	243,2447
Desviopadrão(BR3)	2238,0751	14554,9014	217,3256
Desviopadrão(BR4)	1981,8318	10501,2124	194,9404
Desviopadrão(BR5)	587,2880	10134,5628	130,5622

As células em negrito da tabela acima mostram as situações onde a média do atributo é menor que o desvio padrão, o que se dá pela grande variação dos valores de um atributo dentro de uma mesma classe.

Com a análise estatística descritiva dos atributos, é possível então representar a grande semelhança entre os objetos de classes diferentes. Isso se dá principalmente porque os detalhes que diferenciam as lesões são bastante tênues e de difícil identificação visual.

5.5 Etapas dos processos de classificação

A classificação projetada aqui foi dividida em etapas de acordo com o tipo de lesão e as características extraídas.

A primeira etapa consistiu da classificação de *clusters* considerando os descritores geométricos como dados de entrada para o classificador; a segunda etapa foi a classificação dos nódulos a partir dos descritores geométricos e tendo como resposta a classificação da lesão em 5 classes seguindo a padronização de laudos BI-RADS. E a terceira etapa foi uma

separação das lesões de acordo com grau de intensidade de pixels, o que representa uma quantização de densidade da mama na área da lesão.

5.5.1 Experimentos com Redes Bayesianas para Classificação de *Clusters* de Microcalcificações

Foram feitas investigações preliminares do funcionamento e comportamento de Redes Bayesianas, através da utilização de um *software* de simulação de redes Bayesianas.

A partir das RIs investigadas foram testados alguns classificadores. Seguindo estudos prévios da literatura foram escolhidos *softwares* de domínio público encontrados na Internet. Depois de vários testes escolheu-se os que mais se adequaram ao problema da classificação de *clusters* de microcalcificações e à base de dados.

Para os testes utilizou-se os atributos descritores geométricos extraídos para as RIs contendo *clusters*. Os atributos foram divididos em grupos por ordem de importância, verificada a partir de análises estatísticas e PCA.

Os resultados apresentados a seguir foram conseguidos através do *software Bayesware Discoverer - Trial Edition* (<http://www.bayesware.com>). Os autores do *software* tratam redes bayesianas como um formalismo da representação do conhecimento. Como já citado anteriormente, uma rede bayesiana é um grafo onde os nós representam variáveis estocásticas, pois a transmissão de um valor para uma variável é representada por uma distribuição de probabilidade, e os arcos representam as dependências entre as variáveis. Uma variável pode assumir um conjunto de valores finitos (<http://bayesware.com/resources/support.html>).

Há dois componentes principais em uma rede bayesiana:

- Componente qualitativo: uma estrutura de grafo representa o modelo de dependência entre as variáveis no domínio;
- Componente quantitativo: um conjunto de distribuições de probabilidades condicionais que modela os efeitos das variáveis pais em cada variável filha.

O primeiro componente codifica a associação de independência condicional entre as variáveis no domínio. O caminho tradicional de aprendizado da estrutura de dependência é

desenvolver uma série de testes de independências condicionais e a estrutura de rede resultante retorna a representação das independências condicionais significantes descobertas nos dados. No entanto o procedimento de aprendizagem do *Bayesware* funciona da seguinte maneira: considerar que todas as possíveis alternativas de modelos que podem ser gerados para representar os dados constituem um conjunto de exposições dos dados mutuamente exclusivos, e adicionar uma distribuição de probabilidades para eles. Como não são conhecidas, a princípio, as exposições, assume-se que a distribuição de probabilidades é uniforme, para representar o fato que todos os modelos são igualmente parecidos. Então estas probabilidades *a priori* uniformes serão atualizadas com base em observações dos dados para obter a probabilidade *a posteriori* do modelo de dados. Esta tarefa é desenvolvida através de dois elementos:

- métrica escore (“Scoring metric”): uma medida para associar o *fitness* de um modelo para os dados, ou seja de quanto provável é um modelo para ser o responsável para representar os dados, que neste caso é a probabilidade *a posteriori* de um modelo dado os dados. Esta probabilidade é mais facilmente calculada conhecendo um outro valor de verossimilhança (*marginal likelihood*). Este valor pode ser muito pequeno e alterar a precisão numérica do computador digital, então é usado *log* da verossimilhança (*marginal likelihood*);
- procedimento de busca: um procedimento de busca explorando o conjunto de possíveis modelos. O modelo apropriado é o de máxima verossimilhança (*marginal log likelihood*).

5.5.2 Classificação de *clusters* de microcalcificações e nódulos por Rede Neurais *Backpropagation*

A partir dos atributos extraídos das imagens contendo *clusters* de microcalcificações, as estruturas foram classificadas em *clusters* “suspeitos” e “não suspeitos” através de uma rede neural *Backpropagation*. A arquitetura da rede foi definida com duas camadas escondidas, uma camada de entrada e uma de saída. A primeira camada recebendo um atributo extraído da imagem, as camadas intermediárias variando entre 10 e 50 neurônios e 2 neurônios na camada

de saída. O número de neurônios na camada de entrada foi definido pelo número de características usadas na representação de cada padrão. Nas camadas intermediárias o número de neurônios pode ser variado em busca de encontrar a melhor configuração, de modo que foram feitos vários experimentos com diversas configurações nestas camadas

Devido à necessidade de uma interface amigável, o algoritmo foi desenvolvido em *Delphi 7.0*, que é uma linguagem que oferece muitos recursos visuais.

A configuração das camadas seguiu o esquema de conexão inter-camadas, que ocorre entre neurônios de diferentes camadas. A propagação do sinal inter-camadas foi feita de maneira *feedforward*, que permite que a informação seja distribuída entre os neurônios em apenas uma direção.

A RNA implementada é de aprendizado supervisionado. A função de ativação usada foi a sigmóide. O algoritmo da rede *Backpropagation* foi detalhado no capítulo 4.

A interface que desenvolvemos possibilitou a escolha dos números de neurônios e a configuração dos parâmetros de treinamento, como a taxa de aprendizagem, termo “*momentum*” e definição do erro mínimo. A Figura 5.19 (a) e (b) mostra a interface usada para o treinamento da rede.

(a)

Figura 5.19 (a) Interface para a etapa de treinamento da Rede *Backpropagation*, configurações; (b) Interface no final do treinamento.

A rede usada para a classificação de *cluster* foi adaptada e utilizada para a classificação de nódulos a partir dos descritores geométricos.

Com a interface, foi possível uma adaptação das configurações de arquitetura da rede, pois a camada de saída para nódulos contém 5 neurônios, seguindo a base canônica para representar as 5 possíveis classes da lesão de acordo com a padronização de laudos mamográficos BI-RADS.

Da mesma maneira como para os *clusters*, a rede foi treinada com variações de arquitetura quanto ao número de camadas escondidas e ao número de neurônios nestas camadas, também variando-se os parâmetros de treinamento.

Para a inserção dos dados na entrada da rede, foi feita uma normalização dos valores para que ficassem numa escala de 0 a 1. Isso foi feito dividindo-se cada valor pelo maior valor extraído para aquela categoria à qual ele pertence - considerado o valor absoluto da divisão.

Depois de treinados, os pesos da RNA foram guardados em arquivos do tipo texto para serem usados nos testes em que a rede classificou os padrões que não entraram no conjunto de treinamento.

5.5.2.1 Treinamento e testes da RNA

O treinamento foi dividido em várias etapas. Os conjuntos de imagens foram dividido da seguinte maneira: 70% de toda base de dados foi usada no treinamento, e os 30% restantes para testes. Isso tanto para classificação de *clusters*, quanto para classificação de nódulos. Com essa divisão da base de dados, usou-se o modo "padrão"⁷ de correção de pesos durante o treinamento. Numa segunda experiência, o treinamento foi feito apenas com o conjunto de imagens de nódulos.

Para cada experimento foram treinadas várias configurações do número de neurônios nas camadas escondidas.

Os dados de entrada para a rede são recuperados de tabelas que armazenam os atributos de cada RI no processo de extração de atributos, para isso usou-se o *Database Desktop 7.0* do *Delphi*.

Depois de treinada, os pesos e limiares da rede são armazenados em arquivos texto, que são usados na fase de testes. Na fase de teste, o classificador recupera as informações guardadas em arquivo texto com as quais a rede é inicializada, as informações de um vetor de entrada são propagadas pela rede e a saída da rede é a resposta do classificador.

5.5.3 Classificação de Nódulos por Grau de Intensidade

A partir dos atributos de intensidade extraídos das regiões de interesse foi utilizado um algoritmo de "clusterização" afim de dividir as RIs por grau de densidade conforme a padronização utilizadas em laudos mamográficos. Então foram considerados 4 valores iniciais: até 25% de densidade, de 25% a 50% de densidade, de 50% a 75% de densidade e mais que 75% de densidade.

⁷ No modo "padrão" a correção dos pesos ocorre a cada apresentação de um exemplo a rede.

Conforme descrito no item 4.5 do capítulo 4, foi implementado um algoritmo de clusterização onde, o usuário através de uma interface aciona um botão e então é calculado o padrão de densidade da RI em questão.

Os algoritmos de "clusterização" são métodos que dividem um conjunto de n observações em g grupos (*clusters*), tal que membros do mesmo grupo sejam mais parecidos que membros de um grupo diferente (Ripley, 2000).

O número de grupos pode ser pré associado ou pode ser decidido pelo algoritmo. Formalmente um algoritmo de "clusterização" deve mapear $c : \{1, \dots, n\} \rightarrow \{1, \dots, g\}$ associando um grupo com todos os exemplos.

Foi implementado o *K-means*, que é uma técnica de classificação baseada no critério de mínima dispersão intraclases. *K-means* começa com N objetos, caracterizados por seus respectivos vetores de atributos, e tenta classificá-los em K classes. Entretanto, o número de classes tem que ser previamente conhecido. Este método requer K protótipos iniciais P_i , os quais foram estimados automaticamente.

Como um método de particionamento, o algoritmo divide os exemplos dentro de um número pré associado de *clusters*. Para dados em um espaço Euclidiano, pode associar-se o centro de um *cluster* $\mathbf{m}$ para cada grupo, que foram 4. Então, escolher os centros dos *clusters* e dos grupos tais que seja mínima a distância da soma dos quadrados de cada exemplo para o centro do *cluster*.

$$\text{Min}_{c, \mathbf{m}_j} \sum ||\mathbf{x}_i - \mathbf{m}_{c(i)}||^2 \quad (5.2)$$

A minimização sobre os centros dos *clusters* é simples: basta escolher o centro do *cluster* j para ser a média dos exemplos associados ao *cluster* j ; assim, conhecer os *clusters* c é suficiente para definir os centros dos *clusters*. Todos os exemplos são associados simultaneamente a seus centros de *clusters* mais próximos, cada centro de *cluster* movido para a média do grupo, e esse processo é repetido. Um grupo pode ser descartado pelo algoritmo, e então pode ser escolhido menos que K *clusters*. A cada nova entrada de um exemplo são recalculados os centros dos *clusters*. A Figura 5. 21 mostra a interface para a "clusterização".

(a)

(b)

Figura 5.21 Interface de “clusterização”: (a) com centros de classe calculados para parte do conjunto de dados; (b) valores de centros atualizados com inserção de novos exemplos

5.6 Apresentação dos resultados da classificação

Foi desenvolvida uma interface onde os resultados da classificação da RI selecionada pelo usuário é mostrada juntamente com a opção de se inserir informações extras de anamnese. Com isso os resultados poderão ser visualizados com informações mais detalhadas possibilitando uma melhor análise para o diagnóstico.

Classificação (Classification)

AbrirImagem | **Abrir_Im_Comparação**

Genética: Não

Terapia com Hormônios: Não

Menopausa: Sim

Idade: 65

Localização: QIE

Sugestão de Classificação

Categoria BI-RADS:

Cluster: suspeito

Grau de densidade:

[Image Preview 1] [Image Preview 2]

(a)

Classificação (Classification)

AbrirImagem | **Abrir_Im_Comparação**

Genética:

Terapia com Hormônios:

Menopausa:

Idade:

Localização:

Sugestão de Classificação

Categoria BI-RADS:

Grau de densidade:

(b)

Figura 5.22 - Interface de apresentação dos resultados: (a) resultados de *clusters* e (b) resultados de nódulos.

O classificador de estruturas mamográficas pode ser sintetizado através da Figura 5.23. Através do esquema está representado todos os procedimentos implementados para a classificação final.

Figura 5.23 - Esquema do processo de classificação Híbrida

Pelo esquema da Figura 5.23 é possível demonstrar as etapas e processos híbridos a partir da representação dos dois tipos de classificadores, *K-Means* e rede *Backpropagation*, ou seja, a combinação de classificadores. A utilização de diferentes representações dos padrões, descritores de intensidade e descritores geométricos, nas entradas dos classificadores. Por fim uma última condição para a caracterização de sistemas híbridos pode ser observada através da estratégia de combinação dos dados.

6

RESULTADOS E DISCUSSÕES

"Erros são, no final das contas, fundamentos da verdade. Se um homem não sabe o que uma coisa é, já é um avanço do conhecimento saber o que ela não é."
(Carl Jung)

Os resultados aqui apresentados correspondem aos testes feitos com as técnicas de classificação que compõem o classificador híbrido, além de experimentos com redes bayesianas, embora tal modelo não tenha sido incluído no esquema classificador final, pelos motivos que serão expostos mais adiante.

6.1 Resultados Obtidos com as Redes Bayesianas na Classificação de *Clusters* de Microcalcificações

Primeiramente foi usado uma base de dados de 60 imagens, com 20 atributos extraídos das RIs contendo *clusters* de microcalcificações, para a construção de uma rede bayesiana (Patrocínio, 2004). Os resultados da validação da rede foram resultantes da técnica *Cross-Validation*⁸ que é um dos recursos do próprio software. Os resultados dos testes são gravados em um arquivo separadamente e estão ligadas a seguir.

⁸ *Cross-Validation* (validação cruzada), onde um mesmo conjunto de dados é dividido de maneira aleatória várias vezes, e para cada divisão a rede é treinada e testada.

A Tabela 6.1 mostra os resultados da validação, cuja base de dados foi dividida em 6 conjuntos de dados e foram testados 10 vezes. O resultado a ser considerado é a média de todos os 10 testes com os 6 diferentes conjuntos, também disposto na tabela.

Tabela 6.1 - Resultados do *Cross-Validation* da rede gerada com a base de dados contendo todos os 20 atributos.

Validação – resultados (%)						
TESTES	Base1	Base2	Base3	Base4	Base5	Base6
Teste1	70	50	70	40	90	80
Teste2	60	60	80	70	60	50
Teste3	80	60	60	50	70	60
Teste4	90	70	50	80	50	70
Teste5	70	80	80	70	60	70
Teste6	60	50	70	60	70	80
Teste7	80	70	70	70	70	60
Teste8	90	80	60	50	80	50
Teste9	70	60	90	80	50	70
Teste10	70	50	60	60	50	50
Acerto Médio		66,33 %				

Os resultados mostrados na Tabela 6.1 foram inferiores aos esperados, pois o acerto médio ficou abaixo dos 70%. Daí a necessidade para gerar redes dividindo o conjunto de atributos.

A Tabela 6.2 apresenta os resultados da validação da rede gerada com apenas 11 atributos: momento_{11} , momento_{20} , momento_{12} , momento_{30} , $Hu1$, $Hu5$, perímetro, irregularidade, área, área_cluster e compacidade; que foram a mais significativa mostrada pelas análises descritas nas sessões anteriores.

Tabela 6.2 - Resultados do *Cross-Validation* da rede gerada com a base de dados contendo apenas 11 atributos.

Validação – resultados (%)						
TESTES	Base1	Base2	Base3	Base4	Base5	Base6
Teste1	70	80	90	80	80	90
Teste2	90	90	80	70	80	70
Teste3	90	80	90	90	70	70
Teste4	80	80	90	90	90	80
Teste5	70	90	90	80	90	90
Teste6	80	90	80	90	90	90
Teste7	90	80	80	90	90	90
Teste8	90	70	70	80	80	80
Teste9	80	80	80	90	90	90
Teste10	80	90	90	80	70	80
Acerto Médio				83,16%		

Os resultados mostrados na Tabela 6.2 foram bem melhores pois a taxa de acerto médio superou os 80%. Mesmo assim outras tentativas foram feitas, como por exemplo, utilização do PCA para seleção de atributos.

A Tabela 6.3 apresenta os resultados de uma rede gerada com uma base de dados contendo os 5 componentes principais da base de dados (compacidade, área, irregularidade, momento₁₁ e momento₁₂). Cada elemento desta tabela indica a porcentagem de acerto nos 10 conjuntos de testes considerados. Foram considerados 5 componentes principais porque eles representam 96% da variância dos dados originais.

Tabela 6.3 - Resultados do *Cross-Validation* da rede gerada com a base de dados contendo 5 componentes principais.

Validação - resultados (%)						
TESTES	Base1	Base2	Base3	Base4	Base5	Base6
Teste1	60	70	60	80	80	70
Teste2	80	70	70	90	70	90
Teste3	70	60	80	70	90	90
Teste4	70	60	90	90	80	80
Teste5	80	80	90	70	80	80
Teste6	90	80	70	50	80	70
Teste7	90	90	90	70	60	90
Teste8	90	80	90	90	70	60
Teste9	80	90	70	60	90	80
Teste10	80	90	80	90	70	70
Acerto Médio			77,66%			

A porcentagem média de acerto mostrada na Tabela 6.3 (77,7%) é um pouco menor que a da Tabela 6.2 (83,2%). No entanto, em relação ao número de atributos maior como na Tabela 6.1, que considerou os 20 atributos originais, obtendo-se uma taxa de acerto de 66,3%, a utilização da técnica PCA produziu um resultado melhor.

Contudo, há necessidade de investigação mais detalhada em busca de resultados mais precisos, pois em se tratando de diagnóstico médico, o resultado tem que ser o mais confiável possível para ser utilizado eficientemente como segunda opinião por especialistas.

6.2 Resultados Obtidos com as Redes *Backpropagation* na Classificação de *Clusters* de Microcalcificações

Nessa etapa da classificação, os resultados obtidos a partir de RNA *Backpropagation* são representados através da metodologia das curvas ROC, previamente descrito no capítulo 4, além das porcentagens de acertos e uma análise estatística em termos de sensibilidade e especificidade, que quantifica os erros em falso-negativo (FN) e falso-positivo (FP) e os acertos

em verdadeiro-positivo (VP) e verdadeiro-negativo (VN). Com as curvas ROC, é possível saber a probabilidade do sistema acertar em casos positivos e em casos negativos, tendo assim uma medida de eficiência através da área sob a curva. Um sistema com alta taxa de falso-negativo significa que muitos sinais não estão sendo detectados (pacientes com a doença são identificadas como se não tivessem), assim como alta taxa de falso-positivo, significa que está sendo detectado sem que exista (pacientes que não tem a doença, são identificadas como se tivessem). É também a forma mais usual para descrever performance de sistemas na área médica.

Foram treinadas redes com diferentes configurações utilizando uma base com 213 RIs, todas contendo *clusters* : 96 *clusters*, cujo laudo indicava lesão com algum grau de suspeita de malignidade e 117 com laudo indicando lesão não suspeita. Foi utilizada uma metodologia conhecida como validação cruzada, que consiste em dividir a base em 10 conjuntos aleatórios de dados, sendo 70% (67 "suspeitos" e 82 "não suspeitos") dos dados utilizados para treinamento e 30% (29 "suspeitos" e 35 "não suspeitos") para testes. Dessa forma, cada configuração da RNA é treinada e testada 10 vezes, cada vez com uma divisão diferente da base de dados. Foram testadas 3 diferentes redes cujas configurações são mostradas na Tabela 6.4.

Tabela 6.4 – Configurações das Redes testadas

REDES Configurações	Neurônios Entrada	Neurônios Escondida 1	Neurônios Escondida 2	Neurônios Saída	Médias Ciclos	Média Erro mínimo
REDE 1	20	50	-	2	28596347	0,09647
REDE 2	20	40	15	2	125365412	0,099857
REDE 3	20	30	10	2	75896327	0,099234

A média de resultados dos 10 testes feitos para cada uma das 3 redes estão descritos nas Tabela 6.5, 6.6 e 6.7. Cada linha das tabelas corresponde a média de acertos dos 10 testes feitos para cada conjunto de dados.

Tabela 6.5 - Resultados de 10 testes para a rede1

TESTES	VP	VN	FP	FN	Precisão	Erro	Acerto%
TesteBase1	0,8857	0,8966	0,1034	0,1143	0,8906	0,1094	89,0625
TesteBase2	0,9143	0,8621	0,1379	0,0857	0,8906	0,1094	89,0625
TesteBase3	0,9429	0,7931	0,2069	0,0571	0,8750	0,1250	87,5000
TesteBase4	0,9714	0,7586	0,2414	0,0286	0,8750	0,1250	87,5000
TesteBase5	1,0000	0,7241	0,2759	0,0000	0,8750	0,1250	87,5000
TesteBase6	0,9714	0,7931	0,2069	0,0286	0,8906	0,1094	89,0625
TesteBase7	0,9143	0,8621	0,1379	0,0857	0,8906	0,1094	89,0625
TesteBase8	0,9429	0,8966	0,1034	0,0571	0,9219	0,0781	92,1875
TesteBase9	1,0000	0,7586	0,2414	0,0000	0,8906	0,1094	89,0625
TesteBase10	0,8000	0,6552	0,3448	0,2000	0,7344	0,2656	73,4375
Média	0,9343	0,8000	0,2000	0,0657	0,8734	0,1266	87,3438

A média dos testes feitos com a rede1 mostra o seu bom desempenho, com alguns dos testes apresentando uma taxa de respostas falsas-negativas significativa, embora a média dessa taxa tenha sido baixa. Isso ocorreu porque, em alguns dos testes a taxa de ocorrência de falso-negativo foi zero. Esse é um dado bastante importante, pois um diagnóstico falso -negativo, na prática, significa que a doença não foi identificada e portanto não será tratada, diminuindo as chances de cura, pois a doença existe.

Tabela 6.6 - Resultados de 10 testes para a rede2

TESTES	VP	VN	FP	FN	Precisão	Erro	Acerto %
TesteBase1	0,9429	0,8276	0,1724	0,0571	0,8906	0,1094	89,0625
TesteBase2	0,9429	0,7931	0,2069	0,0571	0,8750	0,1250	87,5000
TesteBase3	0,9714	0,8621	0,1379	0,0286	0,9219	0,0781	92,1875
TesteBase4	1,0000	0,7586	0,2414	0,0000	0,8906	0,1094	89,0625
TesteBase5	1,0000	0,7586	0,2414	0,0000	0,8906	0,1094	89,0625
TesteBase6	0,9429	0,7931	0,2069	0,0571	0,8750	0,1250	87,5000
TesteBase7	0,9143	0,8276	0,1724	0,0857	0,8750	0,1250	87,5000
TesteBase8	1,0000	0,9310	0,0690	0,0000	0,9688	0,0313	96,8750
TesteBase9	0,9714	0,7241	0,2759	0,0286	0,8594	0,1406	85,9375
TesteBase10	0,8857	0,6207	0,3793	0,1143	0,7656	0,2344	76,5625
Media	0,9571	0,7897	0,2103	0,0429	0,8813	0,1188	88,1250

Em relação aos resultados da rede1, a rede2 apresentou um pequeno acréscimo na média de acertos. Uma das justificativas para isso é que taxas de falso-negativo iguais a zero ocorreram em mais testes que para rede1. A média das taxas de respostas verdadeiras-negativas diminuiu em relação à rede1, o que fez aumentar a média da taxa de respostas falsas-positivas.

Tabela 6.7 - Resultados de 10 testes para a rede3

TESTES	VP	VN	FP	FN	Precisão	Erro	Acerto %
TesteBase1	0,8857	0,7931	0,2069	0,1143	0,8438	0,1563	84,3750
TesteBase2	0,8857	0,8621	0,1379	0,1143	0,8750	0,1250	87,5000
TesteBase3	0,9143	0,7241	0,2759	0,0857	0,8281	0,1719	82,8125
TesteBase4	0,9714	0,7241	0,2759	0,0286	0,8594	0,1406	85,9375
TesteBase5	1,0000	0,6552	0,3448	0,0000	0,8438	0,1563	84,3750
TesteBase6	0,9143	0,7586	0,2414	0,0857	0,8438	0,1563	84,3750
TesteBase7	0,9429	0,7241	0,2759	0,0571	0,8438	0,1563	84,3750
TesteBase8	0,7714	0,9655	0,0345	0,2286	0,8594	0,1406	85,9375
TesteBase9	0,8000	0,8966	0,1034	0,2000	0,8438	0,1563	84,3750
TesteBase10	1,0000	0,6207	0,3793	0,0000	0,8281	0,1719	82,8125
Media	0,9314	0,7517	0,2483	0,0686	0,8500	0,1500	85,0000

A rede3 foi a que apresentou menor acerto médio, com taxa de 85%. Contudo as respostas falsas-negativas mantiveram-se em pequena taxa como na rede1 e na rede2. A média das taxas falsas-positivas teve pequeno aumento, o que pode explicar a queda da média de acertos.

A Tabela 6.8 apresenta um resumo comparativo dos desempenhos das redes e também a área sob a curva ROC (representado por A_z na tabela). Como já comentado anteriormente, esse valor de A_z mede a eficiência do sistema. Em um sistema ideal, o valor da área sob a curva ROC deve ser 1, visto que a probabilidade de acertos nos casos positivos é 1, assim como a probabilidade de acertos nos casos negativos. Portanto, num sistema real, quanto mais próximo de 1 for A_z , melhor a eficiência do sistema.

Tabela 6.8 - Resumo dos desempenho das 3 rede, mais a área sob a curva ROC

Clusters	VP	VN	FP	FN	Precisao	Erro	Az	Acerto%
REDE1	0,9429	0,8069	0,1931	0,0571	0,8813	0,1188	0,9348	88,1250
REDE2	0,9714	0,8000	0,2000	0,0286	0,8938	0,1063	0,9407	89,3750
REDE3	0,9314	0,7517	0,2483	0,0686	0,8500	0,1500	0,9131	85,0000

Na Tabela 6.8 é possível observar que os valores de A_z ficaram acima de 0,9 para as três redes, destacando-se a rede2. Também é possível observar que os resultados obtidos com os testes para as redes *Backpropagation* foram bem superiores aos testes realizados com o simulador de redes bayesianas apresentados no item 6.1. Dessa maneira, essa foi a rede escolhida para integrar o classificador, ficando como a ferramenta para classificação de *clusters* em suspeitos e não suspeitos.

As figuras 6.1, 6.2 e 6.3 mostram as curvas ROC da rede1, rede2 e rede3 respectivamente.

Figura 6.1 - Curva ROC que representa a rede1.

Figura 6.2 - Curva ROC que representa a rede2.

Figura 6.3 - Curva ROC que representa a rede3.

6.3 Resultados Obtidos com Redes *Backpropagation* na Classificação de Nódulos

Assim como para a classificação de *clusters* utilizando redes *Backpropagation*, para a classificação de nódulos, foram treinadas redes com diferentes configurações utilizando uma base com 155 RIs: 33 classificadas na categoria BI-RADS 1; 33 classificadas na categoria BI-RADS 2; 36 classificadas na categoria BI-RADS 3; 30 classificadas na categoria BI-RADS 4; e

23 classificadas na categoria BI-RADS 5. A classificação prévia das RIs foi seguindo laudos dos radiologistas coletados durante o trabalho.

Também para a classificação dos nódulos, utilizou-se validação cruzada, dividindo a base de dados em 70% para treinamento das redes e 30% para testes e essa divisão foi feita 10 vezes, selecionando aleatoriamente 10 diferentes conjuntos de treinamento e teste. A Tabela 6.9 mostra as configurações testadas.

Tabela 6.9 - Configurações das Redes testadas

REDES Configurações	Neurônios Entrada	Neurônios Escondida 1	Neurônios Escondida 2	Neurônios Saída	Médias Ciclos	Média Erro mínimo
REDE 1	20	15	20	5	28596347	0,09647
REDE 2	20	15	30	5	125365412	0,099857
REDE 3	20	15	10	5	75896327	0,099234

Os resultados dos testes feitos com cada uma das redes estão descritos nas Tabelas 6.10, 6.11 e 6.12.

Para a classificação de nódulos, os resultados trazem uma taxa a mais, que é a porcentagem de acerto das classes. Na classificação por categorias BI-RADS, a única categoria que indica ausência de sinal, ou seja, imagens negativas, é a categoria BI-RADS 1. Por isso dois tipos de respostas estão sendo consideradas: (a) respostas verdadeiras-positivas e verdadeiras-negativas e (b) respostas quanto a classe da lesão, pois um verdadeiro-positivo pode estar classificado em uma das 4 últimas categorias BI-RADS (2 a. 5). Isso significa que o classificador foi avaliado em termos de dois tipos de informação de saída que devem ser apresentadas:

- classificação como contendo ou não um nódulo (a inexistência do nódulo leva automaticamente à categoria BI-RADS 1);
- constatando a existência do nódulo, classificação da RI numa das outras demais categorias BI-RADS, em função das características do achado mamográfico;

Então, nas tabelas a seguir, ao lado dos índices de VP, VN, FP e FN, são apresentadas algumas taxas de acerto das redes, definidas da seguinte forma:

- Acerto total: é a soma dos acertos VP e VN;
- Acertos das classes: é o número de acertos em cada uma das classes BI-RADS

Tabela 6.10 - Resultados dos 10 testes da rede1

TESTES	VP	VN	FP	FN	%AcetoTotal	%AcertoClasse	Precisão	Erro
TesteBase1	0,9459	0,6000	0,4000	0,0540	87,2340	68,0851	0,8723	0,1270
TesteBase2	0,8648	0,7000	0,3000	0,1351	82,9788	70,2128	0,8298	0,1700
TesteBase3	1,0000	0,5000	0,5000	0,0000	89,3617	78,7234	0,8936	0,1060
TesteBase4	0,8378	0,6000	0,4000	0,1621	78,7234	65,9574	0,7872	0,2120
TesteBase5	1,0000	0,6000	0,4000	0,0000	91,4894	72,3404	0,9149	0,0850
TesteBase6	0,8918	0,8000	0,2000	0,1081	87,2340	80,8511	0,8723	0,1270
TesteBase7	0,9729	0,5000	0,5000	0,0270	87,2340	76,5957	0,8723	0,1270
TesteBase8	0,7567	0,9000	0,1000	0,2432	78,7234	74,4681	0,7872	0,2120
TesteBase9	0,7837	1,0000	0,0000	0,2162	82,9788	74,4681	0,8298	0,1700
TesteBase10	0,9459	0,7000	0,3000	0,0540	89,3617	68,0851	0,8936	0,1060
Média	0,9000	0,6900	0,3100	0,1000	85,5319	72,9787	0,8553	0,1446

O desempenho da rede1 pode ser resumido pelas médias. O acerto médio dos testes foi em torno de 85% enquanto o acerto médio das classes ficou próximo de 73%, embora o teste de número 6 tenha atingido 80% de acerto das classes.

A taxa das respostas verdadeiras-positivas foi alta. Nos testes 3 e 5, por exemplo, todos os sinais foram corretamente classificados. Entretanto, dos casos negativos, parte significativa foi classificada erroneamente em vários testes, resultando numa taxa média baixa de verdadeiro-negativo de 0,7. Isso implica aumento de respostas falsas-positivas.

Tabela 6.11 - Resultados dos 10 testes da rede2

TESTES	VP	VN	FP	FN	%AcetoTotal	%AcertoClasse	Precisão	Erro
TesteBase1	0,8919	0,5000	0,5000	0,1081	80,8511	70,2128	0,8085	0,1915
TesteBase2	0,9730	0,5000	0,5000	0,0270	87,2340	72,3404	0,8723	0,1277
TesteBase3	0,8378	0,6000	0,4000	0,1622	78,7234	78,7234	0,7872	0,2128
TesteBase4	0,7838	0,8000	0,2000	0,2162	78,7234	63,8298	0,7872	0,2128
TesteBase5	0,9730	0,4000	0,6000	0,0270	85,1064	74,4681	0,8511	0,1489
TesteBase6	0,9730	0,5000	0,5000	0,0270	87,2340	78,7234	0,8723	0,1277
TesteBase7	0,7838	0,8000	0,2000	0,2162	78,7234	72,3404	0,7872	0,2128
TesteBase8	0,7838	0,9000	0,1000	0,2162	80,8511	76,5957	0,8085	0,1915
TesteBase9	0,8108	1,0000	0,0000	0,1892	85,1064	68,0851	0,8511	0,1489
TesteBase10	0,9189	0,8000	0,2000	0,0811	89,3617	70,2128	0,8936	0,1064
Media	0,8730	0,6800	0,3200	0,1270	83,1915	72,5532	0,8319	0,1681

Nos testes realizados com a rede2 o acerto médio foi inferior ao da rede1. As médias das taxas de falsos-positivos e falsos-negativos tiveram um pequeno acréscimo e a média dos acertos das classes manteve-se muito próximo ao da rede1.

Tabela 6.12 - Resultados dos 10 testes da rede3

TESTES	VP	VN	FP	FN	%AcetoTotal	%AcertoClasse	Precisão	Erro
TesteBase1	0,8649	0,6000	0,4000	0,1351	80,8511	72,3404	0,8085	0,1915
TesteBase2	0,9459	0,5000	0,5000	0,0541	85,1064	70,2128	0,8511	0,1489
TesteBase3	0,8108	0,7000	0,3000	0,1892	78,7234	68,0851	0,7872	0,2128
TesteBase4	0,8378	0,7000	0,3000	0,1622	80,8511	70,2128	0,8085	0,1915
TesteBase5	0,9189	0,6000	0,4000	0,0811	85,1064	72,3404	0,8511	0,1489
TesteBase6	0,8919	0,6000	0,4000	0,1081	82,9787	76,5957	0,8298	0,1702
TesteBase7	0,8649	0,7000	0,3000	0,1351	82,9787	70,2128	0,8298	0,1702
TesteBase8	0,8378	0,8000	0,2000	0,1622	82,9787	70,2128	0,8298	0,1702
TesteBase9	0,7568	0,9000	0,1000	0,2432	78,7234	74,4681	0,7872	0,2128
TesteBase10	0,8108	0,7000	0,3000	0,1892	78,7234	72,3404	0,7872	0,2128
Média	0,8541	0,6800	0,3200	0,1459	81,7021	71,7021	0,8170	0,1830

O desempenho da rede 3 foi o pior entre as 3 redes, embora os valores das médias tenha ficado próximos aos das redes anteriores. O média de acerto das classes caiu, não atingindo 72%, e a média de acertos também caiu, não atingindo 82%.

As curvas ROC mostradas nas Figura 6.4, 6.5 e 6.6 representam as performances da rede1, rede2 e rede3 respectivamente.

Figura 6.4 - Curva ROC REDE1

Figura 6.5 - Curva ROC REDE2

Figura 6.6 - Curva ROC REDE3

As áreas sob as curvas ROC para cada uma das redes foram: $A_z(\text{Rede1}) = 0,865$; $A_z(\text{Rede2}) = 0,845$; e $A_z(\text{Rede3}) = 0,835$, evidenciando também a melhor performance da rede1, embora A_z não tenha atingido 0,9.

Na tentativa de melhorar a performance da classificação de nódulos, outras redes foram testadas, e foi então considerado mais um atributo como entrada para o classificador: quantização de densidade.

O grau de densidade de uma imagem é um dos pontos analisados nos laudos e constitui uma informação importante, além de, nos casos de imagens de nódulos, a densidade ser um fator ainda mais significativo, pois é uma das maneiras de caracterização deste tipo de lesão.

Através da metodologia descrita no item 5.6 do capítulo 5, as regiões foram separadas por grau de densidade através do *K-Means* em 4 grupos: o grupo 1, de imagens não densas e baixa densidade (predominância do parênquima lipo-substituído); grupo 2, com imagens de densidade não tão baixas (parênquima apresenta-se parcialmente lipo-substituído); grupo 3, de imagens com densidade significativa; e grupo 4 imagens com alta densidade. Esses conceitos de quantidade de densidade, na prática médica, são bastante subjetivos, causando grande confusão na classificação mesmo por especialistas, que frequentemente dão respostas diversas para uma mesma imagem ou até se contradizem ao analisar a mesma imagem várias vezes. Por isso, através dos atributos de intensidade extraídos das imagens utilizou-se uma normalização dos grupos para representá-los como entrada da rede *Backpropagation* em termos de porcentagem do grau de densidade que cada um dos grupos representam. Assim, o grupo 1 foi representado, de maneira a indicar que pertence ao grupo de imagens com até 25% de densidade, grupo 2 de forma a indicar que pertence ao grupo com até 50% de densidade, grupo 3 indicando que pertence ao grupo de de 51% a 75% e grupo 4 que pertence ao grupo com mais de 75% chegando à 100% de densidade.

As novas redes testadas seguem agora as configurações descritas na tabela 6.13., onde basicamente foram alterados apenas o número de neurônios na camada de entrada.

Tabela 6.13 - Configurações das Redes testadas

REDES Configurações	Neurônios Entrada	Neurônios Escondida 1	Neurônios Escondida 2	Neurônios Saída	Médias Ciclos	Média Erro mínimo
REDE 1	21	15	20	5	56251245	0,0994114
REDE 2	21	15	30	5	82687452	0,097254
REDE 3	21	15	10	5	39258697	0,098142

As Tabelas 6.14, 6.15 e 6.16 descrevem os resultados dos testes feitos com as 3 redes da tabela 6.13.

Tabela 6.14 - Resultados dos 10 testes da rede1 considerando densidade

Nodulos	VP	VN	FP	FN	%AcetoTotal	%AcertoClasse	Precisão	Erro
TesteBase1	0,9700	0,7000	0,3000	0,0270	91,4894	70,2128	0,9100	0,0850
TesteBase2	0,8900	0,8000	0,2000	0,1081	87,2340	72,3404	0,8700	0,1277
TesteBase3	0,9700	0,6000	0,4000	0,0270	89,3617	76,5957	0,8900	0,1064
TesteBase4	0,8900	0,6000	0,4000	0,1081	82,9787	74,4681	0,8300	0,1702
TesteBase5	0,9700	0,6000	0,4000	0,0270	89,3617	70,2128	0,8900	0,1064
TesteBase6	0,9400	0,9000	0,1000	0,0541	93,6170	76,5957	0,9300	0,0638
TesteBase7	1,0000	0,7000	0,3000	0,0000	93,6170	78,7234	0,9300	0,0638
TesteBase8	0,8100	0,8000	0,2000	0,1892	80,8511	80,8511	0,8000	0,1915
TesteBase9	0,8100	0,9000	0,1000	0,1892	82,9787	76,5957	0,8200	0,1702
TesteBase10	0,8900	0,8000	0,2000	0,1081	87,2340	65,9574	0,8700	0,1277
Media	0,9100	0,7400	0,2600	0,0838	87,8723	74,2553	0,8800	0,1213

Os resultados realizados com a rede1 mostrou melhoras com essa nova configuração dos atributos de entrada, considerando a densidade da imagem, em relação à rede1 considerando apenas os 20 atributos descritores geométricos anteriores. A média de acertos das classes aumentou mais de um ponto percentual, tendo o teste 8 registrado próximo de 81% de acerto das classes.

A média geral de acertos também foi maior que na rede1 anterior, chegando próximo de 88%, o que refletiu num aumento dos verdadeiros-negativos.

Tabela 6.15 - Resultados dos 10 testes da rede2 considerando densidade

TESES	VP	VN	FP	FN	%AcetoTotal	%AcertoClasse	Precisão	Erro
TesteBase1	0,9189	0,8000	0,2000	0,0811	89,3617	74,4681	0,8936	0,1064
TesteBase2	1,0000	0,8000	0,2000	0,0000	95,7447	74,4681	0,9574	0,0426
TesteBase3	0,8919	0,8000	0,2000	0,1081	87,2340	76,5957	0,8723	0,1277
TesteBase4	0,8108	0,9000	0,1000	0,1892	82,9787	68,0851	0,8298	0,1702
TesteBase5	0,9459	0,8000	0,2000	0,0541	91,4894	72,3404	0,9149	0,0851
TesteBase6	1,0000	0,7000	0,3000	0,0000	93,6170	80,8511	0,9362	0,0638
TesteBase7	0,8649	0,7000	0,3000	0,1351	82,9787	74,5106	0,8298	0,1702
TesteBase8	0,8378	1,0000	0,0000	0,1622	87,2340	74,4681	0,8723	0,1277
TesteBase9	0,8378	0,8000	0,2000	0,1622	82,9787	72,3404	0,8298	0,1702
TesteBase10	0,9459	0,9000	0,1000	0,0541	93,6170	74,4681	0,9362	0,0638
Media	0,9054	0,8200	0,1800	0,0946	88,7234	75,1064	0,8872	0,1128

Os resultados da rede2 com os 21 atributos mostrou melhoras em relação não apenas à rede1 com 21 atributos, mas também às redes com 20 atributos. A média de acertos das classes passou de 75% de acerto.

A média geral de acertos ficou próxima de 89%, o que também se refletiu no aumento da média de verdadeiros-negativos, assim como na rede1 com 21 atributos, embora tenha sido um pouco menor a média de verdadeiros-positivos que na rede1 com 21 atributos.

Tabela 6.16 - Resultados dos 10 testes da rede3 considerando densidade

TESTES	VP	VN	FP	FN	%AcetoTotal	%AcertoClasse	Precisão	Erro
TesteBase1	0,8378	0,8000	0,2000	0,1622	82,9787	78,7234	0,8298	0,1702
TesteBase2	0,8649	0,6000	0,4000	0,1351	80,8511	74,4681	0,8085	0,1915
TesteBase3	0,8378	0,6000	0,4000	0,1622	78,7234	70,2128	0,7872	0,2128
TesteBase4	0,8649	0,6000	0,4000	0,1351	80,8511	74,4681	0,8085	0,1915
TesteBase5	0,8919	0,9000	0,1000	0,1081	89,3617	74,4681	0,8936	0,1064
TesteBase6	0,9189	0,7000	0,3000	0,0811	87,2340	76,5957	0,8723	0,1277
TesteBase7	0,9189	0,5000	0,5000	0,0811	82,9787	76,5957	0,8298	0,1702
TesteBase8	0,8108	0,6000	0,4000	0,1892	76,5957	80,8511	0,7660	0,2340
TesteBase9	0,9459	0,8000	0,2000	0,0541	91,4894	74,4681	0,9149	0,0851
TesteBase10	1,0000	0,6000	0,4000	0,0000	91,4894	68,0851	0,9149	0,0851
media	0,8892	0,6700	0,3300	0,1108	84,2553	74,8936	0,8426	0,1574

Finalmente, os resultados da rede3 com da Tabela 6.16 mostram média de acerto das classes próximo de 75%, e, embora menor que o da rede2 com 21 atributos, foi melhor que das redes com 20 atributos.

A média geral de acertos ficou em torno de 84%, com redução das médias de verdadeiros-positivos e verdadeiros-negativos, comparadas com os resultados das 2 avaliações anteriores.

Um resumo do desempenho de todos testes feitos com RIs de nódulos podem ser visualizado na tabela 6.17, bem com os valores de A_z para cada uma das redes, o que facilita a análise da melhor configuração de rede.

Tabela 6.17 - Resumo dos desempenho das 3 redes, mais a área sob a curva ROC

Médias	VP	VN	FP	FN	%AcertoTotal	%AcertoClasse	Precisão	Erro	Az
Rede1	0,9000	0,6900	0,3100	0,1000	85,5310	72,9780	0,8500	0,1440	0,8650
Rede1 - densidade	0,9160	0,7400	0,2600	0,0800	87,8720	74,2550	0,8800	0,1210	0,906
Rede2	0,8730	0,6800	0,3200	0,1300	83,1910	72,5530	0,8300	0,1680	0,8450
Rede2 - densidade	0,9050	0,8200	0,1800	0,0900	88,7230	74,2550	0,8800	0,1130	0,9250
Rede3	0,8540	0,6800	0,3200	0,1500	81,7020	71,7020	0,8100	0,1830	0,8350
Rede3 - densidade	0,8890	0,6700	0,3300	0,1100	84,2550	74,8900	0,8400	0,1570	0,8535

Através da Tabela 6.17 é possível notar que as redes considerando o atributo de densidade melhoraram as performances da classificação. Em todos casos, as médias de acertos das classes aumentaram quando considerado o atributo de densidade. Exceto para as médias de verdadeiro-negativo e falso-positivo para as redes3, em que, considerando a densidade, os valores pioram sutilmente. Todas as outras respostas melhoraram com a adição da densidade.

Os valores de A_z mostram bem a melhora da performance da classificação das redes com 21 atributos, pois nesses casos, os valores foram acima das redes com 20 atributos, principalmente nas redes 1 e 2, cujos A_z ficaram acima de 0,9. Com destaque para a rede2 que atingiu 0,92 e por isso foi a escolhida para integrar o classificador final.

As curvas ROC das redes de 21 atributos são mostradas nas Figuras 6.7, 6.8 e 6.9 respectivamente a rede1, rede2 e rede3.

Figura 6.7 - Curva ROC REDE1_densidade

Figura 6.8 - Curva ROC REDE2_densidade

Figura 6.9 - Curva ROC REDE3_densidade

7

CONSIDERAÇÕES FINAIS

*"Eu poderia estar passando todo esse tempo aprendendo a voar.
Há tanta coisa para aprender!"
(Richard Bach)*

As investigações contidas nesse trabalho e a metodologia desenvolvida foram extensas, na tentativa de se conseguir mais uma ferramenta que possa auxiliar no diagnóstico precoce do câncer de mama, sobretudo para lesões encontradas em mamas densas, cujo padrão de parênquima não é característica apenas de mulheres jovens, mas também de mulheres em idades mais avançadas submetidas às atuais terapias de reposição hormonal. O problema é que, para mamas densas, torna-se ainda mais difícil o rastreamento da doença através da mamografia, devido ao baixo contraste das imagens.

7.1 Investigação da análise de lesões mamográficas segundo especialistas

Através da revisão bibliográfica sobre tipos de lesões mamárias e suas visibilizações na mamografia, foi possível entender um pouco da análise feita pelos especialistas ao examinarem um mamograma, e o que eles procuram para caracterizar uma lesão de acordo com um grau de importância para a detecção do câncer de mama.

A padronização de laudo BI-RADS que os serviços de mamografia vêm utilizando auxilia para a formalização do raciocínio seguido pelo médico para análise de um mamograma. Apesar disso, a análise continua sendo bastante subjetiva e a identificação de uma lesão como suspeita ou não suspeita, ou mesmo sua detecção, depende muito da experiência do médico. Essa é mais uma motivação para o uso de ferramentas automáticas para o auxílio ao diagnóstico, que tentam fornecer uma análise baseada em dados mais objetivos através de valores extraídos da própria imagem.

A quantização de densidade do parênquima mamário através da mamografia, como pode-se observar pela literatura, é uma tarefa bastante complicada pela subjetividade da análise visual dos médicos. Devido ao baixo contraste da imagem e também fatores físicos de formação da imagem mamográfica que influenciam na sua qualidade, é difícil relacionar numa escala linear a densidade ótica da imagem com a densidade da mama.

7.2 Atributos Extraídos das Imagens Mamográficas

Geralmente os trabalhos de classificação de achados mamográficos encontrados na literatura enfatizam a classificação de microcalcificações isoladas e não de *clusters*, ou então classificam *clusters* a partir de atributos extraídos de cada microcalcificação individualmente. Além disso, trabalhos que classificam nódulos geralmente utilizam apenas atributos da imagem em níveis de cinza, e não uma imagem binarizada com o nódulo segmentado.

Tanto em relação aos *clusters* quanto aos nódulos é de fundamental importância a caracterização de formato, e no caso dos *clusters*, de distribuição das microcalcificações, pode ser um importante indicativo da suspeita de malignidade, conforme o que foi descrito no capítulo 2.

Sendo assim, foi necessário investigar um conjunto de atributos que descrevessem essas características de modo adequado. A partir da teoria dos momentos, verificou-se que a descrição geométrica completa dos *clusters* de microcalcificações era bastante eficiente, pois possibilitava extrair características que descrevem exatamente o que é analisado na prática médica, como distribuição, distância, área, forma e irregularidade.

A descrição dos *clusters* com momentos e os outros descritores de forma serviram para mostrar que, mesmo não extraíndo características diretamente de cada microcalcificação pertencente ao *cluster*, foi possível representar bem os casos da base de dados, além de evidenciar a relevância de uma análise detalhada destes achados antes do encaminhamento para o exame histológico. Além disso, comprovou-se que características geométricas dos *clusters* estão relacionadas com a questão da malignidade, como sugerido na literatura.

A teoria dos momentos também foi eficiente na descrição geométrica dos nódulos, pois foi possível separá-los de acordo com as possibilidades de seus formatos e contornos de acordo com uma das categorias BI-RADS.

Os atributos de intensidade contribuíram principalmente para a classificação dos nódulos, mas também serviram para quantizar a densidade local encontrada em mamografia, a análise do padrão textural e a variação de intensidade de pixels na imagem, medidas essas que forneceram informações importantes e mais objetivas da densidade da imagem.

Os nódulos foram representados por atributos híbridos, tem sido utilizados tanto descritores geométricos extraídos das imagens binarizadas, como atributos de intensidade extraídos de imagens originais, em níveis de cinza..

7.3 Classificação automática híbrida das lesões

Muitos trabalhos são encontrados na literatura descrevendo classificação automática de lesões mamárias. A maioria trata o problema como uma classificação considerando apenas duas classes, e abordando uma questão chave que é a classificação dos achados em maligno e benigno. Uma ferramenta automática que fornece como resultado uma afirmação de malignidade ou benignidade pode ser um problema sério, pois o responsável pelo laudo é o médico que analisa a imagem e a ferramenta automática deve servir apenas como uma espécie de segunda opinião a esse profissional. Além disso, tais classificadores não costumam tratar a lesão como pertencente a uma das 5 classes BI-RADS, nem mesmo se conhece até aqui trabalho que utiliza atributos híbridos, extraídos tanto da imagem original em níveis de cinza como da imagem binarizada para a classificação de uma mesma lesão nas classes BI-RADS.

A combinação de classificadores para a solução da classificação automática de lesões mamárias tem sido bastante explorada, conforme trabalhos descritos no capítulo 4. A classificação híbrida envolve quatro categorias: múltiplos classificadores, usando uma representação idêntica do padrão; múltiplos classificadores, usando representações distintas do padrão; estratégias de combinações dependentes dos dados e classificadores de múltiplos estágios.

As redes neurais artificiais correspondem a técnica que permitem flexibilidade de aplicação e eficiência no reconhecimento de padrões, conforme comprovado pela literatura investigada. As redes *Backpropagation* mostraram grande robustez na classificação tanto de nódulos quanto de *clusters*.

A fase de treinamento das redes *Backpropagation* é o processo mais importante para que se garanta uma boa performance na fase de testes, onde a rede assume capacidade de generalização ao receber novos padrões, o que depende muito do conjunto de dados de treinamento pela necessidade de representar a diversidade inter e intra classes. Para se conseguir uma boa performance de treinamento existem diversas técnicas e uma delas, a validação cruzada (*cross-validation*), foi comprovada como eficiente, principalmente para as redes treinadas para a classificação de nódulos, cujos números de exemplos de cada uma das classes eram diferentes, não existindo muitos exemplos.

Para a classificação supervisionada, as redes *Backpropagation* foram eficientes tanto para um problema de apenas duas classes (a classificação dos *clusters*), quanto para problema de várias classes, como na classificação dos nódulos entre as 5 categorias BI-RADS.

Como já citado anteriormente, a subjetividade da quantização da densidade da mama pela visibilização do mamograma impediu a utilização de classificação supervisionada para a separação das RIs, considerando os atributos de intensidade. Com isso, o uso de técnica de “clusterização” *K-means*, que é uma técnica de classificação não supervisionada, mostrou-se eficiente para a separação das lesões. Os centro dos grupos, aos quais as imagens foram agregadas, serviram de base para uma normalização de valores em função de porcentagens para que o grau de densidade de cada grupo de imagem fosse um valor representativo. Assim, esse dado serviu como um novo atributo, que foi inserido na entrada das redes *Backpropagation* para a classificação de nódulos.

Além disso, o *K-means* fornece uma informação de semelhança para que uma imagem possa ser comparada com outra.

A análise dos resultados da classificação com base nas curvas ROC possibilita o detalhamento da performance do classificador através da quantização não apenas de acerto total entre as classes, mas também na análise de detecção de um sinal, seja uma informação como a presença ou não de um padrão em uma classe, seja uma estrutura presente no padrão. Por exemplo: na classificação dos *clusters* o sinal é a suspeita (positiva) de malignidade; portanto, a classe não suspeita é representada pela ausência de sinal, caracterizado por um padrão verdadeiro-negativo; para os nódulos, o sinal é a presença ou não da estrutura na RI.

Os testes das redes mostraram que, ao acrescentar o atributo que representa a densidade, a performance da classificação teve melhora significativa, comprovada através dos valores de A_z apresentados na tabela 6.18 do capítulo 6.

As duas técnicas de classificação utilizadas nesse trabalho caracterizam uma classificação automática híbrida considerando atributos também híbridos, que é uma proposta inovadora para a solução de classificação automática de achados mamográficos via processamento de imagens.

A utilização de técnicas de classificação como as redes bayesianas implica no uso de variáveis discretas. Apesar da longa investigação dessa técnica, os métodos de discretização de variáveis contínuas utilizados para inserção dos dados na rede não são bem detalhados na literatura, o que dificultou o desenvolvimento de um procedimento de construção da rede com os dados utilizados nesse trabalho, que são essencialmente valores contínuos gerados pelos procedimentos de extração de atributos. Por isso os testes realizados com redes bayesianas foram feitos com simuladores (*softwares* fechados), impedindo que se verificassem as etapas de discretização e formatação das rotas e dos nós das redes. Embora as redes bayesianas sejam capazes de apresentar um detalhamento dos resultados através da interface gráfica que é possível se criar, de forma a apresentar a dependência entre as variáveis e a classe escolhida para cada padrão, a compreensão gráfica só acontece quando as informações são discretizadas de maneira conhecida.

As *Backpropagation* são ferramentas que possibilitam o uso de variáveis contínuas e de classe com distribuição não normal, que foi justamente a forma dos dados disponíveis para a classificação automática de achados mamográficos conforme a formalização dos padrões feitas nesse trabalho.

A representação de padrões por atributos como foi feita gera um conjunto de variáveis contínuas para descrever de maneira quantitativa uma estrutura e não de maneira apenas qualitativa, que é a forma como um padrão deve ser representado numa rede Bayesiana.

O método do *K-means*, técnica não supervisionada de "clusterização" para separar padrões, foi empregado pois suas características de formalização do procedimento de separação permite a utilização do conhecimento prévio do número de classes, que era preciso ser empregado no agrupamento das imagens por intensidade. Além disso, é uma técnica simples de ser implementada e bastante conhecida, com eficiência comprovada em várias aplicações descritas na literatura.

7.4 Conclusões Finais

Esse trabalho utilizou classificação híbrida com múltiplos classificadores usando representações distintas de padrões agregados à dependência de variáveis.

O classificador híbrido final considera 2 redes *Backpropagation* (técnica de classificação supervisionada) e uma técnica não supervisionada e "clusterizador" *K-means*. As redes *Backpropagation* são utilizadas de um modo quando a lesão analisada é um *cluster*; considerando a caracterização geométrica da lesão através dos atributos descritores geométricos, e de modo diverso quando a lesão é um nódulo. Para classificação dos nódulos, é utilizada a resposta do *K-means* como um dos atributos de entrada da rede *Backpropagation* que, dessa forma, passa a considerar não apenas a caracterização geométrica das lesões, mas também uma informação de densidade da RI.

Os resultados obtidos ficaram próximos a 89% de acerto total (verdadeiros-positivos mais verdadeiros- negativos) para nódulos e ultrapassaram 75% de acerto da classificação entre as classes BI-RADS produzindo valor de A_z superior a 0,92. Na classificação dos *clusters* os acertos totais da classificação ficaram próximos de 90%, com A_z superior a 0,94. Para ambas lesões, as taxas de respostas falsas-negativas ficaram abaixo de 0,1, o que significa baixo erro em relação a não detecção da doença quando o sinal está presente. Para a classificação dos *clusters* a taxa de respostas verdadeiras-positivas foi de 0,97 comprovando a identificação da

maioria das lesões suspeitas, e a taxa de respostas verdadeiras-negativas foi de 0,80 demonstrando que poucos *clusters* não suspeitos são classificados erroneamente como um falso-positivo. Os resultados obtidos para os nódulos, a taxa de respostas verdadeiras-positivas ficou um pouco acima de 0,90 representando que o classificador identificou a maior parte das imagens que apresentavam realmente uma lesão, e a taxa de respostas verdadeiras-negativas foi de 0,82 o que significa que as RIs que não apresentavam nódulos foram também, em sua maior parte, classificadas corretamente.

Embora a média de acertos da separação de um caso em termos de uma das 5 categorias BI-RADS, na classificação dos nódulos, não tenha atingido 80%, as taxas de verdadeiro-positivo foram altas, chegando ao valor 1 em alguns testes que, para esses casos, indica ausência de resposta falsa-negativa do classificador. Em alguns testes foi a taxa verdadeira-negativa que atingiu valor 1, descartando a hipótese de resposta falso-positiva. As informações de verdadeiro-positivo, verdadeiro-negativo, falso-positivo e falso-negativo são extremamente importantes e podem ser mais significativas do que simplesmente a informação de acertos totais da classificação, pois os erros falso-positivo e falso-negativo tem pesos diferentes em se tratando de diagnóstico médico e dependem do que é considerado como sinal. Apesar dos dois erros acarretarem danos, o falso-negativo pode ser relacionado com o aumento das chances de morte, visto que significa que a doença não foi detectada, embora exista, o que implica a falta de tratamento ou tratamento tardio, enquanto o falso-positivo implica principalmente tratamento desnecessário e custo aumentado.

Essa média de acertos inferior a 80% pode ser justificada com o pequeno e desigual conjunto de padrões representando cada uma das categorias BI-RADS.

A pouca quantidade de exemplos de *clusters* dividida pelas categorias BI-RADS impediu que fosse testada a classificação considerando as 5 categorias e não apenas suspeitos e não-suspeitos de malignidade, termos utilizados para sugerir ao especialista, como uma segunda opinião, que se preste mais atenção à lesão classificada como suspeita, ou nos casos de resultados não suspeitos, que se preste mais atenção na tentativa de evitar o exame histológico desnecessário.

Na prática, sugere-se que a utilização do classificador na rotina dos serviços de mamografia seja testada primeiramente através dos convênios já existentes entre o LAPIMO e hospitais, de maneira a ser ajustados alguns parâmetros de interface às necessidades que forem sendo identificadas com o uso do sistema.

Como parte do esquema CAD que vem sendo desenvolvido pelo grupo, o classificador híbrido finaliza uma etapa importante do CAD, que é utilizar as técnicas desenvolvidas tanto para a detecção de *clusters* como de nódulos, quantificar as lesões detectadas através da extração e análise dos atributos e classificá-las, fornecendo uma informação mais detalhada sobre as possibilidades de diagnóstico através do achado mamográfico.

O modelo desenvolvido pode ser ainda ajustado através de inserção de atributos para a classificação mais apurada das imagens em níveis de cinza, como mais atributos de intensidade nas redes *Backpropagation*. Além disso, as informações de anamnese da paciente para fornecer previsões da associação de um caso aos riscos de malignidade poderiam ser utilizadas através de investigações mais detalhadas de inferências bayesianas. Dessa maneira, a saída fornecendo informações auxiliares para que o especialista faça a tomada de decisão quanto ao diagnóstico seria mais completa. Mesmo assim, as informações fornecidas pelo classificador desenvolvido já são relevantes e apresenta um certo grau de detalhamento, como grau de densidade, categoria BI-RADS, disponibilidade de comparação de RIs com graus de densidade diferentes, além de informações de anamnese que estejam disponíveis em relação a possibilitar a apresentação na interface de saída de resultados a cada caso avaliado.

8

REFERÊNCIAS BIBLIOGRÁFICAS

Referências bibliográficas citadas no texto.

(ACR, 1995)

American College of Radiology: Breast imaging reporting (BI-RADS) *Second Edition*. Reston (VA): American College of Radiology, 1995

(Allan, 2001)

Allan, R. & Kinsner, W. – “A study of microscopic images of human breast disease using Competitive Neural Network” – *IEEE*, 0-7803-6715-4/01, 2001.

(Andersen, 19850)

ANDERSEN, C.H; ESSICK, G.K. & SIEGEL, R.M. – “Encoding of spatial location by posterior parietal neurons”, *Science*, V.230, pp. 456-458, 1985.

(Anderson, 1995)

ANDERSON, J. A. *An Introduction to Neural Network*. The MIT Press, Massachuts Institute of Technology, Massachusts, Cambridge 1995

(Andreassen, et. al., 1991)

Andreassen, S.; Jensen, F. V.; Olesen, K. G.; - "Medical expert systems based on causal probabilistic networks" - *Int. J. Biomed. Comput.*, 28:1-30, 1991.

(Antsakalis, 1992)

Antsakalis, P. J., “Neural Networks in Control Systems”, *IEEE Control Systems Magazine*, vol. 12, pp. 8-10, April 1992,

(Bacchus, 1994)

Bacchus, F.; "Using first-order probability logic for the construction of Bayesian networks" - In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 219-226, July 1994.

(Ballard, et. al., 1982)

Ballard, D. H.; Brown, C. M. - *Computer Vision* - Prentice-Hall Inc, Englewood Cliffs, New Jersey, 1982

(Benatti, 2003)

BENATTI, R. H. (2003). *Elaboração de banco de imagens mamográficas digitalizadas*. 64p. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

(Blane, 2002)

Blane, C. E.; et. al. - "Decreasing Rate of Fatty Involution at Screening Mammography" - *AUR*, 2002; 9: 895-898, USA..

(Box , 1992)

Box, G. E. P. & Tiao, G. C. - *Bayesian Inference in Statistical Analysis* - Wiley Classics Library - 1992

(Bruce, et. al. 1997)

Bruce, L. M.; Kalluri, R. - "An Anaysis of the Contribution of Scale in Mammographic Mass Classification" - Proceedings 19th International Conference – IEEE/EMBS, p. 1609-1612, October 1997.

(Bruce, et. al., 1999)

Bruce, L. M.; Adahami, R. R.. - "Classifying Mammographic Mass Shape Using the Wavelet Transform Modulus-Maxima Method" - IEEE Transactions on Medical Imaging, v. 18, n. 12, p. 1170-1178, December 1999.

(Bruce, 1999b)

Bruce, L. M.; Kallergi, M.. - "Effects of Image resolution and Sgmentation Method on Automated Mammographic Mass Shape Classification" - Proceedings. SPIE, v. 3661, p. 940-947, February 1999.

(Burnside, 2000)

Burnside, E.; Rubin, D.; Shachter, R. - "A Bayesian Network For Mammography", *Med. Phys.* 27, 4-12, 2000.

(Byng, 1998)

Byng, J. W.; et. al. - "Analysis of Mammographic Density and Breast Cancer Risk from Digitized Mammograms" - *RadioGraphics* 1998; 18: 1587-1598, RNSA 1998, Chicago -USA.

(Cameron, 1978)

Cameron, J.R.; Skofronick, J.G. - *Medical Physics* - New York, Wiley -1978.

(Chan, 1999)

Chan, H.P.; Sahiner, B.; Wagner, R. F.; Petrick, N.- "Classification Design For computer-aided diagnosis: effects of finite sample size on the mean performance of classical and neural network classifiers". *Med. Phys.*, v. 26, p. 2654-2668, 1999.

(Cheng 1997)

Cheng, J., Bell, D.A. e Liu, W. An Algorithm for Bayesian Belief Networks construction from data. In Proceedings of AI & STAT'97, pp. 83-90, Flórida, 1997.

(Cheng 1998)

Cheng, J., Bell, D. A. e Liu, W. Learning belief networks from data: An efficient approach based on information theory. Technical Report, 1998.

(Chow 1968)

Chow, C.K., and Liu, C.N. Approximating probability distributions with dependence trees. In: IEEE Transactions on Information Theory, vol.14, nr. 3, p. 462-468, 1968.

(Cooper, 1992)

Cooper, G. F. & Herskovits, E. - *A Bayesian Method For The Induction Of Probabilistic Networks from Data*. Machine Learning, 9, 309-347, 1992.

(Costa, 2001)

Costa, L.F.; César, R.M. - *Shape analysis and classification: theory and practice* - CRC Press, 2001.

(Cover et. al., 1991)

Cover, T. M. & Thomas, J. A.; *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

(Curry III, 1990)

Curry III, T.S.; Dowdey, J.E.; Murry Jr., R.C. *Christensen's Physics of Diagnostic Radiology*. Lea & Febiger, Philadelphia, 4th Edition 1990

(Daubechies, 1992)

Daubechies, I.; *Tem Lectures On Wavelets*, SIAM, Philadelphia, 1992.

(Deriche, 1990)

Deriche, R. - "Fast algorithms for low-level vision"- *IEEE Transactions on pattern Analysis and Machine Intelligence*, 1(12): 78-88, Jan. 1990.

(Doi, et. al., 1995)

Doi, K.; Giger, M. L.; Nishikawa, R.M.; Hoffman, K.R.; Macmahom, H.; Schmidt, R. A; Metz, C.E., "Recent Progress in Development of Computer-Aided Diagnosis (CAD) Schemes In Radiology". *Medical Image Technology*, v. 13, n.. 6, p. 822-35 , Nov. 1995

(Doi, 1996)

Doi, K. - "Perspectives on digital image analysis in medical image. Potencial usefulness of computer-aided diagnosis"- *Anais do III Fórum Nacional de Ciência e Tecnologia em Saúde*, v.2, p. 795-796, 1996.

(Dryden, I. L. et. al., 1993)

Dryden I. L. & K.V. Mardia. 1993. *Multivariate shape analysis*. Sankhya 55:460-480.

(Duda, et. al., 1973)

Duda, R.O; Hart, P.E.H. - *Pattern Classification and Scene Analysis* - John Wiley & Sons, New York, 1973.

(Elliott, 1913)

Elliott, E.B. - *An introduction to the Algebra of Quantics*, Chelsea Publishing Company - Bronx - New York, 1913

(Egan, 1980)

EGAN, R.L.; McSWEENEY, M.B.; SEWELL, C.W. (1980). *Intramammary calcifications without an associated mass in benign and malignant diseases*. Radiology, v.137(1), p. 1-7.

(Evans, 1981)

Evans, A.L. - *The evaluation of medical images*. Adam Hilger Ltd, Bristol, Great Britain, 1981.

(Escarpenati 2001)

ESCARPINATI, M. C.; SCHIABEL, H.; GOES, C. E. - "Interpretação e manipulação de imagens digitais visando ao processamento em mamografia" - XIII Congresso Argentino de Bioingenieria - Taffe del Valle, Tucuman (ARG) - 26-29/09/2001.

(Faugeras, et. al., 2002)

Faugeras, O & Keriven, R. - "Variational principles, Surfaces Evaluation, PDE's, Level set methods and the Stereo Problem" - 5th IEEE EMBS International Summer School on Biomedical Imaging - France 2002; IEEE Catalog Number: 02EX582.

(Ferreira, et. al., 2001)

Ferreira, C. B. R.; Borges, D. L. (2001). - "Automated Mammogram Classification Using a Multiresolution Pattern Recognition Approach" - XIV Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2001) . Florianopolis (Brazil), p.76-83, October 15-18.

(Friedman, 1996)

Friedman, J. (1996). On bias, variance, 0/1-loss, and the curse of dimensionality. Data Mining and Knowledge Discovery, 1.

(Friedman, 1996a)

Friedman, N. and Goldszmidt, M. (1996). Building classifiers using Bayesian networks. In Proceedings AAAI-96 Thirteenth National Conference on Artificial Intelligence, Portland, OR, pages 1277-1284. AAAI Press, Menlo Park, CA.

(Friedman, 1996b)

Friedman, N., Goldszmidt, M. Learning Bayesian Networks with Local Structure. Computer Science Division, University of California, Berkeley, CA, 1996.

(Friedman, 1997a)

Friedman, N., Geiger, D. e Goldszmidt, M. Bayesian Networks Classifiers. Machine Learning, 29, pp. 131-161, 1997.

(Fukunaga, 1990)

Fukunaga, K. *Introduction to Statistical Pattern Recognition*, Second Edition, 1990.

(Gomes, 1997)

Gomes, J.; Goldesntein, S.; Velho, L.; "Wavelets: Teoria Software e Aplicações". *21^o Colóquio Brasileiro de Matemática* IMPA – RJ 1997.

(Gonzales, 1992)

GONZALEZ, R.C.; WINTZ, P. *Digital Image Processing*. Addison-Wesley Publishing Company, United States of America, 1987

(Gose, 1996)

Gose, E.; et. al. - *Pattern Recognition and Image Analysis* - Prentice Hall, NJ, 1996

(Gram, 1997)

Gram, I.T.; Funkhouser, E.; Tabár, L. - "The Tabár classification of mammographic parenchymal patterns" - *European Journal of Radiology* v.24, p. 131-136. 1997

(Gur, 2004)

Gur, D. et. al. - "Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of Computer-Aided Detection System" - *Journal of National Cancer Institute*. Vol.96, No. 3, Oxford University Press, February 4, 2004.

(Gurevich, 1964)

Gurevich, G.B – *Foundations of the Theory of Algebraic Invariants* – Nordoff, Groningen, The Netherlands – 1964.

(Hall, 1979)

Hall, E. L. - *Computer Image Processing And Recognition* – Academic Press – NY - 1979

(Hadjiiski, 1999)

Hadjiiski, L.; Sahiner, B.; Chan, H.P.; Petrick, N.; Helvie, M. "Hybrid Unsupervised-supervised Approach for Computerized Classification of Malignant and Benign Masses on Mammograms", *Proceedings of SPIE* v.3661, p.464-473. 1999

(Haralick, 1973)

Haralick, R.M.; Shanmugan, K.; Dinstein, I.; "Textural features of images classification". *IEEE Trans. Systems Man. Cybernetics* SMC-3, p.610-621, 1973.

(Haralick, 1979)

Haralick, R.M. - "Statistical and Approaches to Texture" - *Proceedings of the IEEE*, v.67, n° 5, May, p.786-804, 1979

(Haykin, 1994)

Haykin, S., "Neural Networks – A Comprehensive Foundation", IEEE Computer Society Press, pp. 138-229, 1994.

(Hertz, 1991)

Hertz, J.; Krogh, A.; Palmer, R. G. - *Introduction to the Theory of Neural Computation* - Addison-Wesley Publishing Company, June, 1991

(Heckerman 1994)

Heckerman, D., Geiger, D., Chickering, D. M. Learning Bayesian networks: The combination of knowledge and statistical data. Technical Report MSRTR9409 (Revised), Microsoft Research, Advanced Technology Division, July 1994.

(Herskovits 1990)

Herskovits, E. H. e Cooper, G. F. Kutató: An entropydriven system for construction of probabilistic expert systems from databases. Piero Bonissone Ed., Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence, pp. 5462, Cambridge, Massachusetts, 1990.

(Heckerman 1995)

Heckerman, David. A Bayesian Approach to Learning Causal. Technical Report MSR-TR-95-04, Microsoft Research, March, 1995. Documento disponível em:

<http://www.research.microsoft.com/research/dtg/heckerma/TR-95-04.htm>.

(Heckerman 1995a)

Heckerman, D., Geiger, D., and Chickering, D. (1995a). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197--243.

(Heckerman 1995b)

Heckerman, D., Mandani, A., and Wellman M. Real-world applications of Bayesian networks: Introduction. *Communications of the ACM*, vol. 38, no.3 1995.

(Heckerman 1995c) Heckerman, D., Shachter, R. A definition and graphical representation for causality. In: Besnard and Hanks

(Heywang - Köbrunner, 1999)

Heywang - Köbrunner, S. H.; Schreer, I; Dershaw, D. D.; Frasson A. - *Mama - Diagnóstico por Imagem* - Editora Revinter (1999)

(Hu, 1961)

Hu, M. K. - "Pattern Recognition by Moments Invariants" – Proc. of the IRE, V. 49, N.9 - p. 1428, 1961.

(Hu, 1962)

Hu, M. K. - "Visual Pattern Recognition by Moments Invariants" - IRE Trans. Inf. Theory, V.8 - pp. 179-187 , 1962.

(Jain, 1995)

Jain, R. – *Machine Vision* – MIT Press & McGraw-Hill Inc. – 1995.

(Kahn, 1997)

Kahn, C.; Roberts, L.; Shaffer, K. A.; Haddawy, P., "Construction of a Bayesian Network for Mammographic Diagnosis of Breast Cancer", *Comput. Biol. Med.* Vol. 27, No. 1, pp. 19-29, 1997

(Kittler, 1997)

Kittler, J.; Hojjatoleslami, A. & Windeatt, T. - "Strategies for combining classifiers employing shared and distinct pattern representations". *Pattern Recognition Letter*, 18, 1373-1377, 1997.

(Klir, 1988)

Klir, G.J. & Folger, T.A - *Fuzzy sets, Uncertainty and Informations* - Prentice Hall, Englewood Cliffs - New Jersey, 1988.

(Knapp, 1996)

Knapp, R. B. - "Fuzzy Sets and Pattern Recognition" -
<http://www.engr.sjsu.edu/~Knapp/HCIFUZZY/fuzzy.html>.

(Kohonen, 1984)

Kohonen, T., *Self-Organization Formation and Association Memory*, first edition, Springer-Verlag, Berlin-Heidelberg, 1984.

(Kononenko, 2001)

Kononenko, I. - "Machine learning for medical diagnosis: history, state of the art and perspective" - *Artificial Intelligence in Medicine*. 23 (2001): 89-109.

(Kopans, 2000)

Kopans, D. B. - *Imagem da Mama* – 2ª edição Revinter, 2000.

(Kovacs, 1996)

Kovacs, Z. L.; - *Redes Neurais Artificiais: Fundamentos e Aplicações* - 2. Ed.
Acadêmica, São Paulo, 1996

(Koehler 1998)

Koehler, C. Uma Abordagem Probabilística para Sistemas Especialistas. Dissertação de Mestrado. Curso de Pós-Graduação em Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis-SC, (Brasil), 1998.

(Krause 1998)

Krause, P.J. Learning Probabilistic Networks. Philips Research Laboratories Tech. Report., 1998.

(Krzanowski, 1988)

Krzanowski, W. J. - *Principles of multivariate analysis: a user's perspective*. - Clarendon Press, Oxford - 1988.

(Lam, 1994)

Lam, W. & Bacchus, F. - "Learning Bayesian Belief Networks. An approach based on the MDL principle". *Computational Intelligence*, 10, 269-293, 1994.

(Lanchenbruch-1975)

Lanchenbruch, P. A. - *Discriminant Analysis* – New York: Hafner, 1975.

(Langley, 1992)

Langley, P.; Iba, W.; Thompson, K. - "An Analysis Of Bayesian Classifiers". In *Proceedings, Tenth National Conference on Artificial Intelligence* (pp. 223-228). Menlo Park, CA: AAAI Press, 1992.

(Lanyi, 1986)

Lanyi, M. - *Diagnosis and Differential Diagnosis of Breast Calcifications* - Röntgeninstitut Kaiserstrabe, German, 1986

(Le Gal, 1984)

Le Gal, M.; Chavanne, G.; Pelier, D. - "Valeur diagnostique des microcalcifications roudées écouvertes par mammographies." - *BULL Cancer* v.71, p.57-64 1984

(Lin, 1995)

Lin, C.T. & Lu, Y-C; "A Neural Fuzzy Systems with Linguistic Teaching Signals" - *IEEE Trans. On Fuzzy Systems*, vol 3, no 2, May, pp. 169-189, 1995.

(Madigan 1995)

Madigan, D., Garvin, J., and Raftery, A. (1995). Eliciting prior information to enhance the predictive performance of Bayesian graphical models. *Communications in Statistics: Theory and Methods*, 24:2271--2292.

(Madigan 1995a)

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63:215--232.

(Mallat, 1992)

Mallat, S.; Zhong, S.; "Characterization of Signal from Multiscale Edges" - *IEEE Transaction Pattern Anal. Mach. Intell.*, v.14, p.710-732. Julho 1992.

(Marr)

Marr, D. - *Vision* - Freeman, São Francisco, (1982).

(Martí, 1998)

MARTÍ, J.; CUFÍ, X.; REGINCÓS, J.; ESPAÑOL, J.; BARCELÓ, C. (1998). Shape-based feature selection for microcalcification evaluation. *Proc. SPIE: Image Processing*, v. 3338, p. 1215-1224

(Mascarenhas, 1989)

Mascarenhas, N. D.A.; Velasco, F.R. D. - *Processamento Digital de Imagens* - Editorial Kapelusz S.A., Buenos Aires, Argentina, 1989.

(Masters, 1994)

Masters, T. - *Signal and Image Processing with Neural Networks a C++* - Sourcebook, John Wiley & Sons, Inc, New York, 1994.

(Mc-Culloch, 1943)

Mc-Culloch, W. S. and PITTS, W.H. A logical calculus of ideas immanent in nervous activity. *Bull Math. Biophys*, pages 5:115-133, 1943

(Metz, 1986)

Metz, C.E. - "ROC Methodology in Radiologic Imaging" - *Investigate Radiology*, v.21,n.9, p. 720-33. Sep. 1986

(Meyer, 1993)

Meyer, Y. - *Wavelets: Algorithms and Applications* - Philadelphia, PA: SIAM., 1993.

(Minsky, 1969)

Minsky, M. L. And Papert, S. A - *Perceptrons* - MIT Press, Cambridge, Massachets, 1969.

(Mitchell, 1988)

Mitchell Jr., G. W.; Basset, L. W. - *Mastologia Prática* - Revinter, 1988.

(Mitchell, 1997)

Mitchell, T. M. - *Machine Learning* - WCB/McGraw-Hill, 1997.

(Monteiro, 1999)

Monteiro, L. R. et al. - *Princípios de Morfometria Geométrica* - Holos Editora - 1999

(Nunes, 2001)

NUNES, F. L. (2001). *Investigações em processamento de imagens mamográficas para auxílio ao diagnóstico de mamas densas*. 208p. Tese (Doutorado) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos.

(Patrocínio, 2000)

Patrocínio, A.C, et al. - "Investigation of clustered microcalcification features for an automated classifier as part of a mammography CAD Scheme" - Cd Rom Proc. of TheWorld Congress On Medical Physics And Biomedical Engineering, Chicago(USA), 4 pp, July, 23-28

(Patrocínio, 2002)

Ana Claudia Patrocínio, et al. - "Análise de atributos extraídos de nódulos mamográficos para identificação de formatos" - XVIII Congresso Brasileiro de Engenharia Biomédica, São José dos Campos (SP), 08-12/09/2002.

(Patterson, 1995)

PATTERSON, D.W. - *Artificial Neural Networks - Theory and Applications* - Prentice Hall, Singapore, 1995

(Poliker, 1996)

Polikar,R. <http://www.public.iastate.edu/~rpolikar/WAVELETS/waveletindex.html>.

(Pratt, 1991)

PRATT, W. K. *Digital Image Processing*. A Wiley-Interscience Publication, 1991

(Prokop, 1992)

PROKOP, R. J. & REEVES, A. P. - A Survey of Moment-Based Techniques for Unoccluded Object Representation and Recognition, - CVGIP: Graphical Models and Image Processing, V. 54, N.5 - pp. 438-460, 1992.

(Quadros, 2003)

Quadros, L. G. A. - "Classificação Bi-Rads. para laudos de mamografia" - *Revista Brasileira de Mastologia*, p 147, Vol. 13 - nº 4 - 2003

(Rissanen, 1978)

Rissanen, J. - *Modeling by shortest data description*. Automatica, 14, 465-471, 1978.

(Roseblatt, 1958)

Roseblatt, F. - "The Perceptron: A probabilistic model for information storage and organization in brain" - *Psychological Review*, pages 65:386-408, 1958

(Roseblatt, 1960)

Roseblatt, F. - "Perceptron simulation experiments" - *Proceedings of the Inst. of Radio Eng.*, p. 48: 301-309, 1960

(Rumelhart, 1986)

Rumelhart, D. E.; Williams, R.J. - *Learning internal representations by error propagation. Parallel distributed processing*, v 1, p.318-62, 1986

(Santaella, 2002)

César H. M. Santaella, et al. "Investigação de Atributos para Classificação de Nódulos em Mamogramas Digitalizados Utilizando Transformadas *Wavelets*" - XVIII Congresso Brasileiro de Engenharia Biomédica, São José dos Campos (SP), 08-12/09/2002

(Santos, 2002)

Santos, V.T; Schiabel, H.; Góes, C.E. "A Segmentation Technique to Detect Masses in Dense Breast Digitized Mammograms". *Journal of Digital Imaging*, Vol 15, Suppl. 1, p. 210-213, may, 2002.

(SBMFGO, 1998)

I Reunião de Consenso da Radiologia - Padronização dos laudos mamográficos Colégio Brasileiro de Radiologia, Sociedade Brasileira de Mastologia e Federação das Sociedades de Ginecologia e Obstetrícia; 1998.

(Schalkoff-1989)

Schalkoff, R. J.; *Digital Image Processing*. John Wiley & Sons, Inc., New York, 1989

(Schiabel, 2003)

SCHIABEL, H; MEDEIROS, R.B. (2003). A CAD scheme structured with a quality assurance program in mammography. In: WORLD CONGRESS ON MEDICAL PHYSICS AND BIOMEDICAL ENGINEERING. *Proceedings on CD-ROM*, Sydney, Australia, August 24-28.

(Schiabel, 2004)

SCHIABEL, H.; VIEIRA, M.A.C.; MARTINEZ, A.C. (2004) Um estudo sobre o processo de aquisição na formação e no processamento de imagens mamográficas. In: III LATIN AMERICAN CONGRESS ON BIOMEDICAL ENGINEERING. *Proceedings on CD-ROM*, João Pessoa, Brasil, Sept. 22-25.

(Shen et. al. , 1993)

Shen, L.; Rangayyan, R.M.; Desautels, J.E.L. - "Detection and classification of mammographic calcifications" - *International Journal of Pattern Recognition and Artificial Intelligence* v.7(6), p.1403-1416 1993

(Sickles, 1986)

Sickles, E. A. - "Breast calcifications: mammographic evaluation" - *Radiology* v.160, p.289-293 1986

(Strachan, 1990)

Strachan, N.J.C.L.; Nesvadba, P. & Allrn, A.R; - "Fish species recognition by shape analysis of images" – *Pattern Recognition*, V.23, N.5, pp. 539-54, 1990.

(Tao, 1998)

Tao, E. Y.; Ornes, C.; Sklansky, J. "Fully automatic detection of microcalcifications in digital mammography". *SPIE 98 - Image Processing*, v. 3338, p. 1450-1458, 1998.

(Toumoulin, 2002)

Toumoulin, C. et al. - "Diversity and Emergency in Biomédical Imaging" - *5th IEEE EMBS International Summer School on Biomedical Imaging* - France 2002; IEEE Catalog Number: 02EX582.

(Trosset, 2001)

Trosset, M. W. - *An Introduction to Statistical Inference and Data Analysis* - College of William & Mary, Williamsburg, VA 23187-8795, 2001.

(Vieira, 2003)

Vieira, M.A.C.; Escarpinati, M. C.; Schiabel, H.; Medeiros, R.B.; Alves, F.F.R. - "Avaliação de Qualidade em Equipamentos Mamográficos Hospitalares do Estado de São Paulo", *Anais em CD-ROM do VIII Congresso Brasileiro de Física Médica*, Porto Alegre (RS), maio 13-16, 2003.

(Wang, 1999)

Wang, X. H et. al. - "Computer-Assisted Diagnosis of Breast Cancer Using a Data-Driven Bayesian Belief Network" - *International Journal of Medical Informatics*. Pp 116-126, 54 – 1999.

(Wang, 2003)

Wang, X. H.; et. al.- "Automated Assessment of the Composition of Breast Tissue Revealed on Tissue- Thickness-Corrected Mammography" – *AJR*2003; 180:257-262.

(Widrow, 1960)

Widrow, B.; Hoff, M. E. - "Adaptive switching circuits" - *IRE WESCON Conversation Record*, p. 96-104, 1960.

(Wolbarst, 1993)

Wolbarst, Anthony Brinton; *Physics of Radiology*, Appleton & Lange, 1993.

(Wang, 1999a)

Wang, X. H.; Zheng, B.; Good, W.F., "Computer-Assisted Diagnosis Of Breast Cancer Using A Data-Driven Bayesian Belief Network", *International Journal of Medical Informatics* (54), pp. 115-126, 1999.

(Wang, 1999b)

Wang, X.H.; Zheng, B.; Chan, Y.H.; Good, W.F. - "Optimizing The Feature Set A Bayesian Network For Breast Cancer Diagnosis, By Using Genetic Algorithm Techniques". *Proc. Spie*, V. 3661, P. 1574-1580, 1999.

(Wang, 2003)

Wang, X. H.; et. al. - "Automated Assessment of the Composition of Breast Tissue Revealed on Tissue- Thickness-Corrected Mammography" – *AJR*2003; 180:257-262.

(Young, 1986)

Young, T.Y. & Fu, K. S. - *Handbook of Pattern Recognition and Image Processing* - Academic Press, USA, 1986.

(Zheng, 1995)

Zheng, B.; Chang, Y. H.; Gur, D. – " Adaptive Computer-Aided Diagnosis Scheme of Digitized Mammograms" – *Academic Radiology*, Vol.3, October 1996

(Zheng, 1997)

Zheng, B. et al; - " Adequacy Testing of Training Set Sample Sizes in the Development of a Computer-assisted Diagnosis Scheme" – *Acad Radiol* 1997; 4: 497-502

(Zheng, 1999)

Zheng, B.; et. al. – "Feature Selection for Computerized Mass Detection in Digitized Mammograms by Using a Genetic Algorithm" – *Acad. Radiology* 1999; 6: 327-332.

(Zheng, 2004)

Zheng, B.; et al. – " Computer- Aided Detection Schemes: The Effect of Limiting the Number Of Cued Regions in Each Case" – *ARJ*2004 - 182:579-583.

(Zhou, 2001)

Zhou, C.; et. al. - "Computerized image analysis: Estimation of Breast Density on Mammograms" - *Med. Phys.* Vol. 28 (6), pp. 1056-1069, June 2001