Campus de São Carlos

**ESQUEMA CLASSIFICADOR DE AGRUPAMENTOS DE MICROCALCIFICAÇÕES
MAMÁRIAS UTILIZANDO REDES NEURAIAS ARTIFICIAS**

Autora: Ana Cláudia Patrocínio

Orientador: Prof. Associado Homero Schiabel

UNIVERSIDADE DE SÃO PAULO

**ESCOLA DE ENGENHARIA
DE SÃO CARLOS**

**ESQUEMA CLASSIFICADOR DE
AGRUPAMENTOS DE MICROCALCIFICAÇÕES
MAMÁRIAS UTILIZANDO REDES NEURAIAS
ARTIFICIAIS**

DEDALUS - Acervo - EESC

31100017120

Ana Claudia Patrocínio

Serviço de Pós-Graduação EESC/USP

EXEMPLAR REVISADO

Data de entrada no Serviço: 19 / 12 / 00

Ass.: *[Signature]*

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

ORIENTADOR: Prof. Dr. Homero Schiabel

São Carlos

2000

Class.	TESE-EEEC
Cutt.	4183
Tombo	004101

st 1129171

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

P314e Patrocínio, Ana Claudia
Esquema classificador de agrupamentos de
microcalcificações mamárias utilizando redes neurais
artificiais / Ana Claudia Patrocínio. -- São Carlos,
2000.

Dissertação (Mestrado) -- Escola de Engenharia de
São Carlos-Universidade de São Paulo, 2000.
Área: Engenharia Elétrica.
Orientador: Prof. Dr. Homero Schiabel.

1. Classificação. 2. Esquemas CAD. 3. RNA.
4. Processamento de imagens. 5. Mamografia. I. Título.

FOLHA DE APROVAÇÃO

Candidata: Tecnóloga ANA CLAUDIA PATROCINIO

Dissertação defendida e aprovada em 01-11-2000
pela Comissão Julgadora:

Prof. Doutor **HOMERO SCHIABEL (Orientador)**
(Escola de Engenharia de São Carlos - Universidade de São Paulo)

Prof. Doutora **ROSELI APARECIDA FRANCELIN ROMERO**
(Instituto de Ciências Matemáticas e Computação - Universidade de São Paulo)

Prof. Doutora **ANA MARIA MARQUES DA SILVA**
(Universidade Federal de Santa Maria)

Prof. Associada **VILMA ALVES DE OLIVEIRA**
Coordenadora do Programa de Pós-Graduação
em Engenharia Elétrica

JOSÉ CARLOS A. CINTRA
Presidente da Comissão de Pós-Graduação da EESC

*Aos meus pais
Neusa e Guilherme*

“O Amor se descobre através da prática de amar e não das palavras”

Agradecimentos

Ao Professor Dr. Homero Schiabel pela competente orientação, incentivo e apoio, essências para a realização deste trabalho. Minha gratidão.

Aos meus pais, Neusa e Guilherme, pelo apoio financeiro, pelo respeito ao meu trabalho, por compreenderem os momentos de ausência e pelo amor que sentem por mim. *Amo vocês!*

Aos meus avós, Odila e Orlando, minha irmã Carla, meu cunhado Wilson pelo apoio, compreensão, carinho e amor. *Amo vocês também!*

Ao Professor Dr. João Fernando Marar pelo incentivo, pelo apoio e auxílios na aprendizagem de RNA, pelos conselhos de mestre e as lições de ética e profissionalismo. Meu Muito OBRIGADA de coração.

À minha amiga Dra. Evelina M. A Neves pela sua amizade e companheirismo, obrigada também pelas discussões sobre Momentos e dicas durante a implementação do trabalho.

Ao Rodrigo H. Benatti pelo coleguismo, amizade e grande auxílio durante os testes deste trabalho.

Ao pessoal do LADI: Maurício, Mogi, Michele, Rodrigo, Claudio, Vivian, César, Ricardo, e especialmente a Fátima pelas discussões, dicas e artigos. A todos pelo coleguismo, pelas horas de descontração e pela paciência comigo e com os as minhas "redes neuróticas". Meu Muito OBRIGADA!

Às minhas amigas do coração, Frederica B. Lorenzetti e Patrícia B. Grizotto pela nossa amizade, pelo incentivo, pelos compreenderem meus momentos de ausência, pelo carinho e apoio durante todos esses anos. Vocês são muito especiais.

À querida Dra. Berenice B. Lorenzetti, pelo incentivo e exemplos de profissionalismo, competência e amor a pesquisa.

Às minhas queridas amigas Patricia Laurenti, Luciana P. Costa, Viviane M. Poletti, Veridiana, Valéria e Mércia de Souza pelo carinho, respeito e compreensão.

À minha amiga e companheira de república Renata A T. Arantes, pela harmonia, companheirismo e tolerância com o meu despertador "tocando" insistentemente todas as manhãs.

A todos que de alguma maneira contribuíram para a realização deste trabalho.

À CAPES pelo apoio financeiro.

A DEUS, pela minha vida e minha capacidade de amar. Obrigada pela minha felicidade!

SUMÁRIO

LISTA DE FIGURAS	i
LISTA DE TABELAS	iv
Resumo	v
Abstract	vi
1 Introdução	1
1.1 O Câncer de Mama	1
1.2 Detecção e classificação do Câncer de Mama por sistemas computadorizados	3
1.3 Motivação	4
1.4 Objetivos	4
1.5 Disposição do Trabalho	5
2 Microcalcificações: Características Patológicas e Radiográficas	7
2.1 Visualização de Estruturas de Interesse em Mamogramas	7
2.1.1 A Mamografia e o Sistema Mamográfico	8
2.1.2 Resolução de Contraste e Resolução Espacial	12
2.2 Microcalcificações e o Câncer de Mama	13
2.2.1 Características Patológicas e Características Radiográficas	15
2.2.2 Critérios Utilizados no Diagnóstico Radiológico de “Clusters” de Microcalcificações	20
2.3 Conclusão	22
3 Reconhecimento de Padrões Aplicados em Sistemas CAD	24
3.1 Definição e Representação de uma Imagem Digital	24
3.2 O Sistema CAD	26
3.2.1 Curvas ROC	26
3.3 Reconhecimento de Padrões	29
3.3.1 Extração de Características	32
3.3.1.1 Descritores de Forma	33
3.3.1.2 Os Invariantes de Momentos de Hu	36
3.3.1.3 Interpretação Geométrica dos Momentos	41

3.3.1.4 Outros Atributos de Forma	45
3.3.2 Seleção de Características	46
3.3.3 Classificação	47
3.4 Conclusão	49
4 Redes Neurais Artificiais	50
4.1 Conceitos Básicos	50
4.2 Um modelo de RNA genérico	51
4.3 Neurônio de Mc-Culloch e Pitts	55
4.4 Redes Neurais Supervisionadas	56
4.5 O Perceptron	57
4.6 "Backpropagation"	59
4.6.1 As Fases do Algoritmo "Backpropagation"	61
4.6.2 A Matemática do "Backpropagation"	63
4.7 Conclusão	68
5 Sistemas Computadorizados de Classificação de Estruturas Mamográficas	70
5.1 Classificação de Estruturas Mamográficas Baseada em características extraídas com métodos de Visão Computacional	71
5.2 Utilização de Dados do Histórico do Paciente para a Classificação do Diagnóstico	87
5.3 Conclusão	94
6 Materias e Métodos	95
6.1 O Banco de Imagens	96
6.2 Sistema de Detecção de Microcalcificações	97
6.3 Esquema Classificador	100
6.3.1 Extração de Características	102
6.3.2 Seleção de Atributos	106
6.3.3 Organização dos Vetores de Entrada da RNA	110
6.3.4 Implementação da RNA	110
6.3.5 Treinamento da RNA	112
6.3.6 Classificação	113

7 Resultados e Discussões	115
7.1 Resultados Preliminares	115
7.2 Resultados do <i>Grupo de Imagens I</i>	118
7.3 Resultados do <i>Grupo de Imagens II</i>	124
7.4 Precisão e Erro obtidos com análise das curvas ROC	130
8 Conclusões	134
8.1 Atributos Extraídos dos Agrupamentos	134
8.2 Classificação por RNA	135
8.3 Conclusões Finais	137
8.4 Sugestões para Trabalhos Futuros	138
9 Bibliografia	139

LISTA DE FIGURAS

Figura 2.1 - Posições de exposição para exames mamográficos: a) incidência CC; b) incidência MLO	9
Figura 2.2 -Imagens mamográficas a) Mama mal posicionada e mal comprimida; b) Mama corretamente posicionada e corretamente comprimida	10
Figura 2.3 Ponto Focal Aparente	11
Figura 2.4- Sistema mamográfico	12
Figura 3.1 Passos que segue o reconhecimento de padrões	30
Figura 3.2 - Etapas de reconhecimento de padrões em um sistema CAD ([JIANG-98])	31
Figura 3.3 - Monômios de até terceira ordem que multiplicam a função $f(x,y)$ no cálculo dos momentos (Hall - 1978)	36
Figura 3.4 – Posição padrão para Invariância à Rotação ([MASTERS - 94])	39
Figura 3.5 - Processos de um sistema classificador ([FUKUNAGA-90])	48
Figura 4.1 - Gráficos das funções de ativação mais usadas para a modelagem de	53
Figura 4.2 Neurônio Biológico X Modelo Matemático	55
Figura 4.3 Perceptron	57
Figura 4.4 Rede "Feedforward" com uma camada escondida ("multilayer-perceptron")	63
Figura 4.5 Rede "Backpropagation"	65
Figura 5.1 - Esquema CAD proposto por ([SHEN-93])	72
Figura 5.2 - a)Esquema de classificação automático; b) detecção de "cluster" manual, respectivamente([JIANG-98])	78
Figura 5.3 - Sistema Computadorizado de Veldkamp ([VELDKAMP-99])	85
Figura 5.4 – quadro das características extraídas dos mamogramas ([LO JY-97])	88
Figura 5.5 - Esquema CAD híbrido ([ROGOVA-99])	92
Figura 6.1 - a) ROI de um mamograma digitalizado no <i>scanner</i> laser e b) ROI de um mamograma digitalizado no <i>scanner</i> ótico	97
Figura 6.2- Esquema do sistema computacional desenvolvido para detecção de "clusters" de microcalcificações([NUNES-97])	98
Figura 6.3 - Esquema utilizado para segmentação das microcalcificações([NUNES-97])	99
Figura 6.4 – a) ROI antes de ser processada com 256 níveis de cinza e b) ROI processada, imagem binária.	100

Figura 6.5 - Etapas do esquema classificador implementado	100
Figura 6.6 - Interface do procedimento de extração de características de Microcalcificações	103
Figura 6.7 – “cluster” delineado por um retângulo	103
Figura 6.8 -Interface do procedimento de extração de características de "clusters"	105
Figura 6.9 - Distribuição das áreas do "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens I.	107
Figura 6.10 - Distribuição das áreas do "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens II	107
Figura 6.11 - Distribuição dos Invariantes Momentos de Hu 1 dos "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens I	108
Figura 6.12 - Distribuição dos Invariantes Momentos de Hu 1 dos "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens II	108
Figura 6.13 - Distribuição das idades das pacientes das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens I	108
Figura 6.14 - Distribuição das idades das pacientes das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens II	109
Figura 6.15 - Tela do procedimento de treinamento	111
Figura 6.16 - Tela do procedimento de classificação	114
Figura 7.1 Desempenho das redes em função do número de neurônios	117
Figura 7.2 - Curvas ROC do classificador considerando apenas descritores de forma como atributo: a) treinado com 20 neurônios e b) treinado com 40 neurônios.	120
Figura 7.3 - Curvas ROC do classificador considerando descritores de forma e idade como atributos: a) treinado com 20 neurônios e b) treinado com 40 neurônios	121
Figura 7.4 - Curvas ROC do classificador considerando descritores de forma e "densificação" como atributos: a) treinado com 20 neurônios e b) treinado com 40 neurônios	122
Figura 7.5 - Curvas ROC do classificador considerando descritores de forma, idade e "densificação" como atributos: a) treinado com 20 neurônios e b) treinado com 40 neurônios	123
Figura 7.6 - Curvas ROC do classificador considerando apenas descritores de forma como atributos: a) treinado com 35 neurônios e b) treinado com 45 neurônios	126

	127
Figura 7.7 - Curvas ROC do classificador considerando descritores de forma e idade como atributos: a) treinado com 40 neurônios e b) treinado com 50 neurônios	
Figura 7.8 - Curvas ROC do classificador considerando descritores de forma e densificação como atributos: a) treinado com 35 neurônios e b) treinado com 45 neurônios	128
Figura 7.9 - Curvas ROC do classificador considerando descritores de forma e densificação como atributos: a) treinado com 40 neurônios e b) treinado com 50 neurônios	129
Figura 7.10 - Precisão dos resultados obtidos com o <i>Grupo de Imagens I</i> representado pelos pontos azuis e dos resultados obtido com o <i>Grupo de Imagens II</i> representado pelos pontos vermelhos	131
Figura 7.11 - Erros em função de A_z , representados pelos pontos azuis os resultados obtidos com o <i>Grupo de Imagens I</i> e pelos pontos vermelhos os resultados obtidos com <i>Grupo de Imagens II</i>	132
Figura 7.12 - Gráfico de desempenho do classificador para os quatro conjuntos de atributos com variações da Quantidade de neurônios para os exemplos do <i>Grupo de Imagens I</i> e <i>Grupo de Imagens II</i>	133

LISTA DE TABELAS

Tabela 1.1 - Estimativa do número de casos novos e da taxa bruta de incidência de câncer, segundo localização primária, sexo e distribuição percentual. Brasil – 1999	2
Tabela 1.1.1 – Mulheres	2
Tabela 1.1.2 - HOMENS	2
Tabela 2.1 - Estudos das características de microcalcificações	16
Tabela - 2.2 - Classificação de Michele Le Gal (1984)	18
Tabela 2.3 - Critérios práticos de avaliação de Classificação em diagnóstica radiológico	21
Tabela 3.1 – Interpretação geométrica dos Momentos Centrais de Segunda e Terceira Ordem	42
Tabela 5.1 - Escalonamento das características de microcalcificações usado em ([WU-93b])	74
Tabela 5.2 - Características extraídas de regiões de interesse para classificação ([TAO-98])	80
Tabela 5.3 - Características extraídas de regiões de interesse para classificação ([MARTÍ-98])	83
Tabela 5.4 - Dados de Entrada da RNA a partir do banco de dados mamográfico	90
Tabela 6.1 – Características extraídas dos “clusters”	105
Tabela 7.1- Resultados obtidos com o modo padrão de treinamento	116
Tabela 7.2 - Resultado obtido com treinamento "leave-one-out"	116
Tabela 7.3 - Índices obtidos na análise da curva ROC para o <i>Grupo de Imagens I</i>	118
Tabela 7.4 - Índices obtidos na análise com porcentagens para o <i>Grupo de Imagens I</i>	119
Tabela 7.5 - Índices obtidos na análise da curva ROC para o <i>Grupo de ImagensII</i>	124
Tabela 7.6 - Índices obtidos na análise com porcentagens para o <i>Grupo de Imagens II</i>	125

Resumo

Os Esquemas CAD ("Computer - Aided Diagnosis") têm mostrado bons resultados no auxílio ao diagnóstico precoce do câncer de mama. A classificação, nesses esquemas, é algo complexo e abrange investigações não só de técnicas computacionais, mas também das caracterizações citológicas dos achados de interesse clínico.

Por isso, o presente trabalho objetivou o desenvolvimento de um esquema classificador visando a indicação de cada caso como "suspeito" e "não-suspeito", com base em investigações de imagens mamográficas digitalizadas. Nessa investigação em particular, o foco de análise foram agrupamentos de microcalcificações detectadas por técnicas de processamento de imagens. A técnica de classificação utilizada no esquema baseou-se em redes neurais artificiais (RNA) supervisionadas, empregando algoritmo de aprendizagem "backpropagation".

O esquema classificador usando RNA, mostrou a eficiência dos descritores de forma na caracterização dos agrupamentos de microcalcificações e também a influência de atributos extraídos dos laudos das imagens como a idade e a "densificação". Os melhores resultados obtidos - apresentados aqui em forma de porcentagens e também de curvas ROC - mostraram 92% de acerto total com $A_z = 0,96$ aproximadamente, índices compatíveis aos dos melhores classificadores descritos pela literatura.

Abstract

Computer-aided diagnosis (CAD) schemes have shown good results in aiding the early diagnosis of breast cancer. In such schemes, the classification is usually complex and it uses not only computer techniques, but also cytological characterization of the clinical findings.

Thus, this work has aimed to develop a classifier scheme regarding to indicate each case as "suspected" or "non-suspected", based upon digitized mammographic images investigation. This analysis focus was clustered microcalcifications detected by image processing techniques. The classification technique used in the scheme was based on supervised artificial neural networks (ANN), with backpropagation as learning algorithm.

The classifier using ANN has shown the geometric descriptors efficiency for characterizing microcalcifications clusters as well as the influence of features extracted from images reports, as "age" and "density". The best data - presented here by percentage values and also by ROC curves - have shown 92% of correct results, with $A_z = 0,96$, which are comparable to the values from the best classifiers described by literature.

1

Introdução

1.1 O Câncer de Mama

O câncer de mama é provavelmente o mais temido entre as mulheres devido à sua alta frequência e sobretudo ao alto índice de mortalidade, sendo, conforme divulgado pelo Instituto Nacional do Câncer (INCA), o câncer que mais mata as mulheres no Brasil. Além disso, provoca grande impacto psicológico, pois envolve a percepção da sexualidade e a própria imagem pessoal, mais do que em qualquer outro tipo de câncer. A aparição da doença é relativamente rara antes dos 35 anos, mas a incidência cresce rapidamente depois desta idade.

Como ainda não são conhecidas as causas do câncer de mama, não existem métodos bem definidos de prevenção. O que se tenta fazer é uma proteção com o objetivo de diminuir o alto índice de mortalidade entre as mulheres de média e avançada idade. Então a única proteção efetiva é o diagnóstico precoce. Para isto, estão relacionadas pelo menos três estratégias, as quais entretanto estão sujeitas a debates com respeito à eficácia e ao custo/benefício de sua utilização. Estas estratégias são o auto-exame das mamas, o seu exame clínico e a mamografia, que é considerada o exame de maior sensibilidade na detecção precoce do câncer de mama.

Segundo dados fornecidos pelo INCA (vide tabela 1.1), o câncer de mama no Brasil mata seis mil mulheres a cada ano. Uma em cada 12 brasileiras poderá desenvolver a doença. A estimativa é que até o ano 2000 deverão ocorrer 35 mil

novos casos por ano. No Brasil, 75% dos casos diagnosticados estão em fase avançada da doença, com tumores grandes acima de 5cm. Tumores diagnosticados precocemente, porém, têm maiores chances de cura ou maior sobrevida.

Em um estudo publicado pelo *Journal of the National Cancer Institute* de jan/1998, pesquisadores da Ohio State University em Columbus (EUA) afirmaram que o estresse causado pelo diagnóstico e pelo tratamento cirúrgico do câncer de mama diminui a imunidade celular nas pacientes com a doença. Eles concluíram que a redução do estresse pode melhorar a resposta imune e levar a uma melhor recuperação nos pacientes com câncer.

Tabela 1.1 - Estimativa do número de casos novos e da taxa bruta de incidência de câncer, segundo localização primária, sexo e distribuição percentual. Brasil – 1999

Tabela 1.1.1- Mulheres

LOCALIZAÇÃO PRIMÁRIA	Nº DE CASOS NOVOS	TAXA BRUTA/100.000 MULHERES
Mama	31.200	39,58
Colo do Útero	20.652	26,28
Pele, não melanoma	16.800	20,65
Cólon e Reto	9.850	12,54
Estômago	6.750	8,55
Corpo do Útero	5.450	6,89
Pulmão	4.800	6,04
Boca	2.100	2,64
Pele, melanoma	2.050	2,50
Bexiga	1.800	2,28
Esôfago	1.700	2,14
Outras	31.250	40,70
Total	134.400	170,79

Tabela 1.1.2- HOMENS

LOCALIZAÇÃO PRIMÁRIA	Nº DE CASOS NOVOS	TAXA BRUTA/100.000 HOMENS
Pele, não melanoma	19.150	25,00
Pulmão	14.800	18,89
Próstata	14.500	18,49
Estômago	13.600	17,34
Cólon e Reto	9.200	11,66
Boca	5.850	7,52
Esôfago	5.300	6,75
Pele, melanoma	1.800	2,45
Outras	37.550	47,21
Total	127.500	162

As microcalcificações são grandes indicadoras do câncer de mama, como pode ser notado na literatura ([LANYI-86], [MITCHELL-88], [CANELLA-99], [NAGEL-98], [SHEN-93, 94], [CHAN-96], [TAO-98], [EGAN-81], [SICKLES-82 e 86], entre outros artigos), embora muitas delas sejam benignas. A classificação é bastante complicada para os médicos. Por isso ultimamente as pesquisas estão atentando muito para esta área.

A mamografia é considerada o exame mais eficiente de rastreamento do câncer de mama [MITCHELL-88], pois sua sensibilidade é alta. Através da mamografia é possível detectar microcalcificações, estruturas consideradas não palpáveis, geralmente da ordem de 0,5 mm de diâmetro, o que dificulta a visualização dos médicos.

Qualquer fator que coloque o médico em dúvida da presença ou não de uma microcalcificação ou de sua patologia (tumor maligno) faz com que a paciente seja encaminhada para uma biópsia, o que é traumatizante, pois além de dolorido, é um exame invasivo. E caso seja constatado que a paciente realmente não tem a doença, o que acontece com frequência, o gasto envolvido neste erro de diagnóstico, além de alto, é desnecessário.

1.2 Detecção e classificação do Câncer de Mama por sistemas computadorizados

Com intuito de diminuir a mortalidade de mulheres em todo o mundo vítimas do câncer de mama, sistemas computadorizados de auxílio ao diagnóstico estão sendo desenvolvidos por muitos grupos de pesquisa. Estes sistemas auxiliam não apenas na detecção de lesões e microcalcificações, mas também na sua classificação.

A classificação pode objetivar a divisão em estruturas benigna ou maligna, ou apenas em "suspeita" ou "não-suspeita". Estes esquemas são vistos como uma segunda opinião ao médico na tomada de decisão no momento do diagnóstico.

Muitas técnicas computacionais e ferramentas matemáticas são empregadas na construção destes esquemas, tais como Visão Computacional, processamento de imagens, digitalização de imagens, reconhecimento de padrões, Redes Neurais Artificiais. Elas geralmente são combinadas a fim de modelar o raciocínio dos especialistas na busca de assisti-los em consultórios e hospitais. Além disso, a qualidade da imagem deve ser considerada, através dos conceitos e parâmetros utilizados em controle de qualidade dos equipamentos, desde sua formação até a digitalização e as características do próprio digitalizador.

1.3 Motivação

Os erros nos diagnósticos do câncer são freqüentes, devidos às dificuldades de identificação e classificação dos achados mamográficos. Isso é causado muitas vezes pelo tamanho reduzido das estruturas de interesse e pelas várias limitações das imagens e dos próprios equipamentos mamográficos. Com os erros, o aumento no número de pacientes submetidas à biópsia cresce muito, visto que este exame é a única forma de comprovação efetiva do diagnóstico.

Um diagnóstico preciso depende de muitos fatores, sendo um deles, e talvez o mais importante, a experiência do especialista. Somente nos Estados Unidos, o custo destes erros é de aproximadamente 250 milhões de dólares por ano em biópsias de microcalcificações benignas ([ROGOVA-99]). No Brasil não é conhecida esta estimativa, mas acredita-se que não deva ser muito diferente.

1.4 Objetivos

Dentro das linhas de pesquisa do LADI - Laboratório de Análise e Digitalização de Imagens - uma ênfase particular têm sido dada ao emprego de

técnicas de processamento de imagens para desenvolver esquemas computadorizados como coadjuvantes no diagnóstico de lesões mamárias a partir da mamografia. Trabalhos anteriores já levaram à elaboração de um esquema automático de detecção de agrupamentos de microcalcificações ([NUNES-97]), que podem ser de utilidade no direcionamento do diagnóstico final.

A partir do estudo dos diversos modelos de esquemas computadorizados para auxílio ao diagnóstico precoce do câncer de mama, a construção de um esquema classificador para avaliar agrupamentos de microcalcificações, revelou-se uma necessidade e sobre seu desenvolvimento é que versa o presente trabalho. Esses agrupamentos, portanto, devem ser detectados pelo esquema de processamento previamente desenvolvido ([NUNES-97]) e o classificador deverá investigá-los por meio da extração de atributos através de descritores matemáticos.

Como na literatura muitos trabalhos estão usando como técnica de classificação Redes Neurais Artificiais (RNA), o nosso propósito é estudar as RNA supervisionadas a fim de desenvolver um classificador neural para a predição de casos em que foram detectados agrupamentos de microcalcificações como "suspeitos" e "não-suspeitos" com a finalidade de auxiliar na redução dos erros ocorridos no diagnóstico do câncer de mama.

As análises dos resultados obtidos com o classificador deverão basear-se na comparação com dados do laudo médico, os quais também servirão para elaboração das curvas ROC.

1.5 Disposição do Trabalho

Além dessa introdução, os seguintes capítulos são apresentados:

- 2. Microcalcificações: Características Patológicas e Radiográficas** - Características de microcalcificações visualizadas em mamogramas e as diferenças entre malignas e benignas abordadas na literatura, além dos conceitos de formação das imagens.

3. **Reconhecimentos de Padrões Aplicados em Sistemas CAD** - definição de sistemas CAD e como são utilizadas as técnicas computacionais nos esquemas de auxílio ao diagnóstico médico.
4. **Redes Neurais Artificiais** - Conceitos básicos da ferramenta e algoritmos das redes supervisionadas que esclarecem e validam a eficiência da técnica.
5. **Sistemas Computadorizados de Classificação de Estruturas Mamográficas** - Estado da arte dos sistemas de classificação computadorizados.
6. **Materiais e Métodos** - Descrição dos procedimentos utilizados no desenvolvimento do esquema classificador implementado nesse trabalho.
7. **Resultados e Discussões** - Disposição e discussões dos resultados obtidos com o classificador.
8. **Conclusões** - Apresentação das conclusões desse trabalho e algumas sugestões para trabalhos futuros.
9. **Bibliografia** - Referências bibliográficas.

2

Microcalcificações: Características Patológicas e Radiográficas

As microcalcificações são artefatos de grande valor para a identificação precoce do câncer de mama ([MITCHELL-88]). Ao serem detectadas, são analisadas com muito cuidado, pois seu tamanho reduzido dificulta o reconhecimento de características que possibilitam classificá-las corretamente quanto às suas origens, embora o simples fato de elas estarem presentes num mamograma, na maioria dos casos, já é suspeito.

2.1 Visualização de Estruturas de Interesse em Mamogramas

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 de pacientes mais jovens ¹: esse tipo de mama ainda é um problema para o diagnóstico precoce do câncer, em razão da pequena variação de contraste entre os tecidos envolvidos.

A visualização 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

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

qualidade da imagem deve ser boa. O aparelho mamográfico é um dos maiores responsáveis pela boa visualização dos médicos, pois o equipamento deve estar corretamente calibrado. Uma imagem é considerada de boa qualidade quando se consegue a melhor resolução espacial em conjunto com a melhor resolução de contraste possível. Os filmes 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 das microcalcificações ([SHEN-93]).

2.1.1 A Mamografia e o Sistema Mamográfico

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 posicionamentos, 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 centro do campo de exposição ([MITCHELL-88]).

Os sistemas mamográficos têm um acessório chamado de compartimento de compressão, que tem uma função importante. A mama é comprimida por uma placa de um material radiotransparente até que se consiga a menor espessura possível. Na figura 2.1 é possível identificar esse compartimento.

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

A compressão da mama tem três finalidades principais que muito contribuem para a qualidade da imagem. Uma é que, com menor espessura, diminui a radiação espalhada e assim o borramento; outra é que imobiliza a mama e assim dificulta o movimento no momento em que a imagem está sendo formada e, por último, a compressão possibilita a região de interesse ficar mais próxima do detector de imagem (chassi com o filme) e evita a sobreposição de tecidos quando feita de maneira adequada. Na figura 2.2, é possível observar uma imagem de uma mama que não foi corretamente posicionada.

Figura 2.2 -Imagens mamográficas a) Mama mal posicionada e mal comprimida; b) Mama corretamente posicionada e corretamente comprimida

O equipamento dedicado à mamografia é especial, pois a imagem deve ser de alta resolução para que se possam visualizar 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 de raios-X convencional. Os sistemas de raios-X convencionais geralmente usam tubo de tungstênio. O do equipamento mamográfico, porém, é de molibdênio, dadas suas características atômicas. O feixe produzido num tubo de molibdênio é menos absorvido do que um feixe produzido num tubo 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, que são uma espécie de direcionadores do feixe de raios-X e barradores de radiação. Isso ajuda a diminuir a dose de radiação ionizante em outras partes do corpo do paciente e também colabora com a melhoria da imagem. Os filtros impedem que os fótons do feixe que atrapalham na formação da imagem e nada acrescentam para o diagnóstico, sejam espalhados, além de atingirem o 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 de visualizar pequenas estruturas de até 0,3 mm de diâmetro. A figura 2.3 mostra o ponto focal aparente, que é o tamanho que está sendo "visto" pelo plano imagem.

Figura 2.3 Ponto Focal Aparente

O chassi mamográfico apresenta um écran intensificador que, ao contrário do convencional, se posiciona embaixo do filme, como pode ser visualizado na figura 2.4. Os fótons atravessam o filme, chegando pela sua base, atingem o écran, transformam-se em luz visível e refletidos de volta impressionando o filme. Esse posicionamento evita o efeito "crossover" e ajuda na obtenção de uma melhor resolução da imagem. O efeito "crossover" é o fenômeno do filme ser impressionado duas vezes pelo mesmo fóton, de modo que isso causa uma certa penumbra na

imagem, deteriorando a resolução. O que acontece no "crossover" é que o fóton de raios-X, ao interagir com o écran intensificador e se transformar em luz visível, impressiona o filme (quando de dupla emulsão) na primeira camada com emulsão que encontra e pode atravessar e impressionar a camada inferior novamente, provocando a penumbra ([WOLBASRST-93]).

Figura 2.4- Sistema mamográfico

O posicionamento do écran intensificador sob o filme no sistema mamográfico também tem um outro objetivo além de evitar o "crossover", que é o de prevenir 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.

2.1.2 Resolução de Contraste e Resolução Espacial

O contraste 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.

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 anodo 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 ([CURRYIII-90]). 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 e densidades ópticas ([CAMERON-78]).

Além do contraste, a resolução espacial da imagem é um fator de influência para a visualização de detalhes, mesmo se ele apresentar um tamanho suficiente e bom contraste em termos de absorção. A nitidez depende do tamanho do ponto focal, da distância foco-filme e filme-objeto e da resolução do sistema de registro da imagem (geralmente filme).

O sistema de imagem contém ruídos estruturais que podem obscurecer os detalhes de interesse. Os grãos do filme e os fótons de raios-X influenciam significativamente no ruído. O sistema écran-filme é particularmente propenso a ter um fundo ruidoso causado pela diferença entre os tamanhos de grãos dos filmes, radiação inerente (o "quantum") e o ruído estrutural dos écrans intensificadores.

2.2 Microcalcificações e o Câncer de Mama

As calcificações são depósitos de cálcio, que podem ser visualizadas 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 dessas calcificações, torna-se mais fácil identificar a sua procedência ([NÄPPI-99]). Já as microcalcificações são consideradas estruturas menores ou iguais a 0,5mm. Frequentemente são os únicos 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.

Uma das dificuldades a serem superadas na área é a ausência de padronização de conceitos. As definições a respeito de calcificações e microcalcificações e outros termos utilizados pelos especialistas apresentam grande variação na literatura. Em 1980, R.L.Egan *et al.*, ([EGAN-80]), depois de pesquisarem 42.888 estudos clínicos e radiográficos, concluíram que as calcificações apresentavam tamanho variando de 2 a 3 mm. Em 1982, E. A. Sickles ([SICKLES-82]) fez um estudo da detectabilidade de microcalcificações utilizando diversas técnicas de formação de imagens. Nesse estudo, as microcalcificações são simuladas e seus tamanhos estão no intervalo de 0,15 a 0,55 mm. Segundo pesquisa do mesmo autor, descrita em ([SICKLES-86]), as calcificações são as menores estruturas identificadas em um mamograma e são vistas em cânceres de mama como "clusters" (agrupamentos) de minúsculas partículas (menores que 0,5 mm), apresentando formas lineares, curvilíneas ou ramificações.

Em ([LANYI-86]), afirma-se que grande parte de ocorrências clínicas de carcinomas de mama (43% a 49%) é detectada pela presença de microcalcificações, sendo 21% delas menores que 0,25 mm de diâmetro; algumas não são maiores que 0,1mm.

O diagnóstico correto não depende apenas da percepção destas pequenas partículas de cálcio, mas também de interpretações. Somente por uma análise meticulosa dos formatos individuais da cada calcificação e seus "clusters" é possível diferenciar casos malignos ou suspeitos, ou ainda benignos. Portanto, a qualidade da imagem é um fator crítico importantíssimo nos diagnósticos de diferenciações de microcalcificações mamárias.

2.2.1 Características Patológicas e Características Radiográficas

De acordo com Lanyi ([LANYI-86]), 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 feridas (lesões) inclusive mecânicas. As calcificações de doenças tipicamente malignas geralmente são múltiplas e são uniformemente distribuídas no tecido; elas são granulares e têm uma intensidade relativamente baixa; formam-se dentro de dutos lactares dilatados que contêm resíduos celulares.

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:

- microcalcificações de 0,1 a 0,3 mm de diâmetro;
- pouca intensidade;
- formatos irregulares e densidades não uniformes;
- agrupamentos ovais de 2 a 5 mm de diâmetro.

Em seu livro, Lanyi ([LANYI-86]) considera que quanto maior a microcalcificação, menor a suspeita de malignidade. Por outro lado, "clusters" de microcalcificações são freqüentemente o primeiro sinal do câncer de mama visível em mamogramas.

Ele mostra um levantamento da literatura com as principais publicações sobre microcalcificações de mama. Na tabela abaixo está parte deste estudo representado.

Tabela 2.1 Estudos das características de microcalcificações

Autor	Número			Localiza- ção histológi- ca	Especificações finas de diagnósticos diferenciais	Número de neoplasias lobulares (LCIS) incluídas em carcino- mas	Sinais patognomônicos de carcinomas
Colbassini et al. (1982)	15 ^b	40 ^b	55 ^b	Sim	Não especificado	Não especifica do	Polimorfismo; variabilidade de formatos não é típico de carcinoma. Número de microcalcificações por cluster e por cm ² é maior em carcinomas. Não há diferenças significantes entre formatos de "clusters" malignos e benignos. Formatos de "clusters" e a distribuição não foram analisados.
Egan et al. (1980)	115	353	468	Não especifica do	Não especificado	13	Muitas microcalcificações em cluster, variando de muito finas até grosseiras, densidade variável de imperceptível a muito densas, indício em radiografia, provém da presença de carcinomas. Há sobreposição de dependendo da posição de exposição de calcificações em doenças fibrocísticas e carcinomas que não podem ser identificados apenas com mamografia.
Frichbier and Lohbeck (1977)	Não especificado			Não Especifica do	Não especificado	Não especifica do	Grande número de microcalcificações, alto risco de malignidade. Variam de muito pequenas até calcificações grosseiramente visíveis. Polimórficas; de formatos bizarros e irregulares, alta suspeita de malignidade. Configuração irregular é suspeita.

Hoeffken and Lanyi (1973)	Não especificado			Não especificado	Para lesões benignas	Não especificado	Densa; cristalina; granular muito fina; angular ou bizarra. Como estilhaços de pedras
Ingbleby and Gerson-Cohen (1960)	64	Não especificado		Sim	Para lesões benignas	Excluído	Agrupadas, finas, tendem a ser lineares como pontas de agulhas quebradas; grandes calcificações amorfas também ocorrem, mas não são específicas
Lanyi (1977)	60	111	171	Sim	Não especificado	Excluído	Tamanho, número, extensão, densidade, contorno, não tem muito significado para o diagnóstico. Microcalcificações são polimorfas (puntiformes, lineares, formatos irregulares, como poeiras), raramente isoladas. Densamente arranjadas, usualmente agrupadas em formas triangular, trapezoidal, ou em forma de roseta.
Leborgne (1951)	Não especificado			Sim	Para lesões benignas	Não especificado	Agrupadas em clusters; arranjadas de maneira convergente; inumeráveis; puntiformes, parecidas com grãos de sal.
Menges et al. (1973)	13	54	67	Sim	Para lesões benignas	Não especificado	Grande número de microcalcificações, provável carcinoma. Mais do que 5, justifica biópsia. Arranjadas linearmente ou em "clusters".
Millis et al. (1976)	33	27	60	Sim	Para lesões malignas	1	A relação de calcificações para outras características radiológicas, número e distribuição foram estudadas; densidade e formato não foram considerados como características que ajudam a distinguir entre maligno e benigno. Microcalcificações malignas são indistinguíveis das benignas.

Moskowits (1979)	Não especificado	Não especifi- cado	Não especificado	Não especifi- cado	O risco depende do formato das microcalcificações. Particularmente é alto para as ramificadas e lineares.
---------------------	------------------	--------------------------	------------------	--------------------------	--

Segundo trabalhos publicados por Michèle Le Gal ([LE GAL-84]), 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), forma destas microcalcificações, densidade e distribuição.

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

Tabela - 2.2 Classificação de Michele Le Gal (1984)

Tipo	Microcalcificações - Morfologia	% de Malignidade
Tipo I	Anclares, 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

Ainda em ([LANYI-86]) é mostrado que as calcificações ocorridas nos lóbulos e sistema ductal correspondem a 26,3% de todas as calcificações intramamárias encontradas na prática clínica. Num estudo de 1144 microcalcificações encaminhadas para biópsia, 24,5% (281 de 1144 biópsias) foram registradas nos sistemas lobular e ductal. A lesão benigna ou maligna quase sempre representa formas de cavidades nas quais ela ocorre, e isso determina seus formatos e arranjos. Então fica claro que um bom conhecimento de anatomia de mama, patologia e anatomia radiográfica é necessário para estabelecer a localização das microcalcificações e fazer um diagnóstico mais preciso, a fim de diminuir as ocorrências de diagnóstico falso-positivo e falso-negativo. Lanyi também conclui que na análise das microcalcificações, há três fatores característicos fundamentais para se chegar a um diagnóstico mais preciso:

- formato dos "clusters" das microcalcificações;
- formato das microcalcificações individuais;

- número de microcalcificações por "cluster" (informação relevante).

Uma quarta característica é a intensidade (densidade) das microcalcificações na imagem radiográfica. Na prática, a análise dessa característica não é usual porque isso requereria um microdensitômetro para avaliação precisa. Também a intensidade da microcalcificação depende da técnica radiográfica empregada (espessura do filme, temperatura de revelação, corrente no tubo, energia do feixe) e distribuição do tecido da mama (mais gordurosa ou mais densa).

Segundo Lanyi ([LANYI-86]), a presença de pelo menos cinco microcalcificações em um "cluster" já é considerado um caso suspeito.

Já Edward A Sickles em ([SICKLES-86]) considera também que não pode ser descartada a suspeita em casos de "cluster" com três ou quatro microcalcificações em 1cm^2 principalmente se elas forem menores ou aproximadamente iguais a 0,5 mm. Uma maneira de avaliação proposta neste seu trabalho é primeiramente identificar as estruturas com características benignas (calcificações com centros brilhantes, calcificações arteriais, calcificações pós-cirúrgicas, calcificações de fibroadenomas arredondadas) e, não sendo identificadas estas características, a estrutura deve ser considerada suspeita.

Em nosso sistema, consideraremos "cluster" como um agrupamento com pelo menos 3 microcalcificações numa área de 1cm^2 ([NUNES-97]).

Com objetivo de classificar as calcificações e considerando as dificuldades encontradas na realização dessa tarefa, foram feitos alguns estudos abrangendo suas origens, formatos, posição e agrupamento ([MITCHELL-88]). Na prática, porém, algumas características são de difícil identificação por parte dos especialistas, o que justifica o alto número de casos enviados para biópsia, a maneira mais efetiva de comprovação do diagnóstico nos casos suspeitos.

Uma primeira suspeita ocorre com a simples presença de microcalcificações em um mamograma. A princípio considera-se o formato individual de cada microcalcificação e a quantidade delas dentro de uma determinada área avaliada, que geralmente é de aproximadamente 1cm^2 no plano radiográfico, correspondente a 1cm^3 do tecido mamário, mas esta área pode variar assim como o número de microcalcificações para que o achado possa ser considerado "cluster" ([LANYI-86],

[LE-GAL-84], [Sickles-86], entre outros trabalhos que serão citados no capítulo 5). O agrupamento de microcalcificações já é um forte indício de problemas mais sérios, mesmo que ele ainda seja pequeno e com poucas microcalcificações.

Quando identificados esses "clusters", parte-se então para um rastreamento detalhado na tentativa de identificar características de formato de microcalcificações junto com uma investigação do histórico clínico da paciente, cujas informações podem servir como ponto decisivo para o especialista na tomada da decisão de encaminhar para o exame cito-patológico, ou não. Um dos fatores de peso é a investigação de casos de carcinomas em antecedentes familiares da paciente, principalmente de primeiro grau.

A idade também é um dado importante, como pode ser comprovado em ([Lo JY-99]), cujo trabalho resultou num sistema computacional de classificação de microcalcificações, utilizando, além dos atributos de forma, informações do histórico da paciente. Entre essas informações, os autores concluíram que a idade é o fator mais importante para a classificação feita por sistemas computacionais.

2.2.2 Critérios Utilizados no Diagnóstico Radiológico de “Clusters” de Microcalcificações

Segundo a Dra. Ellyette Canella, médica do Instituto Nacional do Câncer e da Santa Casa de Misericórdia e do LABS, Rio de Janeiro ([CANELLA-99]), microcalcificações são encontradas em 42% dos casos de lesões não palpáveis e podem representar o sinal mais precoce de malignidade. Com esta ocorrência, a análise das microcalcificações deve ser feita de maneira consistente e exige conhecimento, critérios e rotinas, como será descrito a seguir.

A análise deve ser feita considerando:

- tamanho da microcalcificação, adotando-se o critério de que quanto menor, mais sugere malignidade, seguindo as medidas já citadas para quantificar estes valores;

- Número, de forma que quanto mais microcalcificações dentro de 1cm^2 , maior a suspeita de malignidade;
- Forma, cujo critério é baseado na classificação morfológica de Michèle Le Gal, como mostra a tabela 2.2 da sessão anterior;
- Densidade, as microcalcificações malignas apresentam variação significativa de densidade entre as partículas;
- Distribuição, as microcalcificações malignas podem estar dispostas em trajetos de ductos ou agrupadas em um pequeno segmento da mama, geralmente são unilaterais.

A partir das informações acima, na prática, o diagnóstico radiológico pode ser auxiliado por uma segunda tabela, na qual cada item de análise está associado a uma pontuação. Essa pontuação foi criada com base nos valores citados por Le Gal na sua classificação morfológica.

Tabela - 2.3 Critérios práticos de avaliação de Classificação em diagnóstica radiológico

Característica	Pontuação 1	Pontuação 2	Pontuação 3
Tamanho	> 1mm	> 0,5mm	< 0,5mm
Número	Até 4 em 1cm^2	5 em 1cm^2	> 5 em 1cm^2
Forma	Monofórmicas Tipo I da Tabela 2.2	Tipo II e III da Tabela 2.2	Tipo IV e V da tabela 2.2
Densidade	Isodensas	Pouca variação de densidade	importante variação de densidade
Distribuição	Esparsas	Agrupadas	Trajeto ductal

Os itens analisados obedecem às respectivas pontuações da tabela 2.2, as quais vão somando-se a cada característica analisada. O critério de avaliação se faz da seguinte maneira: até 5 pontos indica microcalcificação provavelmente benigna; a partir daí quanto maior a pontuação, maior o grau de suspeição - de 6 a 12 pontos a microcalcificação passa a ser considerada suspeita, e de 13 a 15 altamente suspeita ([CANELLA-99]).

É um critério que vem sendo utilizado pelos radiologistas brasileiros, segundo a Sociedade Brasileira de Radiologistas.

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 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. Seu objetivo principal é diminuir a confusão na interpretação da imagem mamográfica, facilitando também assim o resultado final do exame. Em setembro de 1995, a segunda edição do BI-RADS® foi lançada com algumas modificações, que a prática diária revelou.

O BI-RADS® classifica as imagens mamográficas nas seguintes categorias:

- Categoria I - mamografia normal;
- Categoria II - achados mamográficos benignos;
- Categoria III - achados mamográficos provavelmente benignos;
- Categoria IV - achados mamográficos suspeitos;
- Categoria V - achados mamográficos altamente suspeitos.

2.3 Conclusão

Na prática, algumas características são de difícil identificação por parte dos especialistas, o que justifica o alto número de casos enviados para biópsia. A maior dificuldade na interpretação das imagens radiológicas é o baixo contraste e o tamanho geralmente reduzido das características de malignidade no estágio inicial. Muitos erros ocorrem no diagnóstico do câncer de mama, principalmente em relação ao diagnóstico precoce, que aumenta sensivelmente as chances de cura.

Um dos sinais mais importantes do câncer de mama é a presença de "clusters" de microcalcificações e aparecem como um pequeno grupo de objetos granulados.

Com objetivo de classificar as calcificações e considerando as dificuldades encontradas na realização desta tarefa, muitos estudos procuram buscar cada vez mais informações e desenvolver técnicas de avaliação considerando suas origens, formatos, posição, agrupamento, entre outras, para auxiliar os especialistas em tomadas de decisões no diagnóstico a fim de que ele seja mais preciso e com uma relação custo-benefício melhor.

3

Reconhecimento de Padrões Aplicados em Sistemas CAD

Com o intuito de auxiliar na árdua tarefa da avaliação e análise de estruturas visualizadas em mamografias, trabalhos vêm sendo desenvolvidos utilizando sistemas computacionais que empregam técnicas de processamento de imagens e reconhecimento de padrões, com o objetivo de detectar precocemente o câncer de mama. Estes esquemas são os chamados CAD (do inglês "Computer Aided Diagnosis").

3.1 Definição e Representação de uma Imagem Digital

Num CAD as imagens são digitais. Geralmente os filmes são digitalizados com equipamentos de alta resolução e seguem certos critérios com as técnicas de digitalização. A imagem digitalizada é constituída por uma unidade básica que é o pixel ("picture element"), deve ter uma resolução espacial boa, geralmente usa-se de 150 a 600 DPI (pontos por polegada), que indica o tamanho de cada pixel da imagem. Além disso uma resolução de contraste com uma quantidade de níveis de cinza suficiente deve ser utilizada a fim de representar as diferenças entre os tecidos e estruturas presentes na imagem ([DOI-95]).

As imagens constituem uma distribuição espacial de irradiância em um plano. Matematicamente falando, a distribuição espacial de irradiância pode ser descrita como uma função contínua de duas variáveis espaciais.

A imagem é representada matematicamente como uma função bidimensional $f(x,y)$, que pode ser descrita como uma função da intensidade luminosa refletida do objeto, onde x e y são as coordenadas espaciais. O valor de $f(x,y)$ é a intensidade luminosa, ou brilho, da imagem do ponto localizado nas coordenadas x e y ([SCHALKOFF-89]). Assim, uma imagem pode ser representada por uma matriz de pontos:

$$f(x,y) \approx \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,n-1) \\ f(1,0) & f(1,1) & \dots & f(1,n-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(m-1,0) & f(m-1,1) & \dots & f(m-1,n-1) \end{bmatrix}$$

A série de valores de f é chamada de escala de cinza ou cadeia de níveis de cinza. Se f puder assumir apenas dois valores, a imagem é dita binária. A quantidade de f relata propriedades radiométricas da imagem inicial.

Já as imagens em escala de cinza possuem mais do que apenas dois valores de intensidade. As imagens radiográficas são digitalizadas em escala de cinza, comumente em 256 níveis (correspondente a 8 bits), embora alguns autores trabalhem com 1024 níveis (10bits) ou ainda 4096 níveis (12 bits), onde a cor preta é representada pelo valor 0 e o branco pelo valor máximo da escala em questão

Até agora falamos apenas da resolução de níveis de cinza que é responsável, entre outras, pelo contraste. Mas esta deve estar combinada com a resolução espacial, segundo Gonzales em ([GONZALES-87]), que, na prática, é dada pelo número de pontos por polegada que a imagem em questão possui, quantos $f(x,y)$ estão contidos numa área de 1 polegada quadrada do plano da imagem(DPI).

3.2 O Sistema CAD

O processo de tomada de decisão dos especialistas pode ser uma árdua tarefa quando os fatores e os recursos disponíveis e conhecidos não são tão claros. O diagnóstico é uma situação complexa, pois depende de informação de variadas naturezas tais como a experiência médica, indicadores clínicos vindos de imagens, sintomas, laudos patológicos, entre outros. Os sistemas CAD buscam auxiliar os especialistas, atuando como uma segunda opinião ao médico, na tomada de decisão.

Segundo Kunio Doi ([DOI-95]), no sistema CAD a imagem é analisada por computador a fim de servir como segunda opinião ao mastologista e desta forma tentar reduzir o número de falsos-positivos e falsos-negativos dos exames e diagnósticos praticados pelos médicos.

Os sistemas CAD empregam técnicas de processamento de imagens e processamento de informações para uma análise computacional quantitativa. É preciso entender a relevância médica das imagens com base nas suas características. Técnicas como Redes Neurais Artificiais, análise fractal e transformadas "Wavelets" são freqüentemente utilizadas nestes esquemas, o que normalmente exige computadores de alta velocidade ([DOI-96]).

Para avaliar a performance dos sistemas CAD existem algumas técnicas que analisam percentuais de acerto do sistema, como o uso das Curvas ROC (Receiver Operating Characteristic)

3.2.1 Curvas ROC

O método das curvas ROC (Curva Característica de Operação do Receptor) consiste em apresentar para o observador imagens cujo conteúdo pode, ou não, ter 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. Logo, o diagnóstico verdadeiro positivo, que acontece quando o observador percebe o sinal realmente existente na imagem, será representado por $p(S/s)$ e o verdadeiro negativo, quando o observador percebe que não há sinal (inexistente realmente) na imagem, será representado por $p(N/n)$ ([EVANS-81]).

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

Quando o sistema de imagem transfere a informação perfeitamente, as probabilidades de S ou N são um ou zero:

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

Mas num sistema real essa probabilidade passa a ser uma fração.

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-81]).

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 o critério que trará maior quantidade de informação e qual resultará em um menor custo para o paciente.

Numa descrição mais detalhada das curvas ROC, é possível verificar que a máxima informação transferida pelo 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-81]).

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

$$p(\text{erro}) = p(S/n) \cdot p(n) + p(N/s) \cdot p(s) \quad (3.1)$$

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

$$\frac{\delta p(\text{erro})}{\delta p(S/n)} = 0 \quad (3.2)$$

ou ainda quando:

$$\frac{\delta p(S/s)}{\delta p(S/n)} = \frac{p(n)}{p(s)} \quad (3.3)$$

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

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

Entretanto, nesta análise não são considerados os custos do teste nem as conseqüências para a saúde do paciente. É de grande importância a consideração dos custos dos exames para encontrar o ponto ótimo de operação nas curvas ROC, pois com o uso de técnicas adequadas pode-se prevenir a população de um teste posterior mais caro e que ofereça maiores riscos. Os custos são calculados com bases em aspectos econômicos e financeiros do exame diagnóstico, como a experiência do médico, o psicológico do paciente, o custo de manutenção do equipamento utilizado no exame, o gasto com a clínica, entre outros fatores. E ainda devem ser consideradas as conseqüências que um erro no diagnóstico podem ocasionar, que são os custos referentes aos falsos-negativos e aos falsos-positivos. Afinal, em um diagnóstico falso-positivo um paciente pode estar sendo tratado de uma doença que ele não tem e, por outro lado, no falso-negativo, um paciente deixa de receber o tratamento adequado.

Todos esses aspectos citados acima devem ser usados para determinar um ponto ótimo de operação das curvas ROC. Este ponto é chamado de máxima relação de custo/benefício.

A precisão (P) é um índice que indica a fração dos casos para os quais o observador esteja correto. As medidas de precisão são sensibilidade e especificidade.

$$\text{Sensibilidade: } \frac{[\text{número de decisões verdadeiras positivas}]}{[\text{número de casos realmente positivos}]}$$

$$\text{Especificidade : } \frac{[\text{número de decisões verdadeiras negativas}]}{[\text{número de casos realmente negativos}]}$$

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

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

Segundo Metz ([METZ-86]), uma curva ROC representa o compromisso entre sensibilidade e especificidade que pode ser obtido de um sistema de diagnóstico, quando o "threshold" (limiar) é variado. Uma comparação entre os sistemas pode ser feita com as respectivas áreas das curvas de cada sistema em questão.

Outra técnica empregada em Curvas ROC é o FROC ("Free Response Operating Characteristic"). A técnica fornece um número arbitrário de anormalidades por imagens e possibilita que o observador indique o nível de confiança e a localização de toda anormalidade percebida. Esse método é considerado o mais próximo de como os radiologistas interpretam as imagens ([CHAKRABORTY-90]).

3.3 Reconhecimento de Padrões

O termo "Reconhecimento de Padrões" está relacionado com muitos problemas de processamento de informações e o diagnóstico médico é um destes problemas ([FRANÇA-99]).

O reconhecimento de padrões é dividido em passos que podem ser melhor visualizados na figura abaixo.

Figura 3.1 Passos que seguem o reconhecimento de padrões

Nos sistemas onde o padrão a ser reconhecido é visual, o passo um, que está descrito na figura 3.1, é fundamental. A aquisição dos dados é importantíssima para o desempenho do sistema, pois a qualidade dos dados de entrada é um fator limitante para a análise de imagens ([DUDA-79]).

Nesse primeiro passo, a cena ou imagem é formada por sensores que captam a energia luminosa e a transformam em sinais elétricos que, digitalizados, serão manipulados para a extração de informações. A aquisição da imagem para os esquemas CAD envolve 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 sistema, que é a parte de classificação. Geralmente os digitalizadores usados são óticos ou laser. Nos óticos o filme radiográfico é varrido por um feixe de luz branca e os sinais são captados por transmissão. Já nos digitalizadores laser, o filme radiográfico é varrido por laser, neste caso a varredura é mais rápida e a transmissão dos sinais também.

Uma vez representada a imagem matematicamente, como já visto, 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-82, DUDA-73, GONZALES-87, SCHALKOFF-89]). Os padrões nos sistemas CAD são mamogramas digitalizados ou parte das imagens destes mamogramas, conhecidas como regiões de interesse ("ROI"). Destas imagens são extraídas informações sobre microcalcificações isoladas, agrupadas, de textura e massa, incluindo morfologia e intensidade.

O segundo passo consiste em selecionar as características que melhor representem o problema ou o padrão em questão ([BEALE-90]). Alguns métodos estatísticos auxiliam nesta etapa. Esse passo procura eliminar classes altamente correlacionadas, a fim de evitar atributos com informações insuficientes para diferenciar uma classe da outra e evitar que sejam, entre os atributos, identificadas outras subclasses.

O último passo é a classificação propriamente dita. Tem como objetivo decidir em qual classe está inserido um certo padrão, a partir da análise do vetor de características extraídas dos dados de entrada. O diagrama de blocos a seguir mostra um sistema CAD, utilizando reconhecimento de padrões.

Figura 3.2 - Etapas de reconhecimento de padrões em um sistema CAD ([JIANG-98])

3.3.1 Extração de Características

Raramente se usa apenas uma medida de cada padrão para que o classificador decida a qual classe pertence. O mais comum é extrair várias medidas de cada padrão e representá-las através de um vetor, o qual servirá de entrada para o classificador. As características são extraídas através do processamento da imagem.

As características de amplitude da imagem, ou características espectrais, são obtidas 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.

Conforme estudado em ([PRATT-91]), a mediana dos pixels em uma vizinhança de região $W \times W$ é dada como a amplitude do pixel da região em questão, onde numa metade os pixels são menores ou iguais a esta amplitude e na outra metade, os pixels são maiores ou iguais.

As características de textura são também muito importantes, pois elas 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-79]). As características de textura são características do domínio espacial da imagem, que também pode trazer informações como sua altura e largura ([SCHALKOFF-89]). Ainda podem-se extrair características do domínio temporal, que serão medidas que informarão movimentos da imagem.

No caso da extração de características de microcalcificações isoladas e agrupadas, como é preciso obter informações que possibilitem a identificação da malignidade ou não, ou ainda quantificar a suspeita, 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 a associação das estruturas a lesões malignas ou benignas.

3.3.1.1 Descritores de Forma

Existem muitas teorias de descrição e reconhecimento de Forma, cada uma delas se esforçando para explicar algum aspecto do problema. Isso ocorre devido ao fato de podermos conceituar a percepção da forma como uma função de alto-nível ([GOLDSTEIN-98]).

Uma das coisas mais importantes a serem determinadas em uma imagem é a forma dos objetos contidos na mesma. Webster define a forma como “a qualidade de um objeto, a qual depende da posição relativa de todos os pontos que compõem seu contorno ou superfície externa” ([WEBSTER-64]). O Reconhecimento da forma dos objetos em uma cena é facilmente determinado pelo sistema visual humano, mesmo quando o objeto está transladado, rotacionado, escalonado, parcialmente obscuro, distorcido ou visto sob diferentes perspectivas. O reconhecimento invariante de formas é importante para seres humanos em diversas tarefas, apesar de o reconhecimento *variante* ser importante em algumas tarefas. Em alguns problemas a determinação da posição (transformação de translação) ou orientação (transformação da rotação) de um objeto é primordial. Em outros casos, a forma do objeto é essencial e não sua posição e orientação. Várias medidas exibem propriedades invariantes, entretanto, segundo Hall & Park (1986) ([PARK-86]), os Momentos possuem a teoria matemática mais elegante. Existem duas abordagens principais para descrever a forma de uma figura bidimensional fechada. Uma abordagem é extrair descritores de formato baseados no perímetro da figura (contorno). Quando existe a necessidade de extrair detalhes finos da forma do objeto, os métodos baseados em perímetro são bons, embora sejam muito sensíveis à ruídos. Esse efeito pode ser reduzido utilizando-se descritores de forma em toda massa da figura, como fazem os momentos.

O método dos Momentos proporciona uma técnica robusta para a 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 a 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. A forma é freqüentemente representada por uma imagem binária e seus momentos podem ser obtidos através de algoritmos rápidos. ([ANDERSEN - 85]). 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 várias variáveis podem ser caracterizadas por uma seqüência de números, 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-61]) desenvolveu a teoria dos Invariantes de Momentos (IM) e sugeriu a sua aplicação para o reconhecimento invariante de padrões visuais, baseado na teoria dos Invariantes Algébricos ([ELLIOTT - 13]). O termo “invariante” denota uma quantidade (ou um valor) que permanece inalterada sobre 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 passado no trabalho de Carley & Sylvester ([GUREVICH - 64]). Baseado nesse trabalho, Hu apresentou fórmulas para funções invariantes contendo momentos de segunda e terceira ordem, como exemplos da teoria algébrica de invariantes de momentos.

Os invariantes de Momento representam um conjunto completo de descritores de imagem, possuindo, portanto, uma propriedade de fundamental importância para a teoria de Reconhecimento de Padrões ([DUDA - 73], [HALL - 79], [BALLARD - 82], [GONZALES - 87], [PRATT - 91], [JAIN - 95]).

A definição de momentos¹ (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 (3.4) a seguir:

$$m_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q \cdot f(x,y) dx dy \quad (3.4)$$

(p, q = 0, 1, 2, ...)

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 3.3 $x^0 y^0$, $x^0 y^1$, $x^0 y^2$, $x^1 y^0$, $x^1 y^1$, $x^2 y^0$, $x^2 y^1$, $x^3 y^0$, $x^1 y^2$. 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.

¹ Um momento é uma soma ponderada de um padrão $f(x)$ sobre todo o campo de entrada, com os pesos iguais à algum polinômio em x ([WOOD - 96])

Figura 3.3 - Monômios de até terceira ordem que multiplicam a função $f(x,y)$ no cálculo dos momentos (Hall - 1978).

3.3.1.2 Os Invariantes de Momentos de Hu

O conjunto dos invariantes de momento absolutos consistem em combinações não lineares aplicados 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 (IM) ([Hu - 62]), que são

invariantes à translação, rotação e escala. Essas funções são mostradas pela equação 3.5 a seguir.

$$\phi_1 = \mu_{20} + \mu_{02} \quad (3.5-a)$$

$$\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (3.5-b)$$

$$\phi_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad (3.5-c)$$

$$\phi_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} - \mu_{03})^2 \quad (3.5-d)$$

$$\begin{aligned} \phi_5 = & (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}) \cdot [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + \\ & + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03}) \cdot [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \quad (3.5-e)$$

$$\phi_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{03} + \mu_{21}) \quad (3.5-f)$$

$$\begin{aligned} \phi_7 = & (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - \\ & (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \quad (3.5-g)$$

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 - 94]) sugere um conjunto de transformações que atenuam esses problemas e podem ser vistos na equação 3.6 mostrada a seguir:

$$\begin{aligned}
nh1 &= 10 \log(10\phi_1 - 1.2) \\
nh2 &= 3 \log\left(\frac{\phi_2 + 0.0001}{0.0101}\right) \\
nh3 &= 8 \log\left(\frac{\phi_3 + 0.0005}{0.0025}\right) \\
nh4 &= 2 \log\left(\frac{\phi_4 + 0.0000001}{0.0000101}\right) \\
nh5 &= 1000\sqrt[3]{\phi_5} \\
nh6 &= 200\sqrt[3]{\phi_6} \\
nh7 &= 500\sqrt[3]{\phi_7}
\end{aligned} \tag{3.6}$$

Enquanto os IM possuem algumas propriedades que boas características de formato devem ter, nem sempre são usados todos eles na solução de um determinado problema em particular. Em alguns casos, poucos IM de baixa ordem podem refletir em características discriminantes de formato suficientes ([CASTLEMAN - 79]).

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 a normalização em relação à rotação consiste em rotacionar o objeto para uma posição padrão, método que foi definido por Hu ([HU - 62]) 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 a figura é achar o seu eixo principal, ou seja, a direção onde ocorre o maior espalhamento, e rotacioná-lo até que seja 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 3.4 abaixo, extraída de ([MASTERS - 94]) mostra a posição padrão.

Figura 3.4 – Posição padrão para Invariância à Rotação ([MASTERS - 94]).

Essas condições podem ser estabelecidas em termos de seu efeito nos momentos centrais. Para satisfazer as condições acima, os momentos da figura deve possuir as seguintes propriedades:

- $\mu_{11}=0$. Essa condição se verifica quando os eixos principais da figura estão alinhados com os eixos das coordenadas vertical e horizontal.
- $\mu_{20} > \mu_{02}$. Essa condição estabelece que o eixo principal da figura é horizontal.
- $\mu_{30} > 0$. Essa condição assegura que a maior massa do objeto se encontra do lado esquerdo, com sua cauda inclinada à direita.

Existem muitas outras normalizações que são menos comuns, mas que podem ser extremamente úteis, por exemplo, pode-se citar a normalização em relação à reflexão e à relação de aspecto, geralmente aplicada após as normalizações de translação, escala e rotação.

Uma outra propriedade qualitativa dos momentos invariantes é que os componentes da imagem são ponderados progressivamente à medida que suas localizações se afastam do centro de massa da imagem, assim podemos pensar que o kernel $x^p y^q$ dos momentos geométricos decresce à medida que x e y se aproximam de zero. Isso indica que os momentos *salientam* a informação contida próximas às bordas dos objetos. Além do mais, devido à natureza de variação lenta das funções base dos momentos de baixa ordem, os detalhes macroscópicos são relacionados resultando em um efeito de média, que ser interpretado como o valor médio de um dado momento dos objetos contidos na cena ([FREEMAN-88]).

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 fornecerem 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 em uma 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-92]).

Em particular, os invariantes de momentos 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 as 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 a figura é assimétrica, 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)². 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 a forma. Se certas porções de

² Uma forma é denominada possuir simetria rotacional n -arestas (n -fold) se tornar-se 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.

simetria forem perdidas, a simetria pode fornecer pistas para recuperar a forma. Além do mais, a ordem de simetria n representa uma boa característica de forma uma vez que é invariante a escala, a translação e a rotação.

3.3.1.3 Interpretação Geométrica dos Momentos

O uso de Momentos como descritor global 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 interseção das linhas $x = \bar{x}$ e $y = \bar{y}$. Essas linhas são paralelas aos eixos x e y e representam as linhas onde toda a 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 a posição do objeto 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 3.7 ([STRACHAN - 90]):

$$T = \frac{1}{2} \arctan \frac{2(m_{00}m_{11} - m_{10}m_{01})}{(m_{00}m_{20} - m_{10}^2) - (m_{00}m_{02} - m_{01}^2)} \quad (3.7)$$

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 - 61]). 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 - 61]).

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

Momentos Centrais	Interpretação Geométrica
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. Em uma 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. Em uma 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 de 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 a 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 a 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 ($X_{l/2}$) do objeto.

- Se $\mu_{30} < 0$ CG à direita de $X_{l/2}$.
- Se $\mu_{30} > 0$ CG à esquerda de $X_{l/2}$.
- Se $\mu_{30} = 0$ $\bar{x} = X_{l/2}$.

• 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 caracteriza a variabilidade de uma função densidade e mede 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 Giração³ sobre um eixo é a distância do eixo à linha onde toda a 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 (3.8) e (3.9) abaixo:

$$RG_x = \sigma_x = \sqrt{\frac{\mu_{20}}{m_{00}}} \quad (3.8)$$

$$RG_y = \sigma_y = \sqrt{\frac{\mu_{02}}{m_{00}}} \quad (3.9)$$

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 a massa está centrada, sem alterar o segundo momento sobre a origem e é dado pela equação 3.10 a seguir:

$$RG_{xy} = \sqrt{\frac{\mu_{20} + \mu_{02}}{\mu_{00}}} \quad (3.10)$$

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

³ O raio de Giração pode ser interpretado também como o raio mínimo que inclui a maioria, senão todos os pontos em seu interior.

3.3.1.4 Outros Atributos de Forma

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

Vários esquemas CAD usam estas características, pois em achados mamográficos, por exemplo, o formato, tamanho e distribuição dos objetos, como vistos 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 ([GONZALES-87], [MASCARENHAS-89], [MASTERS-94], [SCHALKOFF-89]).

O perímetro é dado pelo número de pixels do contorno do objeto. Existem várias maneiras de se calcular perímetro e área ([MASCARENHAS-89, GONZALES-87],).

A Compacidade é uma medida derivada do perímetro P e área A de um objeto, esta medida indica o quanto o objeto em análise é sólido. Em ([MARTÌ-98]), mostra-se a extração desta de compacidade de imagens de microcalcificações, usando a equação 3.11.

$$C = \frac{P}{4\pi A} \quad (3.11)$$

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 3.12, pois no seu trabalho foi considerado irregularidade de fronteira como sugestivo de malignidade ([TAO-98]).

$$I = \frac{A}{P^2} \quad (3.12)$$

Muitas outras medidas de formato estão listadas no capítulo 5, conforme vistas nos esquemas CAD, para extrair informações de microcalcificações, regiões de interesse e nódulos. Estas medidas, quando combinadas, representam bem as classes em que se dividem os achados mamográficos.

3.3.2 Seleção de Características

O processo de seleção de atributos tem como função mapear as medidas extraídas dos padrões (imagens) em características significativas ([GONZALES-87, MASCARENHAS-89]). Esse processo tem duas ênfases:

- transformar o espaço de atributos, que reduz a dimensão dos vetores de características em busca de torná-lo o mais significativo possível na representação de uma classe;
- calcular a distância entre as classes, e medidas que forneçam informações que sejam possíveis avaliar a separabilidade entre as classes;

Para a transformação de atributos, técnicas como transformada de Karhunen-Loève e a função discriminante de Fisher são bastante usadas. No cálculo da distância entre as classes, técnicas como medida de Divergência e Distância Jeffries-Matusita (JM) são empregadas ([YOUNG-86]). As técnicas estão descritas detalhadamente em ([MASCARENHAS-89]).

A distância JM é usada para medir separabilidade envolvendo múltiplas classes ([DUDA-73], [MASCARENHAS-89]). Já o conceito da Divergência é definida apenas para duas classes, fornecendo informação média para a discriminação em favor de uma classe contra a outra ([MASCARENHAS-89]).

O Erro de Bayes é uma técnica pela qual se calcula a função densidade de probabilidade, que resultará numa medida que realça o significado do atributo e os reorganiza. Em seguida, deve ser usada uma técnica para se calcular as distâncias entre estes subgrupos de atributos selecionados pelo Erro de Bayes ([FUKUNAGA-90]).

3.3.3 Classificação

A Classificação é dependente dos passos anteriores, pois para obtenção de um bom resultado, é preciso que o sistema tenha desempenhado os dois primeiros passos, extração de características e seleção, com sucesso.

Existem duas categorias na classificação de padrões: numéricas - medidas determinísticas e estatísticas - e as não numéricas, tratadas com os métodos segundo a teoria dos conjuntos "Fuzzy".

O classificador Bayesiano é um classificador estatístico. Já o classificador linear é determinístico, bem como a classificação do vizinho mais próximo. Estes são apenas exemplos dos tipos de classificadores.

As Redes Neurais Artificiais vêm sendo usadas como técnica de classificação para reconhecimento de padrões e, conforme foi observado na literatura, vêm apresentando um bom desempenho em problemas associados a sistemas CAD, como classificação de estruturas de mama quanto a lesão maligna ou benigna. Por isso, essa técnica será tratada em mais detalhes, no próximo capítulo.

Na figura 3.5, pode ser visualizada, a partir do esquema em blocos, a projeção de um sistema de classificação.

Figura 3.5 - Processos de um sistema classificador ([FUKUNAGA-90])

Cada passo de um sistema classificador é importante, pois influencia no seu desempenho. Depois da obtenção dos dados, as amostras são normalizadas e organizadas, mas diferentes tipos de dados requerem diferentes tipos de normalização e registro. Em seguida, deve ser medida a separabilidade das classes dos dados através de uma medida de espaço. Visto que não é apropriado neste estágio assumir uma forma matemática para a estrutura de dados, o processo de estimação deve ser não paramétrico ([FUKUNAGA-90]).

Os passos seguintes incluem análise de características com técnicas de extração, agrupamento, testes estatísticos e a escolha do classificador, que será de acordo com as informações extraídas nos estágios anteriores.

É comum em sistemas CAD o uso de classificadores paramétricos, em especial as Redes Neurais "Multi-layer Perceptron" (MLP) com o algoritmo "Backpropagation" ([JIANG-96], [CHITRE-94], [ZHANG-94], [JIANG-98], [WU-

93a], [NISHIKAWA-95], entre outros), que é baseado em métodos estatísticos, conhecidos por métodos de aproximação estocástica.

Alguns trabalhos usando estes tipos de classificadores em sistemas CAD serão discutidos no capítulo 5.

3.4 Conclusão

As técnicas citadas neste capítulo são de grande importância para os sistemas CAD. Estes sistemas reúnem muitas técnicas e ferramentas matemáticas e computacionais, na tentativa de abstrair o complexo raciocínio médico nas tarefas de tomada de decisões em diagnósticos por imagem.

O processamento de imagem auxilia na visualização e detecção de artefatos que, muitas vezes, não são possíveis de se identificar com a visão humana.

Embora de maneira simplista, o reconhecimento de padrões é usado para mapear as informações extraídas de imagens e amostras a fim de classificá-las com base em características lógicas e técnicas. Os métodos de classificação são estruturados a partir de uma esquematização e modelagem da abstração do raciocínio médico no momento do diagnóstico. Esta, com certeza, não é uma tarefa tão simples de modelagem, tão pouco precisa o bastante, sendo feita com as funções lógicas e aritméticas conhecidas e que a máquina consegue entender.

4

Redes Neurais Artificiais

4.1 Conceitos Básicos

As Redes Neurais Artificiais (RNA) estão envolvidas em problemas relacionados com a captura de estímulos e seu processamento, de forma biologicamente plausível, uma vez que a abordagem conexionista, onde as Redes Neurais Artificiais estão inseridas, parte de estudos da Neurofisiologia ([ANDERSON-95]).

Um neurônio do sistema nervoso é visto como se fosse uma unidade computacional que processa uma operação sobre suas entradas. Assim as Redes Neurais Artificiais são biologicamente inspiradas, pois são formadas por unidades que tentam imitar as funções mais elementares dos neurônios biológicos, apesar da ciência ainda não conhecer plenamente o funcionamento cerebral. Isso leva os pesquisadores desta área a descartarem a plausibilidade biológica, fazendo do cérebro uma metáfora e construindo modelos não prováveis à luz dos conhecimentos correntes. Apesar desta semelhança superficial, as Redes Neurais Artificiais exibem um surpreendente número de propriedades típicas do cérebro ([MASTERS-94]).

O cérebro realiza o processo de aprendizagem através das mudanças nas concentrações químicas do interior do corpo celular do neurônio. Assim como os

neurônios biológicos, em uma RNA, estas mudanças nas concentrações químicas são representadas pela modificação dos valores dos pesos e limiares associados aos neurônios artificiais. Este processo de aprendizagem das RNA pode ser entendido como uma técnica de aproximação de funções sucessivas, discretas no tempo, de modo que a função realizada pela RNA está se aproximando de uma dada função pertencente ao espaço do conjunto de amostras de um problema que, em geral, são conhecidos apenas alguns valores. Quando o processo discreto é finito e o erro associado à aproximação é menor que um valor real previamente definido, dizemos que a rede neural convergiu para a função desejada: isso significa dizer ela "aprendeu" ([MARAR-97]). Esse "aprendizado" se dá a partir de experiências passadas, podendo generalizar quando se deparam com novas situações baseando-se em exemplos já conhecidos.

As Redes Neurais Artificiais 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, etc.. Mais aplicações para as redes neurais foram surgindo à medida que elas foram sendo investigadas e, então, passaram a ser empregadas na soluções de outros tipos de problemas como compressão de dados, reconhecimento de padrões, otimização combinatória e classificação, entre outros.

4.2 Um modelo de RNA genérico

O primeiro passo para o desenvolvimento de uma rede é a coleta de dados relativa ao problema e a separação destes em um conjunto de treinamento e um conjunto de teste. Os dados coletados devem ser significativos e cobrir de maneira ampla o domínio do problema ([PATTERSON-96]).

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 à sua utilização na rede. Estes pré-processamentos referem-se aos passos anteriores à classificação, como foi citado na discussão dos estágios de um sistema classificador apresentado na figura 3.3, no capítulo anterior.

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 e enviam 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 esquematizadas 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 no 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-94]).

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 (conjunto de exemplos).

O estado de ativação da rede, termo bastante citado na literatura, mostra a representação do estado de um sistema em um determinado instante de tempo em forma de vetor $a_i(t)$ que representa o estado de ativação da rede.

A função de saída está associada com cada unidade (u_i) e mapeia o atual estado de ativação ($a_i(t)$) em um sinal de saída ($O_i(t)$). Assim $O_i(t) = f_i(a_i(t))$, sendo f_i uma função de transferência (função de ativação).

As funções de ativação podem ser também do tipo "Siebert", ou ainda rampa. As funções de ativação "Siebert" podem ser representadas por uma função sigmóide (logística) e pela tangente hiperbólica que são simétricas e possuem todas as derivadas contínuas. As funções de ativação mais usadas podem ser vistas na figura 4.1 ([KOVACS-96]).

Figura 4.1 - Gráficos das funções de ativação mais usadas para a modelagem de neurônios.

Para que cada unidade tenha efeito sobre a outra, sua ativação deve exceder um certo valor "threshold" (limiar). A função de saída determina se a unidade transmite informações para outras unidades e a intensidade dessa transmissão.

Um padrão de conectividade entre as unidades define o conhecimento do sistema e determina como a rede responde a uma certa entrada. Os padrões de conectividade são resumidos em uma matriz de conectividade ou matriz de pesos $[w]$.

As RNAs podem ser "Feedforward", ou seja, são as redes de fluxo direto: todos os pontos de conexões seguem em uma direção, propagação para frente. Todos os nós estão conectados entre si: os neurônios de uma dada camada i estão conectados com todos os neurônios da próxima camada e com todos os neurônios da camada anterior. Outra maneira de propagação de um sinal na rede é "Feedback", que possibilita que o fluxo de informações além de correr em uma direção seja recorrente ([FU-97]).

As RNAs também diferem quanto a 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-91]). Neste trabalho, a aprendizagem supervisionada, será tratada com ênfase, dada sua importância e eficiência na aplicabilidade de problemas como o que queremos resolver, segundo estudos bibliográficos incluídos neste contexto.

4.3 Neurônio de Mc-Culloch e Pitts

Em ([Mc-CULLOCH-43]) foi proposto o primeiro neurônio artificial. A figura 4.2 mostra o modelo proposto pelos autores Mc-Culloch e Pitts e faz uma comparação entre o neurônio biológico e o artificial.

Figura 4.2 Neurônio Biológico X Modelo Matemático

Na mesma figura 4.2, está explícita a comparação que se pode fazer da realidade biológica com um modelo artificial. A resposta de um neurônio biológico diante de um estímulo depende se ele será disparado ou ficará inativo de acordo com a concentração química no interior do corpo celular. Estas diferenças de concentração são chamadas sinapses. Desta forma, os modelos matemáticos procuram representar as sinapses com as modificações dos pesos nas redes.

O funcionamento do modelo de McCulloch e Pitts pode ser resumidamente descrito da seguinte maneira:

$$Saída(x) = 1 \text{ se e somente se } \sum_{j=1}^n w_j x_j - \theta > 0$$

$$Saída(x) = 0 \text{ se e somente se } \sum_{j=1}^n w_j x_j - \theta \leq 0$$

onde $Saída(x)$ é a saída do neurônio e x_j e w_j são as entradas e pesos das conexões respectivamente, n é o número de entradas e θ é o "threshold".

Os pesos são representados pela matriz W , onde cada elemento w_{ij} corresponde a um valor referente à conexão de cada unidade da camada i para as unidades da camada j .

O cálculo de *Saída* (x) tem o objetivo de ser análogo ao que ocorre na soma da célula nervosa. A função de ativação não atua diretamente sobre as entradas, mas sobre uma soma ponderada dos pesos das conexões pelas respectivas entradas.

4.4 Redes Neurais Supervisionadas

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. Então os pares de entrada são apresentados à rede: para cada entrada a rede produz uma saída (saída obtida pela rede), a qual é comparada com a saída desejada para a dada entrada. Caso a saída da rede seja diferente da saída desejada, a rede calcula um erro, que é um valor que representa o quão distante está a sua resposta daquela desejada. Este erro é usado para que os pesos da rede sejam ajustados até que a saída da rede seja aproximadamente igual a saída desejada, e então o erro é minimizado. Este processo de minimização dos erros exige um circuito especial, que é chamado de "professor", daí o nome *redes de aprendizado supervisionado*.

Alguns fatores devem ser considerados na fase de aprendizagem, como o tempo gasto por iteração e o número de iterações para cada par de entrada alcançar o erro mínimo.

Um algoritmo bastante usado no treinamento de redes supervisionadas é o "Backpropagation", pois considera os parâmetros citados, de maneira a alcançar uma aprendizagem ótima. O algoritmo "Backpropagation" será comentado mais adiante, já que, conforme mostra a literatura vem sendo de grande utilidade em aplicações de classificações e reconhecimento de padrões, e fez parte do nosso trabalho.

4.5 O Perceptron

Um dos primeiros modelos de redes neurais, o "perceptron", foi sugerido por Rosenblatt ([ROSENBLATT-58, ROSENBLATT-60]). Ele modela um neurônio realizando uma soma ponderada de suas entradas e enviando o valor um para a saída, caso a soma seja maior que algum valor inicial ajustável, e valor zero caso contrário ([RICH-88]).

Figura 4.3 Perceptron

Na aprendizagem em neurônios simples, analisada em termos cognitivos, há uma tendência em se aprender o comportamento recompensado e esquecer o penalizado.

O teorema de Roseblatt é dado por: seja S um subconjunto do conjunto de amostras de um problema e $G : S \rightarrow \{0,1\}$. Existe um "perceptron" com n entradas que aproxime uma dada função se, e somente se, S é linearmente separável com relação a G ([ROSENBLATT-58]).

O teorema sobre os "perceptrons", acabou resultando em um algoritmo de aprendizagem e provando que qualquer função linearmente separável pode ser representada por uma única camada de neurônios, entretanto se for uma função linearmente não separável, o algoritmo pode oscilar indefinidamente e não convergir.

O algoritmo de aprendizagem do Perceptron pode ser descrito da seguinte forma:

- 1- Consideram-se valores iniciais de pesos e limiares:

$$W_{ij}(t), 0 \leq i \leq n$$

$$T_j (T \text{ é o threshold})$$

Deve-se atribuir valores aleatórios aos pesos e "thresholds";

- 2- Apresentar as entradas e saídas desejadas em valores binários onde x_i são as entradas e z_i são as saídas desejadas;
- 3- Calcular o efeito da entrada na saída

$$Y(t) = f_n \left[\sum_{i=0}^n W_i(t) x_i(t) \right]$$

onde

$Y(t)$ é a saída da rede

$$f_n(v) = 1 \Rightarrow v \geq 0$$

$$f_n(v) = 0 \Rightarrow v < 0$$

W_i são os pesos;

A função de ativação ($f_n(v)$) utilizada é a "Heaviside" (rampa);

- 4- Recalcular os pesos

$$Y(t) = 0, \quad Z(t) = 1 \Rightarrow W_i(t+1) = W_i(t) + x_i(t)$$

$$Y(t) = 1, \quad Z(t) = 0 \Rightarrow W_i(t+1) = W_i(t) - x_i(t)$$

$$Y(t) = Z(t), \quad \Rightarrow W_i(t+1) = W_i(t)$$

Para $1 \leq i \leq n$,

onde n é o número de unidades de entrada, $Y(t)$ é a saída da rede e $Z(t)$ é o alvo ("Target") ou saída desejada.

O modelo de Rosenblatt com relação ao de McCulloch e Pitts apresentou algumas vantagens, tais como: 1) estímulos podendo assumir valores contínuos em oposição aos valores apenas binários e 2) introdução de uma lei de treinamento para os "perceptrons".

Como solução do problema de convergência para as funções não linearmente separáveis vale citar a Regra Delta, que foi introduzida logo depois do "perceptron" por Widrow e Hoff ([WIDROW-60]). O método busca minimizar o erro médio quadrado entre a saída desejada e a saída obtida pela rede. Com a Regra Delta para funções não linearmente separáveis é possível obter uma boa convergência dos dados.

Os passos do algoritmo da Regra Delta de Widrow consistem em:

- 1- Inicializar os pesos com valores pequenos e aleatórios;

- 2- Calcular o produto escalar dos pesos pelas entradas, onde $pe = \sum w_i'x_i$, para cada unidade $(x_1, x_2, x_3, \dots, x_n)$ do conjunto de treinamento, pe é a saída obtida pela rede;
- 3- Calcular o erro, $\delta = \text{saída desejada} - pe$;
- 4- Ajustar os pesos através da equação: $w_i^{\text{nov}} = w_i^{\text{antigo}} + \eta \delta x_i$, sendo η a taxa de aprendizagem¹ (intervalo $[0,1]$), usada para controlar a estabilidade e a convergência dos pesos da rede, pois estabelece o tamanho do passo que se dará naquela direção.
- 5- Voltar para o passo 2 até o fim das amostras e convergência da rede.

A Regra Delta se diferencia do *Perceptron* quanto ao uso ou não da função *heaviside* (no passo 2).

Para problemas da separabilidade não linear foi demonstrado por Minsky e Papert ([MINSKY-69]) que redes com múltiplas camadas de neurônios do tipo "perceptron" (MLP, "Multilayer Perceptron") apresentam um bom desempenho. Mas tanto o algoritmo de aprendizagem do "perceptron" quanto o algoritmo da Regra Delta não funcionam para as redes de múltiplas camadas, pois o ajuste de pesos nas unidades das camadas escondidas torna-se complicado por não existir um erro direto para a medida δ , entre as unidades das camadas escondidas e a saída desejada.

O algoritmo "backpropagation" é uma das soluções para a aprendizagem das redes MPL ([RUMELHART-86]).

4.6 "Backpropagation"

O termo "Backpropagation" (BP) refere-se ao método de aprendizagem no qual os pesos de conexão (e muitas vezes o "*Thresholding*" de ativação do neurônio)

¹ A taxa de aprendizagem foi proposta por Hebb, que se baseou no princípio de que a aprendizagem poderia ser reduzida a um processo inteiramente local, em que a intensidade das conexões sinápticas é alterada em função dos detectáveis localmente ([KOVÁCS-96]).

da rede são ajustados. Durante a operação, todo o fluxo de informação é propagado de maneira "Feedforward" ([FU-97]).

O algoritmo "Backpropagation" é uma aplicação no contexto de aprendizagem em Redes Neurais Artificiais (RNA). Ele é baseado em um método estatístico proposto em 1951 por Hebert Robbins e Sutton Monro ([ROBBINS-51]), que denominaram o método de aproximação estocástica. Várias aplicações desta 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, apresentado no livro PDP (*Parallel Distributed Processing*) ([RUMELHART-86]).

A regra de aprendizagem 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 a camada de saída e propaga-se este 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 estes erros através da rede permite que os pesos entre todas as camadas sejam ajustados corretamente. Por exemplo, mudanças na entrada de uma camada qualquer da rede acarretam mudanças nas saídas destas unidades. Estas mudanças são suavizadas se a variação do "threshold" é suave. A variação do *threshold* é medida pela inclinação da curva da função de ativação ([HERTZ-91]).

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

4.6.1 As Fases do Algoritmo "Backpropagation"

O treinamento BP por ser do tipo supervisionado, requer que à cada vetor de entrada seja apresentado com um vetor de saída desejada correspondente.

No processo de aprendizagem denomina-se "ciclo" a apresentação de todos os N pares (entrada e saída) do conjunto de treinamento. A correção dos pesos em um ciclo pode ser feita de duas maneiras:

- modo padrão, onde a correção dos pesos acontece a cada apresentação à rede de um exemplo do conjunto de treinamento. Cada correção de pesos baseia-se somente no erro do exemplo apresentado naquela iteração, assim, em cada ciclo, ocorrem N correções;
- modo *Batch*, onde apenas uma correção é feita por ciclo. Todos os exemplos do conjunto de treinamento são apresentados à rede, seu erro médio é calculado, e a partir deste erro, fazem-se as correções dos pesos.

Um método bastante usado para aplicações em reconhecimento de padrões com as redes "Backpropagation", é o "leave-one-out", onde o conjunto de dados com N elementos é dado como entrada para a rede e apenas um elemento (par de entrada) fica de fora. A rede é então treinada com $(N-1)$ elementos e testada com o elemento que ficou de fora, isso se repete N vezes e cada vez é deixado de fora um elemento do conjunto de dados para servir de teste ([CHAN-97]).

Outro método é o "cross-validation", em que o conjunto de dados é dividido aleatoriamente em conjunto de treinamento e conjunto de testes. Para impedir a dependência no caso de treinamento e teste, o procedimento é repetido várias vezes com diferentes divisões do conjunto de dados.

Os pesos da rede devem ser inicializados com pequenos valores aleatórios, distribuídos uniformemente em um intervalo definido para evitar que a rede fique saturada precocemente com valores grandes de pesos([HERTZ-91]).

Na prática, é mais utilizado o modo de treinamento padrão devido ao menor armazenamento de dados, além de ser menos suscetível ao problema de mínimo local, devido à pesquisa de natureza estocástica que realiza. Em contra partida, no

modo *batch* se tem uma melhor estimativa do vetor gradiente, tornando o treinamento mais estável. A eficiência relativa dos dois modos de treinamento depende do problema que está sendo tratado.

Quanto ao tempo de treinamento, vários fatores podem influenciar a sua duração, porém sempre será necessário utilizar algum critério de parada. O critério de parada do algoritmo BP não é bem definido e geralmente é utilizado um número máximo de ciclos. Mas, devem ser considerados a taxa de erro médio por ciclo, e a capacidade de generalização da rede.

Os passos de mais alto nível do algoritmo "Backpropagation" são:

1. Inicializar os pesos com pequenos valores aleatórios;
2. Selecionar um par do conjunto de treinamento, apresentando o vetor de entrada às entradas da rede.
3. Ativar a rede para produzir sua saída.
4. Calcular o erro entre a saída obtida (ou seja, a saída produzida pela rede) e a saída desejada (do conjunto de treinamento).
5. Ajustar os pesos da rede com o intuito de minimizar o erro.
6. Repetir os passos de 1 a 5 sobre o conjunto de treinamento, até que o erro global (para todo o conjunto de treinamento) seja aceitavelmente baixo.
7. Voltar ao passo 2 e repetir o algoritmo até que todos os padrões do conjunto de treinamento tenham sido treinados.

Conjuntamente, os passos 2 e 3 são conhecidos como passos de propagação para frente. 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. Este erro é então utilizado pelo passo 5 para ajustar os pesos da rede.

A rede só é considerada treinada quando, após várias repetições destes 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 máximo.

A fase de uso da rede "*Backpropagation*" corresponde aos passos 2 e 3 de seu processo de treinamento. Um vetor de entrada é aplicado à rede e este 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 as entradas aplicadas.

4.6.2 A Matemática do "Backpropagation"

Figura 4.4 Rede "Feedforward" com uma camada escondida ("multilayer-perceptron").

Na figura 4.4 considera-se a seguinte legenda:

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

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

k .

ξ_k^μ = entradas;

ξ_i^μ são as saídas desejadas;

O_i = saídas obtidas pela rede;

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

Segundo Hertz em ([HERTZ-91]), os passos do algoritmo podem ser detalhados com o seguinte equacionamento descrito a seguir.

As entradas são introduzidas à rede com um valor particular e são 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, 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:

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

torna-se

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

Na figura 4.5 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 4.5 Rede "Backpropagation"

A regra de atualização é local.

Considerando:

η = taxa de aprendizagem;

δ = erro;

Para calcular as saídas das camadas escondidas temos:

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

onde δ_i^{μ} é definido por:

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

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

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

Este 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. Ele pode ser generalizado 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 descontinuidades 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 reescrita 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 $\frac{1}{2}$.

Outros tipos de função de erro δ podem ser usadas ao invés do erro quadrático, como foi mostrado anteriormente. Uma outra função é a medida de entropia que tem mostrado resolver problemas que não podem ser resolvidos pelo erro quadrático E . A medida de entropia é apropriada quando o conjunto de treinamento são dados probabilísticos ou "Fuzzy", como por exemplo, na associação de sintomas com causas em diagnósticos médicos([HERTZ-91]).

O gradiente descendente é muitas vezes usado para situações cujo problema de decisão é binário (+ 1 ou -1) e então, considera-se qualquer saída positiva ($O_i > 0$) como sendo um sim, e qualquer saída negativa ($O_i < 0$) como sendo um não. Isto funciona bem desde que o problema seja linearmente separável mas não linearmente independente. Winter e Denker ([WINTER-88]) mostraram que uma possível alternativa é modificar a função custo de acordo com o problema que se pretende resolver, só então o gradiente descendente poderá encontrar uma solução se ela existir.

O uso de uma medida de probabilidade no cálculo da função erro é outra proposta para que o gradiente descendente não se comporte como a função do erro quadrático.

Com equacionamento descrito até agora, pode-se detalhar o algoritmo do BP passo a passo com seus respectivos equacionamentos:

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

tal que $V_k^0 = \xi_k^\mu$ 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 as 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

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

para $m = M, M-1, \dots, 2$ até que o delta tenha sido calculado para todas as 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 os padrões.

O algoritmo BP foi implementado neste trabalho. No capítulo 6 estão descritas com mais detalhes a construção de uma rede MLP "Backpropagation" e o uso como classificador

4.7 Conclusão

As Redes Neurais Artificiais são ferramentas que apresentam grande eficiência em muitas aplicações. Elas abrangem um amplo campo de uso e dentre eles o de classificação de padrões em Visão Computacional.

A facilidade de seu uso se dá devido ao alto grau de liberdade para ajustar seus parâmetros de acordo com a aplicabilidade que se deseja.

Em particular, as redes supervisionadas com algoritmos BP são usadas freqüentemente em esquemas CAD, pois são consideradas como boas

classificadoras. Além disso a implementação do algoritmo BP não é tão complexo, e apresenta um certo poder de generalização do conhecimento.

O poder de generalização em reconhecimento de padrões é um fator bastante importante e necessário na maioria das aplicações, além da capacidade de aprender a partir de exemplos, o que facilita o seu uso e ajuste.

Uma das maiores vantagens das Redes Neurais é a otimização computacional, visto que elas empregam um paralelismo no processamento das informações e cada unidade da rede é vista como uma unidade de processamento.

5

Sistemas Computadorizados de Classificação de Estruturas Mamográficas

Vários modelos usando Visão Computacional e Inteligência Artificial têm trazido grandes contribuições para as pesquisas em processamento e avaliação de imagens médicas com o objetivo de auxiliar nos difíceis diagnósticos clínicos. Esquemas CAD em mamografia têm apresentado muitos estudos para classificar microcalcificações (e, eventualmente, "clusters" de microcalcificações) em malignas ou benignas.

Das imagens digitalizadas de mamogramas são extraídas diversas informações de características das microcalcificações e "clusters", sendo classificadas por formato, dimensão, quantidade de microcalcificações em um agrupamento. Informações podem ser obtidas no processamento destas imagens através de morfologia matemática.

Duas categorias de atributos, ou características, são tipicamente usadas nos esquemas de classificação ([ROGOVA-99]):

- métodos de visão computacional que empregam vários intensificadores baseados na extração automática de características das imagens;
- métodos do domínio do conhecimento humano, que usam informações das características da imagem consideradas por médicos especialistas.

5.1 Classificação de Estruturas Mamográficas Baseada em características extraídas com métodos de Visão Computacional

Em 1993 Dhawan et al. ([DHAWAN-93]) apresentaram um trabalho de classificação de microcalcificação de mama baseada em Redes Neurais Artificiais. Foram usados mamogramas digitalizados com 256 níveis de cinza, e consideraram pequenas regiões de 1cm^2 dos mamogramas que apresentavam microcalcificações. Estas regiões foram processadas e foi construído um histograma de segunda ordem para cada uma delas, do qual as características das microcalcificações foram extraídas.

Estas características que serviram como dados de entradas para o treinamento da Rede Neural Artificial consistiam em entropia, contraste de H (histograma), momento angular de segunda ordem de H (ASM de H), momento diferencial inverso de H , correlação de H , média de H , derivação de H , entropia de H_{diff} , onde H_{diff} é diferencial do histograma de segunda ordem, momento angular de segunda ordem de H_{diff} (ASM de H_{diff}), média de H_{diff} . As informações foram divididas em duas classes e dispostas em vetores de 5 posições com cada posição representando uma informação. Para essa divisão os autores utilizaram métodos estatísticos para avaliar quais as características extraídas da imagem eram mais significativas.

A Rede Neural Artificial usada para classificar as microcalcificações em malignas e benignas foi uma "MLP", a configuração da rede foi de 5 neurônios na camada de entrada, 2 na camada escondida e um na camada de saída, com o algoritmo de aprendizagem "Backpropagation". Segundo os autores a rede apresentou bons resultados na classificação de microcalcificações de difíceis diagnósticos em categorias malignas e benignas, uma curva Gaussiana mostrou a distribuição de informações em cada uma das classes.

Liang Shen e seus companheiros, também em 1993, desenvolveram um sistema computadorizado de detecção e classificação de calcificações em mamografias ([SHEN-93]), onde descreveram uma detecção de microcalcificações pelo método de crescimento de região e as classificaram com Rede Neural Artificial. O sistema detectava uma região com potencial de calcificações e as contornavam, de

forma a extrair informações sobre o formato. Um algoritmo de treinamento de redes neurais baseado em "step" foi usado, o "leave-one-out", que usa mesmo conjunto de padrões para treinamento e testes da rede, onde o treinamento é feito K vezes, com (K-1) padrões por vez e o teste é realizado com o padrão que ficou fora do conjunto de treinamento. O método usado para a detecção está ilustrado na figura-5.1.

Figura 5.1 - Esquema CAD proposto por ([SHEN-93])

Os autores extraíram informações de forma, centro de gravidade, e tamanho para determinar os melhores parâmetros. Dentre eles, as características de forma foram extraídas a partir de descritores de momento, descritores de Fourier e solidez, com base nos contornos das regiões e microcalcificações.

Para o treinamento da rede eles usaram 18 mamogramas, com comprovação de biópsia. Estes mamogramas continham 64 calcificações benignas e 79 malignas.

A taxa de acerto obtida na classificação da RNA foi de 94% para benignas e 87% para as malignas. Estes resultados foram considerados bons pelos autores.

Os mesmos pesquisadores, em 1994 ([SHEN-94]), publicaram resultados de pesquisas desenvolvidas no sentido de classificar as calcificações, em benignas e malignas, de acordo com suas formas. Para extrair os contornos das regiões de interesse, foi utilizado o método de crescimento de região. Após a extração do

contorno desejado, foram calculadas algumas medidas: *momentos regulares* - que são os tipos mais populares de momentos e *momentos das regiões de fronteira*. Os primeiros são definidos como:

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy, \quad \text{para } p, q = 0, 1, 2, \dots, \quad (5.1)$$

onde $f(x, y)$ é uma imagem bidimensional e contínua.

Foi sugerido, então, um novo conjunto de momentos, derivado pelos autores, e outros números, como descritores de Fourier e compacidade, foram calculados e analisados para algumas formas básicas. Observou-se que algumas medidas variaram e outras permaneceram constantes de acordo com a forma da estrutura. Após a obtenção desses resultados, o método foi testado para calcificações mamográficas, avaliados 80 mamogramas com dados comprovados por biópsia. No geral, as calcificações associadas com condições benignas apresentaram-se maiores, mais arredondadas, em número pequeno e distribuídas de forma difusa, além de apresentar maior homogeneidade em termos de forma e tamanho. Os autores comentam ainda que uma das maiores diferenças entre calcificações malignas e benignas é a rugosidade (irregularidade) de suas formas.

Os dados obtidos nos trabalhos citados são de extrema importância, uma vez que, além da detecção de microcalcificações em mamogramas, é necessária a sua classificação de acordo com a possibilidade de relacionamento com tumores malignos ou benignos para que seja efetiva a contribuição dos esquemas computadorizados em diagnósticos precoces do câncer de mama.

As Redes Neurais Artificiais (RNA) têm sido aplicadas no campo de imagens médicas como reconhecimento de padrões favorecendo a classificação de pacientes em uma variedade de situações diferentes ([WU-93a]).

O grupo da Universidade de Chicago, coordenada pelo Dr. Kunio Doi vem se destacando há tempos com trabalhos nesta área. Em 1993, apresentaram um trabalho também empregando RNA em mamografia para aplicação em tomadas de decisões no diagnóstico do câncer de mama ([WU-93b]).

Neste trabalho eles inicialmente selecionaram 43 características extraídas de mamogramas que serviram como dados de entrada para a RNA. Esses atributos eram caracterizados dentro de três grupos: características relacionadas a massa, atributos relacionados a microcalcificações e atributos de anormalidades secundárias.

Uma RNA "feedforward" com três camadas usando algoritmo "backpropagation" foi empregada neste estudo. Uma função logística com parâmetros de centro (largura) e amplitude ajustáveis foi usada como função de ativação para cada uma das unidades da rede. A estrutura da rede a princípio era de 43 unidades de entrada, cada qual representando uma característica extraída dos mamogramas, 10 unidades escondidas e 1 unidade de saída, que representava o diagnóstico (1=maligno, 0=benigno).

O conjunto de treinamento foi baseado em 133 casos selecionados de uma atlas mamográfico de casos compilados por Tábar e Dean¹. Os 43 atributos extraídos foram indicados numa escala de 0 a 10, por especialistas experientes em mamografias. Depois disso os valores foram normalizados de 0 a 1, e as taxas para cada caso foram usadas como entrada da rede. Todos os casos tinham a informação histopatológica de malignidade ou não. Na sequência, foram selecionados 60 casos clínicos, confirmados com biópsia, de maneira aleatória para avaliar a performance da RNA.

A tabela abaixo, mostra um exemplo de como foram escalonadas as características de microcalcificações segundo os especialistas.

Tabela 5.1 Escalonamento das características de microcalcificações usado em ([WU-93b])

Atributos	Taxa
Número de microcalcificações	
Número de microcalcificações em "clusters" (10=muitas)	6
Presença de microcalcificações similares pela mama (não agrupadas) (10=microcalcificações não similares)	8
Formato de microcalcificações	
Arredondada a irregular (10=irregular)	8
Presença de microcalcificações lineares (10=lineares)	8
Presença de microcalcificações ramificadas (10=ramificadas)	7
Uniformidade das microcalcificações	
Uniformidade no tamanho (10=não uniforme)	9
Uniformidade no formato (10=não uniforme)	8
Uniformidade na densidade (10=não uniforme)	7
Distribuição de microcalcificações	

¹ Tábar L, Dean PB. Teaching atlas of mammography. 2nd ed. New York, NY: Thieme-Stratton, 1985.

Difusa no "cluster" (10=difusa)	4
Formato do "cluster", de geométrico a irregular (10=irregular)	6
Radiação da microcalcificação ao longo do ducto (10=radiação proeminente)	3
Presença de macrocalcificações	
Associada com microcalcificações (10=macrocalcificação)	0
Macrocalcificações pela mama (10=macrocalcificação)	0

O método de treinamento da RNA usado pelos autores foi o "leave-one-out" para os casos retirados do atlas mamográfico que, segundo eles, não correspondem a um bom conjunto de treinamento para representar a prática, já que os casos do livro eram casos com certeza escolhidos por especialistas e muito bem definidos. Então eles modificaram o método de treinamento para um método convencional "round-robin", que foi aplicado para um grande conjunto de dados, incluindo os casos clínicos e os casos retirados do Atlas mamográfico, e avaliaram a performance do esquema para os dois conjuntos de dados separadamente. Para a avaliação, foi usada uma curva ROC que, com o conjunto de dados retirados do Atlas, resultou em $A_z = 1$. Isso significa que a RNA aprendeu todos os casos. Já com os casos clínicos a curva ROC resultou em $A_z = 0,95$.

Eles ainda compararam com a performance de radiologistas na análise dos mesmos casos empregados na RNA, e concluíram que as características das microcalcificações são bastante diferentes para as malignas e as benignas. Foi feito experimento também com um número reduzido de características, apenas 14 das 43 inicialmente utilizadas e eles comprovaram que o desempenho do classificador continuou semelhante.

Em ([VYBORNÝ-94]) são discutidas algumas técnicas que vêm sendo trabalhadas com mamografia. Técnicas de Visão Computacional e Inteligência Artificial vem sendo empregadas com sucesso em mamogramas para a detecção de características de anormalidades em imagens digital. Os autores citam ainda diversos trabalhos desenvolvidos nos anos 80 e início dos 90 que obtiveram bons resultados com sistemas computadorizados de classificação de microcalcificações e "clusters" em malignos e benignos, podendo concluir que as Redes Neurais Artificiais tem um desempenho melhor do que radiologistas nesta classificação, devido à capacidade de observação humana ser limitada pelo nosso próprio sistema de visão.

Em ([JIANG-96]) é mostrada uma técnica de classificação onde são extraídas informações de características de microcalcificações e seus agrupamentos a partir de mamogramas digitalizados. As imagens dos mamogramas digitalizados são processadas passando pela segmentação, e por um método computadorizado é feita a detecção dos "clusters" e microcalcificações, de onde são extraídas características como volume da microcalcificação, formato, densidade, quantidade de microcalcificações por "cluster" e área de cada microcalcificação. Depois de obtidas estas informações foi implementada uma Rede Neural Artificial que classifica os "clusters" em malignos ou benignos.

A Rede Neural Artificial é do tipo supervisionada e o algoritmo de aprendizagem utilizado foi o "Backpropagation". Para os testes foi usado o método "leave-one-out" que, segundo os autores é um método de teste bastante adequado quando se tem uma base de dados limitada e previamente conhecida.

O artigo mostra resultados satisfatórios com 95% de casos verdadeiros-positivos na classificação com as Redes Neurais Artificiais e este método de classificação tem potencial para ser usado clinicamente pois sua performance de classificação foi considerada superior à dos radiologistas. O resultados foram analisados com curvas ROC.

Em 1998 Yeli Jiang et al. apresentaram um trabalho ([JIANG-98]) onde propuseram um sistema de detecção de microcalcificações e classificação das microcalcificações agrupadas em "clusters" em malignas e benignas, para avaliar o desempenho da classificação em casos de sinais falsos-positivos. O objetivo do esquema era alertar o radiologista quanto à existência de "clusters" de microcalcificações em mamogramas.

A etapa de detecção foi dividida em três passos: primeiro foi aplicado um filtro linear nas imagens com as prováveis microcalcificações com o intuito de torná-las menos ruidosas; no segundo passo eles usaram um "Threshold" global e um local para identificar as microcalcificações na imagem; e no terceiro, foi feita uma análise das características para eliminar os sinais falsos-positivos, conforme está ilustrado na figura 5.2 (a). A figura 5.2 (b) ilustra o processo de classificação usado pelos autores, com uma Rede Neural Artificial MLP "backpropagation".

Como dados de entrada do treinamento da rede foram selecionadas oito características das microcalcificações detectadas, tais como número de microcalcificações em cada "cluster", circularidade e área de um "cluster", área média e volume médio (que significa área vezes a densidade efetiva) de cada microcalcificação, desvio padrão relativo ao contraste (ou densidade efetiva) e ao volume efetivo de cada uma das microcalcificações, e por último eles consideram uma medida de irregularidade de formato. As medidas de área e circularidade do "cluster" não dependem da localização individual das microcalcificações no interior do "cluster" de maneira que apenas com uma análise individual de cada microcalcificação da fronteira do "cluster", o esquema computadorizado pudesse estimar a probabilidade de malignidade do "cluster". As características extraídas foram escolhidas de acordo com a correlação qualitativas com os radiologistas

Eles usaram um banco de dados contendo 100 mamogramas de 53 pacientes, cujos "clusters" de microcalcificações foram confirmados com biópsia, contendo 40 mamogramas com microcalcificações malignas e 60 mamogramas com microcalcificações benignas.

A análise dos resultados deste sistema de classificação foi feita a partir de curvas ROC para medir a sua performance comparada à classificação de radiologistas experientes no assunto. Os resultados apresentados pelos autores mostram um bom desempenho das redes neurais, com $A_z = 0,92$ (área da curva ROC), e um acerto maior do que os radiologistas tanto na classificação de microcalcificações benignas quanto malignas, sendo que para os radiologistas e para as redes neurais a performance cai levemente na classificação das benignas.

Figura 5.2 - a) Esquema de classificação automático; b) detecção de “cluster” manual, respectivamente ([JIANG-98])

Em ([TAO-98]), as seguintes características são usadas para diferenciar microcalcificações malignas e benignas: as benignas geralmente são nebulosas com o centro mais claro, microcalcificações vasculares que aparecem como traços paralelos, formatos tubulares em linha (ou lineares) que são claramente associados com vasos ou ductos. Além daquelas grosseiras ou parecidas com “pipoca” em forma de barra contínua e que podem ser ramificadas. Já as malignas não apresentam formas definidas ou distintas, podem ser associadas a formas arredondadas ou em flocos, heterogêneas com variação de tamanho, normalmente menores que 0,5mm.

No trabalho, citam que formações com aparência nebulosa podem estar associadas com microcalcificações malignas, onde parece haver uma certa contradição, visto que o autor descreve esta característica anteriormente como sendo um atributo de microcalcificações benignas.

O sistema proposto naquele trabalho objetiva classificar “ROIs” em “indicado para biópsia” e “não indicado para biópsia”, diferente de muitos sistemas

computadorizados que classificam diretamente como benigno ou maligno. Segundo os autores, muitos médicos usualmente classificam os mamogramas dentro de 5 classes do BI-RADS - N (Normal), B (benigna), P(provavelmente benignas), S (suspeitas) e M (malignas) . Geralmente, pacientes diagnosticados dentro das classes N e B não são indicados para a biópsia, mas sim que estão nas classes M e S; pacientes pertencentes à classe P são aconselhadas a acompanhamento, ou seja, de acordo com os critérios clínicos e médicos, sugere-se repetir o exame de tempos em tempos (usual a cada seis meses).

Foi desenvolvido então um sistema classificador de pixels de microcalcificações e não microcalcificações e um classificador de regiões de interesse contendo microcalcificações. O primeiro destes dois classificadores é referido como um classificador de segmentação e o segundo como um classificador de diagnósticos. O sistema de diagnóstico depende da eficiência do sistema de segmentação das microcalcificações.

Neste trabalho eles compararam um sistema de diagnóstico automático com outro onde a segmentação e a detecção de microcalcificação são feitas manualmente (por médicos).

Para a aquisição de mamogramas, contaram com o apoio de vários laboratórios, sendo selecionados todos os casos que revelaram microcalcificações para o estudo. Foram usadas imagens de 512X512 pixels com 70 microns de resolução espacial e 12 bits na escala de níveis de cinza. Estas configurações foram usadas depois de uma breve investigação na literatura sobre a influência da resolução na detecção e classificação de microcalcificações, pois alguns grupos citam que mamogramas digitalizados com tamanho do pixel a partir de 100 microns não são adequados para detecção e classificação, uma vez que as microcalcificações são muito pequenas e não seria adequado para a extração de características de forma. O "UCSF/Livermore Laboratory" produz imagens com 42 microns; o banco de imagens do MIAS ("the Mammography Image Analysis Society") é digitalizado com 50 microns e no Departamento de Radiologia da UCLA a aquisição de imagens é feita com 65 microns.

A fase de detecção e segmentação se caracteriza por: remoção de baixas frequências espaciais e supressão de picos; suavização(smoothing) local e "thresholding" bimodal;

erosão e supressão de artefatos; crescimento de região gaussiana; extração de características.

Na tabela 5.2 estão dispostas todas as características extraídas dos mamogramas para o sistema de classificação.

Tabela 5.2 - Características extraídas de regiões de interesse para classificação ([TAO-98])

Nome da característica	Sigla	Fórmula	Significado Clínico
Área	A	N.º de pixel por microcalcificação (classificação ≥ 0)	N.º de pixel por microcalcificação (classificação ≥ 0)
Perímetro	P	N.º de pixel na fronteira central	N.º de pixel por microcalcificação (classificação ≥ 0)
Irregularidade	I	A/P^2	Fronteira irregular sugere anormalidade maligna
Valor de intensidade	MI	Quantidade de brilho do pixel	Microcalcificações brilhantes sugerem anormalidades benignas
Padrão de intensidade	SI	Divergência média do pixel	Grande SI sugere densidade irregular e malignidade
Intensidade exterior	ME	Média de brilho ao longo dos pixels da microcalcificação	Grande ME sugere textura densa e malignidade
Divergência de intensidade exterior	SE	Desvio padrão ao longo de 5 pixels de microcalcificação	Grande Sugere tecido maligno
Contraste	C	$ MI - ME /MI$	Grande contraste sugere anormalidade maligna
Limite	SH	Média da mudança de brilho entre os pixels internos e externos	Grande SH sugere anormalidades benignas
Contraste de superfície interna	B ₁	Média da mudança de brilho entre os pixels internos	grande B ₁ sugere anormalidades benignas
Contrastes central	B ₀	Média da mudança de brilho entre os pixels do centro da microcalcificação	Grande B ₀ sugere anormalidades benignas
Outros contrastes	B ₋₁	Média da mudança de brilho entre os pixels centrais e externos	Grande B ₋₁ sugere anormalidades benignas
Contraste exterior	B ₋₂	Média da mudança de brilho entre os pixels externos	Grande B ₋₂ sugere anormalidades benignas

Índice de concavidade	CI	Área da região entre as microcalcificações e sua concavidade	CI grande sugere malignidade
Forma	SG	$\frac{\int r(\theta) - R d\theta}{R}$ <p>Onde r é a distância radial da borda para o centróide, R é média de r, e θ é a orientação</p>	SG é uma medida do formato da microcalcificação, grande SG sugere textura benigna
Relação(proporção) de aspectos	R	proporção do máximo diâmetro central para o mínimo diâmetro central	Grande R sugere malignidade
Proporção do contraste	CS	C/SI	grande CS sugere microcalcificações verdadeiras
Numero de microcalcificações	N	numero de microcalcificações na região de interesse	Grande N sugere malignidade
Media da distancia entre as microcalcificações	D	Medida de valor alto quando as microcalcificações estão agrupadas.	pequena distancia sugere malignidade.

Eles selecionaram as melhores características utilizando duas técnicas de segmentação computadorizada e mais uma técnica onde a segmentação é feita manualmente. Algoritmo genético e K-NN ([FU-94]) foram usados na seleção das melhores características dentro daquelas já listadas na tabela e mais uma técnica onde a segmentação é feita manualmente, e a classificação foi feita com rede neural artificial que os autores nomearam "RNA - visual". A eficácia do sistema foi medida com curvas ROC.

O sistema foi dividido em três fases de detecção e segmentação, A(manual), B(5 métodos de segmentação citados anteriormente) e C (utilização de algoritmo genético e K-NN para seleção), correspondendo as técnicas de segmentação citadas acima e o resultado obtido foi de redução de casos para biópsia de 72% para A, 40% para B e 68% para C.

Por seu turno, Yu afirma no seu trabalho ([YU-98]), que a aparição de "cluster" de microcalcificação em mamogramas é um dos sinais mais importantes do câncer de mama. Este trabalho apresenta um novo sistema de processamento de imagens de detecção automática de "cluster" em mamogramas digitalizados cujo método de detecção usa transformada "wavelet" e RNA MLP, pixels de possíveis

microcalcificações e um conjunto de características situadas em microcalcificações individuais.

O método proposto traz dois passos principais: (1º) possíveis pixels de microcalcificações em mamogramas são segmentados com base em transformada “wavelet” para usar RNA MLP e rotulados dentro do potencial de microcalcificações individuais pela sua conectividade espacial; (2º) microcalcificações são identificadas com base nas características dos potenciais do objeto para uso de uma RNA MLP como um classificador.

Os autores extraíram 9 características dos mamogramas digitalizados com 12 bits de níveis de cinza e 100 microns de resolução espacial (tamanho do pixel). Os mamogramas resultaram no estudo de 105 “clusters” de microcalcificações, sendo retirados de 40 regiões de interesse, referentes a 21 pacientes.

A fim de reduzir o falso-positivo, os autores usaram as seguintes estratégias: média de níveis de cinza da região; desvio padrão dos níveis de cinza; nível de cinza do fundo de cada objeto e intensidade da borda de cada objeto. Para a extração de informações de contraste foram feitas três medidas: razão entre o fundo e o primeiro plano da imagem; diferença (primeiro plano - fundo) e razão da diferença. Para informações de forma: tamanho de cada objeto e alongação do objeto, que foi definida a partir da razão entre a distância mínima e a máxima entre as fronteiras do objeto.

Um curva FROC foi usada para avaliar a performance. Os resultados mostraram que uma significativa taxa de 90% na detecção é alcançado ao custo de um falso-positivo por imagem usando “wavelet” ortogonal.

Martí et. al., em ([MARTÍ-98]) descrevem uma outra classificação de “clusters” de microcalcificações em mamogramas, ainda nas categorias malignos e benignos.

Cada mamograma contendo um ou mais “clusters” de microcalcificações suspeitas e o estudo foi dividido em dois estágios, um retrospectivo e um prospectivo. No estágio retrospectivo foi usado um conjunto de 146 mamogramas com o objetivo de analisar as características de caráter maligno das microcalcificações e escolher as melhores características para construir um modelo de predição para diagnóstico de malignas. Todos eles tinham a comprovação do diagnóstico por biópsia e depois

foram comparadas com opiniões 3 radiologistas. Os mamogramas foram digitalizados com 600 DPI e com 12 bits de níveis de cinza. As imagens foram segmentadas usando técnicas de crescimento de região, cujo o primeiro pixel era escolhido por radiologistas experientes.

Foram extraídas inicialmente muitas características de microcalcificações individuais a partir da imagem segmentada, descritas na tabela 5.3 a seguir:

Tabela 5.3 - Características extraídas de regiões de interesse para classificação ([MARTÌ-98])

Características	Descrição
Área	Número de pixels em uma microcalcificação.
Perímetro	O total do comprimento da borda da microcalcificação
Compacidade (ou compactação ou solidez)	Derivada do perímetro P e área A de uma microcalcificação, isto é dado por: $\frac{P}{4\pi A}$
Box Min. X,Y; Box Max. X,Y	Coordenadas dos pixels extremos, esquerda, topo, baixo e direita, respectivamente, da microcalcificação
Janela (caixa, polígono, "Feret") X, Y	As dimensões das bordas mínima e máxima na horizontal e vertical de um polígono ou janela que enquadra uma microcalcificação, à partir das medidas anteriores
Diâmetro mínimo	Encontrar o diâmetro mínimo do polígono citado à cima depois de checar todos os ângulos (max = 64)
Diâmetro máximo	Encontrar o maior diâmetro usando o mesmo processo à cima
Diâmetro significativo	Média dos diâmetros encontrados para todos os ângulos checados
Alongação (polígono)	Medida de formato da microcalcificação $\frac{\text{Diâmetro máx.}}{\text{Diâmetro min}}$
N.º de orifícios	Número de orifícios da microcalcificação
Perímetro Convexo	Uma aproximação da convexidade da microcalcificação

Aspereza (rugosidade, irregularidade)	Medida de irregularidade	Perímetro
		Perímetro convexo
Comprimento	Uma medida do comprimento da microcalcificação	
Largura	Uma medida da verdadeira largura da microcalcificação	
Alongação	Comprimento / Largura	
Centróide X, Y	Posição X,Y do centro de gravidade da microcalcificação (medida extraída com Momento)	
Eixo principal	O ângulo pelo qual a microcalcificação tem o menor momento de inércia (eixo de simetria), para microcalcificações alongadas, isto é, alinhado com o eixo mais longo	
Eixo secundário	O ângulo perpendicular ao eixo principal	

Depois de aplicadas técnicas estatísticas para selecionar as características que melhor representavam o problema, o conjunto descrito na tabela 5.3 foi reduzido a: área, compacidade, números de orifícios por área de microcalcificação, alongação (polígono), irregularidade e eixo principal. Também foram adicionadas informações de número de “clusters” por mamograma, sendo “cluster” definido pelos autores como um agrupamento de 4 ou mais microcalcificações em uma área circular de 0,5 cm de diâmetro.

Para a classificação foi usado um modelo estatístico e obtiveram resultados significativos para concluir que as informações de formato de microcalcificação são parâmetros relevantes na classificação de uma estrutura mamográfica quanto à malignidade.

No estágio prospectivo, o modelo foi testado, utilizando 70 padrões. A performance do modelo estatístico foi avaliada comparando-se o resultado do modelo com a biópsia. Os resultados obtidos não foram tão satisfatórios, pois o modelo classificador teve um percentual maior de indecisão do que de acertos, embora o percentual de erros tenha sido baixo.

Um esquema computadorizado para detecção de “clusters” e classificação em “clusters” falsos-positivos e verdadeiros-positivos através de métodos estatísticos foi

desenvolvido por Veldkamp em ([VELDKAMP-99]). O sistema foi dividido em dois estágios : o primeiro de detecção e o segundo de classificação. A classificação é dada a partir de informações extraídas de microcalcificações individuais e características dos “clusters” quantificadas segundo as características das microcalcificações individuais.

A figura 5.3 mostra em diagrama de blocos o esquema desenvolvido pelos autores, cujo desempenho foi analisado com curvas FROC.

Figura 5.3 - Sistema Computadorizado de Veldkamp ([VELDKAMP-99])

A classificação em verdadeiro e falso-positivos é baseada em “clusters” com propriedades específicas. As características são derivadas das informações das microcalcificações individuais que compõem o “cluster”:

- 1- Perímetro, definido como o número de pixels que compõem as fronteiras da microcalcificação (um pixel é representado por um quadrado);
- 2- Área, representada pelo número de pixels da microcalcificação;
- 3- Densidade, definida como: $\frac{P^2}{4\pi A}$
onde P = perímetro e A = área;
- 4- Alongamento: definida como a razão $R_{\max}/R_{\min}$, ou seja, as distâncias máximas e mínimas entre as bordas e o centro de massa da

microcalcificação, onde o centro de massa é encontrado através de momento;

$$5- \text{ Excentricidade, definida como } e = \frac{I_{xx} + I_{yy} - \sqrt{(I_{xx} - I_{yy})^2 + 4 I_{xy}^2}}{I_{xx} + I_{yy} + \sqrt{(I_{xx} - I_{yy})^2 + 4 I_{xy}^2}}$$

onde I_{xx} , I_{yy} e I_{xy} são os momentos de inércia.

- 6- Espessura, calculado com a largura do melhor ajuste retangular;
- 7- Orientação, definida com o ângulo do eixo do menor momento de inércia respectivos ao plano xy;
- 8- Direção, calculada como a direção relativa na qual a microcalcificação está localizada, vista do centro de gravidade do “cluster”;
- 9- Linha, a média da saída do detector de borda *linha/borda* em uma microcalcificação detectada;
- 10- Fundo, o nível médio de intensidade do fundo;
- 11- Primeiro plano, a média de intensidade da microcalcificação detectada;
- 12- Distância, entre as calcificações vizinhas;
- 13- A medida C_i^{me} , dos pixels da microcalcificação com relação aos tecidos do fundo.

As características de “clusters” são relatadas conforme a distribuição das informações locais, descritas acima, no “cluster” detectado. Foi considerado $I = (I_1, I_2, \dots, I_k)^T$ como um vetor contendo as k características locais descritas para as microcalcificações.

Já as características de “clusters” utilizadas foram:

- 1- O valor médio da I_i característica local em um “cluster”, indicado como $m(I_i)$.
- 2- A mediana de I_i em um “cluster”, indicado por $med(I_i)$.
- 3- O desvio padrão de I_i em um “cluster”, indicado por $s(I_i)$.
- 4- O mínimo valor de I_i em um “cluster”, indicado por $min(I_i)$.
- 5- O máximo valor de I_i em um “cluster”, indicado por $max(I_i)$.

Além destas características, foram ainda definidas a área, representada pelo número de pixels de um “cluster”; e o número de calcificações em “cluster”.

As características foram selecionadas por um procedimento de seleção sequencial direta que, segundo os autores, é um método bastante rápido para se obter a lista das combinações de características válidas. Este método procura pelas características que maximizam a distância entre as classes.

Na classificação, para se obter uma FROC, é usado o método do "*leave-one-patient-out*", que consiste em deixar todos os "clusters" de um mesmo paciente fora do conjunto de treinamento, que será o paciente de teste. O conjunto de treinamento é sempre predefinido com o mesmo número "clusters" verdadeiros-positivos e falsos-positivos detectados. O nível de sensibilidade do primeiro estágio – detecção – é ajustado para variar a probabilidade inicial de um pixel ser parte da microcalcificação no modelo de campo aleatório de Markov. No estágio 2 a classificação é ajustada para considerar a taxa de falsos-positivos dos k vizinhos de um "cluster" de entrada. Se o número de verdadeiros-positivos fora dos k vizinhos exceder um certo "threshold", então um "cluster" é considerado como verdadeiro.

Os resultados foram apresentados com curvas ROC e o método avaliou a sensibilidade do sistema para diversas combinações de características.

Os autores concluíram que "clusters" verdadeiros-positivos tem alta sensibilidade, podem conter muitas microcalcificações falsos-positivas. Isto limita o poder de discriminação das características das microcalcificações individuais. Eles também descrevem que a informação de área do "cluster" é uma medida importante no aumento da sensibilidade da curva.

5.2 Utilização de Dados do Histórico do Paciente para a Classificação do Diagnóstico

Uma RNA foi desenvolvida para assistir radiologistas e sugerir na predição de achados não palpáveis como suspeitos. A RNA desenvolvida por Lo J. Y at al. em 1997 ([LOJY-97]), usa informação do histórico da paciente, a idade, e

informações das características da lesão descritas conforme o padrão BI-RADS. Eles selecionaram aleatoriamente 254 paciente que passaram por exame de agulhamento para localização de lesões não palpáveis para uma avaliação retrospectiva, resultando em 266 lesões com diagnóstico histológico. Destas 266 lesões, 170 eram benignas (64%) e 96 eram cânceres (36%). Radiologistas foram questionados para descrever 9 características morfológicas pertencentes às lesões. Os achados mamográficos foram reportados seguindo as descrições do BI-RADS. As primeiras três características mamográficas aplicadas para microcalcificações e calcificações associadas com massa: distribuição de calcificações, número e descrição. Outras quatro características aplicadas somente para massas: contorno da massa, formato, densidade, e tamanho. As duas últimas características poderiam ser usadas para todas as lesões: achados associados (axiliares, adenopáticas) e casos especiais (assimetria dos tecidos mamários). A décima e última característica foi a idade da paciente. Estas características foram extraídas antes do desenvolvimento da RNA. As características baseadas no BI-RADS foram convertidas em valores numéricos para serem aplicados à rede. A figura 5.4 mostra as características representadas numericamente e a apresentação delas à RNA.

Distribuição de Calcificação		Contorno da massa		Achados associados	
Ausência de Calcificações	0	Ausência de massa	0	nenhum	0
Difusa	1	Bem circunscrita	1	lesões na pele	1
Regional	2	Microlobulada	2	hematoma	2
Segmentar	3	obscura	3	pós cirurgica	3
Linear	4	Mal definida	4	Espessamento trabecular	4
Agrupada	5	Especulada	5	Retração do mamilo	5
No. de Calcificações		Tamanho da massa		Retração da pele	
Ausência de calcificações	0	Dada em milímetros		Espessamento da pele	7
<5	1	Formato da massa		Distorção anatômica	8
5 a 10	2	Ausência de massa	0	Axiliatres a demopáticos	9
>10	3	Redonda	1	Casos especiais	
Descrição de calcificações		Oval	2	Nemhum	0
Ausência de calcificações	0	Lobular	3	Nódulos linfáticos intramamários	1
Cálcio como de leite	1	Irregular	4	Assimetria dos tecidos da mama	2
Borda	2	Densidade da massa		Focos de densidade assimétricos	3
Pele	3	Ausência de massa	0	Densidade tubular ou isoladas em ductos dilatados	4
Vascular	4	Contendo gordura	1	Idade	
Esférico	5	Pouco densa	2	Dada em anos	
Sutura	6	Isodensa	3		
Áspero	7	Muito densa	4		
Grande, em forma de haste	8				
Redonda	9				
Distrófica	10				
Puntiforme	11				
Indistinguível	12				
Pleomorfismo	13				
Fina ramificação	14				

Figura 5.4 – quadro das características extraídas dos mamogramas ([LO JY-97])

As características foram apresentadas à RNA, os valores numéricos dos atributos extraídos, conforme mostra a figura acima, foram normalizados no intervalo $[0;1]$.

O modelo computacional desenvolvido foi uma RNA “Backpropagation” de três camadas (sendo apenas 1 camada escondida) que classificou em indicado para exame invasivo e câncer *in situ*. A rede foi treinada com o método “leave-one-out”, a rede usou 10 neurônios na camada de entrada, 15 na camada escondida. A avaliação do desempenho da rede foi feita através da área sobre a curva ROC, e não foi comparada com o desempenho dos radiologistas. Para 96 casos que passaram pela biópsia e eram malignos a rede teve um ótimo desempenho com área da curva ROC $A_z = 0.91$.

Com o objetivo de diminuir o número de biópsias, Fredederick e Floyd ([FREDERICK-98]) construíram um sistema que utiliza além das características extraídas da imagem mamográfica, também dados do histórico da paciente. A motivação deste trabalho, assim como para muitos dos sistemas aqui já citados, é que somente de 15 a 34% dos casos suspeitos de lesões não palpáveis submetidos ao exame de biópsia são comprovados com malignidade.

Eles desenvolveram uma RNA para predição de biópsia para o câncer de mama utilizando como base de dados o “Breast Imaging Reporting Data System” (BI-RADS; American College of Radiology, Reston, VA) e também dados do histórico clínico da paciente.

Os dados considerados para este estudo foram coletados de 478 pacientes, nas quais foram detectadas lesões não palpáveis. Destas pacientes resultaram 500 lesões confirmadas com biópsia, sendo 326 benignas e 174 malignas. O banco de dados então contém 174 casos malignos, 326 casos, benignos e idades das pacientes que varia de 24 a 86 anos, com média de 55,5 anos, de onde foram extraídos 11 atributos que serviram como entrada para o mapeamento de uma saída (benigna/maligna) por RNA. Estes 11 atributos estão listados na tabela 5.4.

Tabela 5.4 - Dados de Entrada da RNA a partir do banco de dados mamográfico

Dados de Entrada	Atributos
1	Distribuição das microcalcificações
2	Número de microcalcificações
3	Descrição da microcalcificação
4	Borda da massa
5	Tamanho da massa
6	Formato
7	Densidade
8	Características associadas
9	Casos especiais
10	Localização
11	Idade

O treinamento tradicional das RNA consiste em aplicar os pares de entrada na rede, calculando assim a diferença entre a saída obtida pela rede e a saída desejada. Este erro é então retropropagado, atualizando os pesos a fim de compensar o erro. Isto se repete até a rede ser considerada treinada, que é quando ela atinge os critérios de parada. A velocidade com que a rede aprende depende da taxa de aprendizagem. A otimização dos erros numa rede pelo algoritmo “Backpropagation” é feita pelo erro médio quadrático (MSE). A área das curvas ROC tem um maior significado como medida de performance do que o MSE, mas essa área não é uma função contínua, característica requerida para a função de otimização pelo treinamento “backpropagation”. Logo deve ser considerada uma aproximação diferente. Os autores usaram também um algoritmo genético ([FU-94]) para seleção de pesos ao invés do tradicional treinamento BP.

Como resultado do classificador, eles obtiveram performance medida com curvas ROC de área igual a 0,84, com especificidade variando entre 38 a 52% para os diferentes modelos de treinamento e configurações das RNA.

Um trabalho desenvolvido por Lo JY et al ([Lo JY-99]), mostra os efeitos dos dados históricos do paciente como características, juntamente com mamografias a serem consideradas por RNA para classificar quanto ao câncer de mama. Os autores avaliaram a contribuição dos dados do histórico médico da paciente para o

prognóstico do câncer de mama com um modelo RNA baseado em achados mamográficos.

Eles usaram três modelos de RNA para efeitos de comparação: primeiramente usaram 10 informações das imagens de mama e dados variados do sistema; o segundo experimento usou as informações das imagens de mama e os dados do sistema mais a idade do paciente; o terceiro usou, além das informações já usadas no segundo modelo, mais sete outras variáveis históricas, para um total de 18 entradas. A eficiência dos modelos e a opinião de radiologista foram avaliados com a área de uma curva ROC; sensibilidade de 100%, 98% e 95%; e os valores de prognóstico positivo.

O resultado dos modelos foram melhores que o parecer dos radiologistas nos três modelos pelas 5 métricas usadas na avaliação dos resultados. A variável “idade da paciente” foi particularmente muito importante. Ao adicionar a idade da paciente no modelo básico de RNA, que usava somente os dados do sistema e as informações das mamografias, obteve-se um significativo aumento de A_z . Nenhuma mudança significativa no valor de A_z ocorreu quando todos os dados de histórico da paciente foram substituídos apenas pela variável idade.

Os autores concluíram que a variável idade da paciente tem papel importante no prognóstico do câncer de mama pelas RNAs e dispensaria as demais variáveis do histórico completo do paciente.

Rogova et al., mostram em ([ROGOVA-99]) um sistema de Visão Computacional como parte de um esquema CAD que os autores chamaram de sistema híbrido, representado na figura 5.5. O sistema de visão computacional utiliza várias análises de imagens e métodos de compreensão que avaliam a performance da extração de atributos de maneira automática que representam diferentes regiões da imagem e a classificação das regiões selecionadas suspeitas para detectar malignidade.

Figura 5.5 - Esquema CAD híbrido ([ROGOVA-99])

Muitos aspectos nos problemas de caracterização de microcalcificação devem ser considerados para o desenvolvimento de um sistema de classificação baseado na intensificação das informações:

- um classificador tem que trabalhar com dados complexos e incertos;
- as características de microcalcificações são variadas e ruidosas;
- as imagens mamográficas têm uma estrutura de textura;
- informações estatísticas podem ser limitadas;

Considerando os aspectos acima, os autores construíram um método de evidências hierárquicas para caracterização de microcalcificações baseado na estrutura da teoria de evidências "Dempster-Shafer" (D-S). Esta teoria é uma ferramenta para representar as medidas de evidências, utilizando a representação da textura de cada microcalcificação individual de um "cluster".

A teoria "Dempster-Shafer" é uma generalização do raciocínio Bayesiano, sendo mais flexível que ele quando o conhecimento está incompleto, quando se tem que tratar com incertezas e ignorâncias. A teoria designa uma convicção que uma hipótese pode ser uma conjunção de muitas "hipóteses atômicas" (*esta microcalcificação ou é maligna ou é benigna*), que reflete muito bem o processo de avaliação das lesões pelos radiologistas. Então, a teoria permite representar somente o conhecimento verdadeiro e não forçar para uma opção incerta.

A idéia do classificador é baseada no simples fato de que “clusters” benignos têm apenas microcalcificações benignas, enquanto “clusters” malignos contêm microcalcificações com grau de certeza malignas e/ou benignas. Os autores não dispunham do resultado de biópsia para as microcalcificações individuais, mas sim para o “cluster” como um todo. Assim pôde-se obter evidências das classes associadas a microcalcificações individuais somente para benignas.

Num nível mais baixo, foi considerada uma imagem MxM contendo uma microcalcificação individual de um determinado “cluster”. Um conjunto de informações caracterizando a textura desta região é usado para computar as evidências de que a microcalcificação, contida na região, é benigna. A evidência é calculada pelo modelo D-S.

Depois de calculadas as evidências para todas as microcalcificações do “cluster” em questão, usavam as K evidências como entrada de uma RNA “Backpropagation” com K unidades de entrada e uma unidade de saída. O número K de evidências escolhidas como entrada para a RNA foi definido empiricamente. A saída da rede foi usada para computar a confiança da decisão da classificação. Esta confiança é usada em um “eleitor inteligente” construído para a combinação dos resultados do domínio de conhecimentos e do classificador de evidências hierárquicas dentro do sistema híbrido.

O domínio de conhecimento, assim como nomeado pelos autores, são informações clínicas da paciente.

O resultado do classificador hierárquico foi comparado com a análise de dois radiologistas. Nesta comparação, mostrou-se que a sensibilidade do classificador foi de 100% e foi maior do que a especificidade tanto do classificador quanto dos dois radiologistas.

5.3 Conclusão

A classificação de “clusters” de microcalcificações de mama em maligno ou benigno tem sido bastante estudada por pesquisadores não só da área médica, mas também tem envolvido profissionais de diversos ramos da ciência. Muitos métodos de classificação foram e estão sendo desenvolvidos.

Uma tendência é a classificação de estruturas mamográficas em suspeitas e não suspeitas com esquemas computacionais, pois classificar em benigno ou maligno, pode ser uma indicação perigosa diante da complexidade das características envolvidas neste problema da caracterização das lesões não palpáveis, principalmente quando não se tem muitas informações e acompanhamento de médicos no desenvolvimento do sistema.

Muitas técnicas de classificação vêm sendo empregadas na tentativa de auxiliar o diagnóstico precoce do câncer de mama, inclusive as RNA, que sua por vez tem mostrado bons resultados na classificação em esquemas CAD, onde os diversos métodos de treinamento mostram a flexibilidade de trabalho da técnica. Elas podem ser configuradas de acordo com os tipos de dados que estão disponíveis.

A resolução espacial, característica citada em muitas descrições dos trabalhos aqui apresentados, é importante na classificação de lesões mamográfica devido ao tamanho reduzido dos achados. O contraste e a escala de níveis de cinza também influenciam na performance dos CAD, já que o contraste e a resolução espacial influenciam conjuntamente na qualidade de uma imagem.

Muitos trabalhos mostram com sucesso a adição de dados do histórico clínico da paciente como parte do conjunto de informações que servirão de entrada de dados para o Classificador, que tem sido, por sua vez, elaborado com frequência baseado nas RNA, dados os bons resultados que essa ferramenta tem proporcionado.

6

Materiais e Métodos

Com base na literatura revisada até então, foi construído um esquema classificador com o objetivo de predição de casos suspeitos e não suspeitos, a partir da avaliação de agrupamentos de microcalcificações; além disso, para uma avaliação mais aprimorada do mamograma, consideramos também a influência das características de cada microcalcificação de um "cluster" isoladamente .

A técnica para classificação empregada foi a de RNA, pois podemos observar nos estudos bibliográficos que tem sido eficiente para resolver problemas de classificação, inclusive de achados mamográficos. As RNA são ferramentas conhecidas como aproximadoras universais de funções e são flexíveis quanto ao aprendizado e adaptáveis a novos problemas.

As imagens mamográficas digitalizadas foram adquiridas de um banco de imagens que vem sendo construído pelo grupo há algum tempo, cujos mamogramas originais são fornecidos pelo Hospital das Clínicas de Ribeirão Preto - FMRP USP e pela Santa Casa de São Carlos.

6.1 O Banco de Imagens

A base de dados utilizada divide-se em dois grupos de imagens com características de digitalização diferentes, o que serviu para avaliarmos também a influência de resolução espacial na classificação, além da eficiência de cada digitalizador.

- *Grupo de Imagens I* - Imagens digitalizadas com um "scanner" ótico UMAX PowerLook II com resolução espacial de 0,05mm e 8 bits de níveis de cinza.
- *Grupo de Imagens II* - Imagens digitalizadas com um "scanner" laser LUMISYS Lumiscan-50 com resolução espacial de 0,15mm e 8 bits de níveis de cinza.

As imagens digitais dos mamogramas completos foram recortadas seguindo orientações dos laudos dos mamogramas originais e selecionadas apenas regiões de interesse (ROIs).

O banco de imagens traz informações da técnica de digitalização, resolução, posição da exposição mamográfica (crânio caudal ou oblíqua), idade da paciente e observações dos laudos médicos, facilitando a busca das imagens que utilizamos durante todo o trabalho.

As imagens cujos laudos indicavam casos suspeitos ou não-suspeitos foram selecionadas, entre elas havia também imagens de mamas densas. Para a separação das imagens entre as classes, usamos apenas informações dos laudos médicos e não de exames histológicos. Muitas imagens referem-se a casos em que a paciente vem sendo acompanhada há mais de 10 anos, o que serviu de parâmetro para os testes do sistema.

Foram utilizadas no total 116 ROIs do *Grupo de Imagens I* (50% indicando casos suspeitos e 50% indicando não-suspeitos) e 86 do *Grupo de Imagens II* (50% a 50% também).

As figuras 6.1a e 6.1b ilustram ROIs do *Grupo de Imagens II* e *Grupos de Imagens I* respectivamente.

Figura 6.1 - a) ROI de um mamograma digitalizado no *scanner* laser e b) ROI de um mamograma digitalizado no *scanner* ótico.

6.2 Sistema de Detecção de Microcalcificações

Depois de seleccionadas e recortadas, as ROIs foram submetidas a um processamento para segmentação e detecção das microcalcificações, que é parte de um esquema de detecção de "clusters" ([NUNES-97]), desenvolvido previamente pelo grupo, resultando em imagens binárias, que foram a base para a extração de características tanto dos "clusters" como das microcalcificações individualmente. O esquema está descrito na figura 6.2 a seguir.

Figura 6.2- Esquema do sistema computacional desenvolvido para detecção de “clusters” de microcalcificações([NUNES-97]).

Cada ROI é submetida aos passos de pré-processamento e processamento para a detecção efetiva das microcalcificações e dos “clusters”.

A segmentação é o processo onde as estruturas de interesse devem ser reforçadas e os ruídos e artefatos devem ser eliminados da imagem. Os passos dessa etapa podem ser visualizados na figura 6.3.

Figura 6.3 - Esquema utilizado para segmentação das microcalcificações [NUNES-97].

1. Subtração da imagem: é um pré-processamento com o objetivo de incrementar a relação sinal-ruído. Consiste em realizar a subtração de uma imagem com os sinais realçados de uma imagem com os sinais suavizados.
2. "Thresholding" global: esse processamento deixa na imagem somente uma porcentagem dos pixels com os maiores níveis de cinza. O objetivo deste passo é eliminar o fundo da imagem.
3. Morfologia matemática: tem o objetivo de eliminar da imagem todos os sinais com menos de três pixels na sua composição, visto que, provavelmente, relacionam-se a ruídos da imagem.
4. "Threshold" local: é o último procedimento. Consiste em eliminar os sinais falso-positivos restantes da imagem, através da verificação do contraste de cada estrutura em relação à sua vizinhança ([NUNES-97]).

As figuras 6.4a e 6.4b ilustram as ROIs antes e depois da segmentação respectivamente.

Figura 6.4 – a) ROI antes de ser processada com 256 níveis de cinza e b) ROI processada, imagem binária.

6.3 Esquema Classificador

O esquema representado na figura 6.5 mostra as principais etapas deste trabalho.

Figura 6.5 - Etapas do esquema classificador implementado

A partir das imagens segmentadas pelo sistema descrito do item 6.2, partimos então para a extração das características dos achados mamográficos. Cada bloco do esquema mostrado na figura 6.5 está descrito a seguir.

- 1- Extração de características: consiste em uma análise detalhada da imagem, percorrendo-a pixel a pixel. Nesta fase foram extraídas características de forma, usando-se descritores matemáticos como momentos, medidas de irregularidade, compacidade, área e perímetro, conforme vistos na literatura ([GONZALES-93], [MATERS-94], [PRATT-91], entre outros). A base de dados foi organizada em Banco de Dados para facilitar as etapas seguintes.
- 2- Na Seleção de atributos usou-se distribuição gaussiana para analisar o comportamento das duas classes, "suspeito" e "não-suspeito", para escolhermos os atributos mais significativos.
- 3- Organização dos vetores de entrada foi feita com o intuito de representar os exemplos das duas classes da melhor maneira possível, foram combinadas as características que mais se distanciavam nas classes e aplicadas nas entradas da RNA, definindo-se, assim, a arquitetura da rede (número de neurônios de entrada, na camada intermediária e de saída).
- 4- Implementação da RNA foi desenvolvida em *Delphi 4.0* devido aos recursos visuais e gráficos, pois desejávamos uma interface amigável para que os dados fossem inseridos na rede de maneira automática, utilizando recursos de Banco de Dados da linguagem *Delphi*.
- 5- O treinamento da RNA utilizou o algoritmo "Backproagation"; a interface possibilitou que parâmetros como taxa de aprendizagem, *momento*, mínimo erro e variações na arquitetura das camadas intermediárias fossem alterados na tentativa de melhorar a performance do treinamento, o que agilizou o processo. Além disso, utilizamos dois modos de treinamento, o modo "padrão" com a base de dados dividida em conjunto de treinamento e de testes e o método "leave-one-out".
- 6- Na fase Classificação, os testes foram feitos apenas com treinamentos que alcançaram os melhores desempenhos, ou seja, com os menores erros mínimos.

6.3.1 Extração de Características

A partir da imagem binária, as características foram extraídas das ROIs que apresentavam mais de três sinais (definição de "cluster" ([SICKLES-86])).

A princípio fizemos um experimento onde consideramos cada microcalcificação do "cluster" individualmente, ou seja, cada microcalcificação foi analisada como um objeto dentro de uma cena (o próprio cluster). Calculamos área, perímetro e as outras duas medidas derivadas destes atributos que são compacidade e irregularidade. Para os "clusters" aos quais pertenciam as microcalcificações, extraímos informações de número de microcalcificações e então o valor médio das áreas, perímetros, compacidades e irregularidades das microcalcificações de cada "cluster". Nesta análise foram usados apenas exemplos do *Grupo de Imagens I*, e a quantidade de exemplos foi reduzida, cerca de 24 ROI, sendo 9 da classe suspeita e 15 da não suspeita.

O algoritmo a seguir mostra o procedimento que adotamos.

Algoritmo:

- 1- Encontrar o primeiro pixel da microcalcificação (branco);
- 2- Verificar as possíveis conexões com outro pixel branco (conectividade 8 e conectividade 4 ([GONZALES-90]));
- 3- Contar pixels conectados;
- 4- Área recebe número de pixels conectados;
- 5- Verificar pixels pertencentes a borda ([MASCARENHAS-89]);
- 6- Perímetro recebe número de pixels da borda;
- 7- Calcular Compacidade ($C := \text{Perímetro} / (4\pi \text{Área})$);
- 8- Calcular Irregularidade ($I := \text{Área} / \text{Perímetro}^2$);

Apenas sinais (possíveis microcalcificações) com áreas maiores que 1 pixel foram considerados para a análise, a fim de evitar o falso-positivo eliminando os ruídos. A figura 6.6 mostra a interface desta parte do sistema para extração de características de microcalcificações.

Figura 6.6 - Interface do procedimento de extração de características de microcalcificações.

Para delinear o “cluster”, são identificados os pixels brancos nas extremidades da ROI que servem como vértices do retângulo que delimita o agrupamento, este retângulo é o objeto em análise. A figura 6.7 mostra o “cluster” já demarcado.

Figura 6.7 – “cluster” delimitado por um retângulo.

Depois de analisados os atributos das microcalcificações individualmente nas duas classes e delineados os "clusters", partimos então para extração de características, considerando o agrupamento como um único objeto, sendo as microcalcificações pontos do objeto. Nesta fase mais características foram extraídas. Outros descritores de formato como momentos foram calculados com o intuito de representar cada padrão adequadamente, pois os momentos consideram toda a massa da figura (objeto), diferente dos descritores de forma baseados em perímetro, que consideram detalhes mais finos e por isso são bastante sensíveis a ruído. Em todos os "clusters" aplicamos a teoria dos momentos e extraímos medidas que descrevem a representação de cada objeto.

Baseando-nos nas linhas do retângulo que delinea o agrupamento, calculamos mais dois atributos que chamamos de "PerímetroCluster" e "ÁreaCluster". Esta medida de perímetro foi usada juntamente com a Área (M_{00}) para calcular Irregularidade (I) e Compacidade (C) nos "clusters". Todas as características extraídas tanto na primeira como na segunda análise estão listadas na tabela 6.1. Cada uma destas características está descrita no capítulo 3. A figura 6.8 mostra a interface do procedimento desenvolvido para a extração das características dos "clusters" nesta segunda etapa.

Já nesta fase usamos 116 ROIs do *Grupo de Imagens I*, sendo 58 suspeitas e 58 não suspeitas, e 86 ROIs do *Grupo de Imagens II*, sendo 43 suspeitas e 43 não suspeitas.

Figura 6.8 -Interface do procedimento de extração de características de "clusters".

Tabela 6.1 – Características extraídas dos "clusters"

Característica	Objeto
Área (A)	Microcalcificação , "cluster"
Perímetro da microcalcificação (P)	Microcalcificação
ÁreaCluster	"cluster"
PerímetroCluster	"cluster"
Hu1(ϕ_1)	"cluster"
Hu2(ϕ_2)	"cluster"
Hu3(ϕ_3)	"cluster"
Hu4(ϕ_4)	"cluster"
Hu5(ϕ_5)	"cluster"
Hu6(ϕ_6)	"cluster"
Hu7(ϕ_7)	"cluster"
Momento ₀₀	"cluster"
Momento ₁₀	"cluster"
Momento ₀₁	"cluster"

Momento ₂₀	"cluster"
Momento ₀₂	"cluster"
Momento ₃₀	"cluster"
Momento ₀₃	"cluster"
Momento ₁₁	"cluster"
Momento ₁₂	"cluster"
Momento ₂₁	"cluster"
Irregularidade	Microcalcificação, "cluster"
Compacidade	Microcalcificação, "cluster"
RG	"cluster"

No "cluster" a área é representada pelo Momento₀₀ utilizado para calcular os invariantes momentos de Hu, Raio de Giração e Irregularidade dos "clusters", cujos equacionamentos foram apresentados no capítulo 3.

Além das características já citadas na tabela 6.1, inserimos também mais duas características de informações retiradas do laudo das imagens: a idade da paciente, que tem sido usada na literatura ([FREDERICK-98], [LOJY-97], [LOJY-98], [LOJY-99], [ROGOVA-99], entre outros), e um parâmetro que indicava se a mama era densa. Para mamas densas a característica "densificação" recebeu o valor 1 e para imagens cujos laudos não se referiam a esta propriedade, "densificação" recebeu o valor 0.

6.3.2 Seleção de Atributos

Nesta fase, não utilizamos nem um método sofisticado para escolher os melhores atributos, pois o objetivo era testar os atributos mais usados na literatura com a nossa base de dados que tem peculiaridades de resolução e técnicas de digitalização e também testar características globais que representassem o objeto "cluster", e não uma cena com vários objetos (microcalcificações).

Mesmo assim, construímos curvas de distribuição Gaussiana para todas as características extraídas ([PATROCINIO-2000]).

As curvas gaussianas foram construídas para mostrar a distribuição das classes para cada característica particularmente.

As curvas Gaussianas foram primeiramente construídas para as características extraídas das microcalcificações consideradas uma a uma. Para os atributos dos clusters também foram construídas as curvas Gaussianas, para os dois grupos de imagens.

Alguns exemplos estão nas figuras 6.9 a 6.14 que mostram algumas das curvas que analisamos para selecionar os atributos que serviriam de entrada para o classificador.

Figura 6.9 - Distribuição das áreas do "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens I.

Figura 6.10 - Distribuição das áreas do "clusters" das duas classes, suspeita (S) representada pela curvas vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens II.

Figura 6.11 - Distribuição dos Invariantes Momentos de Hu 1 dos "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens I.

Figura 6.12 - Distribuição dos Invariantes Momentos de Hu 1 dos "clusters" das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens II.

Figura 6.13 - Distribuição das idades das pacientes das duas classes, suspeita (S) representada pela curva vermelha e não-suspeito (NS) representada pela curva azul para o Grupo de Imagens I.

Figura 6.14 - Distribuição das idades das pacientes das duas classes, suspeita (S) representada pela curva vermelha e não- suspeito (NS) representada pela curva azul para o *Grupo de Imagens II*.

Informações provenientes da literatura nos ajudaram na escolha das características. Em [SHEN-93], foi apresentado um esquema CAD cujos momentos de Hu foram as características mais relevantes, enquanto vários outros trabalhos usam área, irregularidade e compacidade como características ([TAO-89], [MARTÍ-98]).

As características foram escolhidas analisando as curvas Gaussianas e testando os atributos em uma simulação feita no *Matlab* ([PATROCINIO-2000]). O resultado foi a seleção de 10 características como listadas a seguir:

1. Invariante Momento de Hu 1 (ϕ_1);
2. Invariante Momento de Hu 2 (ϕ_2);
3. Invariante Momento de Hu 3 (ϕ_3);
4. Invariante Momento de Hu 4 (ϕ_4);
5. Raio de Giração (RG);
6. Irregularidade (I);
7. Compacidade (C);
8. Área (M_{00});
9. Perímetro do "Cluster" (soma dos lados do retângulo que delinea o "cluster");
10. Área do "Cluster" (área do retângulo que delinea o "cluster") .

6.3.3 Organização dos Vetores de Entrada da RNA

Depois de todas as características extraídas e selecionadas, percebemos a necessidade de normalização dos valores dos atributos, pois eles se diferenciavam bastante de uma característica para outra e a faixa de valores era bastante ampla, o que não é indicado para aplicar como vetor de entrada em um classificador, exigindo então uma normalização dos dados ([DUDA-73], [FUKUNAGA-90], [MASCARENHAS-89]). A normalização foi feita de maneira que todos os valores ficassem entre -1 e 1, dividindo-se os valores pelo módulo do maior, para cada característica.

6.3.4 Implementação da RNA

Uma RNA supervisionada “Multilayer Perceptron” foi implementada seguindo todos os requisitos necessários para a construção de um classificador neural. A necessidade de definição de arquitetura, algoritmo de aprendizagem, escolha da base de dados que representassem bem o problema e a divisão dos exemplos para treinamento e teste é o que se chama na literatura de requisitos para o desenvolvimento destes sistemas de inteligência artificial.

Devido à necessidade de uma interface amigável, o algoritmo foi desenvolvido em *Delphi 4.0*, que é uma linguagem que oferece muitos recursos visuais, embora tenhamos utilizado o *Matlab 5.0* para os primeiros experimentos.

A arquitetura da rede foi definida com duas camadas escondidas, uma camada de entrada e uma 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. A saída também foi variada com um e dois neurônios.

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, usando o algoritmo "Backpropagation". A função de ativação usada foi a sigmóide.

A interface que desenvolvemos possibilitou escolher os números de neurônios e a configurar os parâmetros de treinamento, como a taxa de aprendizagem e definição do erro mínimo. A figura 6.15 mostra a interface usada para o treinamento da rede.

Figura 6.15 - Tela do procedimento de treinamento

Depois de treinados, os pesos da RNA foram guardados em arquivos do tipo texto para serem usados nos testes onde a rede classificou os padrões que não entraram no conjunto de treinamento.

6.3.5 Treinamento da RNA

O treinamento foi dividido em várias etapas. Numa primeira experiência em que a base de dados contava com apenas 44 imagens, 22 com laudos suspeitos e 22 com laudos não-suspeitos (todos os exemplos do Grupo de Imagens I), foram extraídos os primeiros 4 momentos de Hu mais irregularidade e compacidade de todos os padrões. O conjunto de imagens foi dividido da seguinte maneira: 50% de toda base de dados foi usada no treinamento, sendo 25% suspeitos e os outros 25% não-suspeitos, e os 50% que não foram usados no treinamento serviram para a fase de testes. Com essa divisão da base de dados, usamos o modo "padrão"¹ de correção de pesos durante o treinamento.

Numa segunda experiência, usando a mesma base de dados da primeira e as mesmas características, usamos o método "leave-one-out", em que para uma base de dados com N exemplos a rede é treinada e testada $N-1$ vezes. Então, para o treinamento, usamos 43 exemplos e a rede foi testada com o exemplo que ficou de fora. Isso se repetiu até que todos os exemplos fossem testados.

Nestas duas primeiras etapas a rede foi implementada no *Matlab*.

Com a rede implementada em *Delphi*, partimos então para uma base de dados com maior quantidade de exemplos e a apresentação de exemplos mais detalhados e com mais características. Nesta fase, cada padrão foi representado por 10 características de forma, já citadas no item 6.3.1 deste capítulo. O número de padrões foi 116 do *Grupo de Imagens I* e 86 do *Grupo de Imagens II*. Toda a base de dados foi dividida diferentemente, ficando 70% de cada um dos grupos de imagens para a

¹ No modo "padrão" a correção dos pesos ocorre a cada apresentação de um exemplo a rede.

fase de treinamento e os outros 30% para a fase de teste. Aqui também foi usado o modo "padrão" de correção dos pesos.

Nos últimos experimentos incluímos duas novas características, idade da paciente e uma referência quanto à "densificação" da mama. A rede foi treinada e testada incluindo as duas características separadamente e depois juntas.

Para cada experiência citada acima, treinamos várias configurações do número de neurônios nas camadas escondidas também.

6.3.6 Classificação

Uma interface para a classificação também facilitou e agilizou os testes.

O classificador recupera as informações guardadas no arquivo texto (pesos e limiares) com os 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.

Na classificação, o usuário pode ou não usar as informações do histórico da paciente e de "densificação", bem como escolher qual a arquitetura que ele deseja usar do classificador, basta escolher entre as configurações treinadas. A interface oferece estas opções. A interface da classificação é mostrada na figura 6.16.

Figura 6.16 - Tela do procedimento de classificação

Os testes foram feitos e avaliados calculando-se as porcentagens de verdadeiro-positivo, verdadeiro-negativo, falso-positivo e falso-negativo, o que possibilitou a análise de acertos e erros do classificador, calculando também a probabilidade de erro.

Para mostrar o desempenho do classificador, construímos curvas ROC, o que facilitou a apresentação e a avaliação dos resultados.

Adotamos o ponto de operação da curva ROC igual a 0,5 para calcularmos sensibilidade e especificidade de cada uma das curvas construídas.

7

Resultados e Discussões

7.1 Resultados Preliminares

A partir dos primeiros resultados que obtivemos, usando um simulador do *Matlab*, foi possível observar a eficiência da técnica de Redes Neurais Artificiais, avaliar a base de dados e os pré-requisitos para construção da rede, além de conhecermos as finalidades dos parâmetros envolvidos na fase de treinamento, ajudando na compreensão do algoritmo "Backpropagation" para posterior implementação.

Nas tabelas 7.1 e 7.2 estão descritos os resultados que obtivemos na classificação dos "clusters" de microcalcificações nessa fase. Os resultados descritos na tabela 7.1, correspondentes ao modo de treinamento padrão (conforme descrito no capítulo 3), e com a base de dados dividida em dois conjuntos (um com metade do número de exemplos, usado para o treinamento da RNA, e outro com a segunda metade dos exemplos, usado para os testes) mostraram a necessidade de ampliarmos a base de dados. O conjunto de treinamento era bastante restrito, contendo apenas 17 ROIs, o que nos levou a optar pelo treinamento "leave-one-out" que, segundo a literatura, vem sendo uma boa alternativa para o treinamento de RNA quando o número de exemplos é reduzido. Isso nos possibilitou verificar as variações da

técnica em relação à divisão e uso da base de dados para as fases de treinamento e teste do classificador.

Tabela 7.1- Resultados obtidos com o modo padrão de treinamento

	VP (%)	FN(%)	VN(%)	FP(%)	Acerto Total(%)
Teste15	37,5	12,5	41,7	8,3	79,2
Teste20	37,5	12,5	41,7	8,3	79,2
Teste25	33,3	16,7	41,7	8,3	75
Teste30	33,3	16,7	41,7	8,3	75
Teste35	12,5	37,5	37,5	12,5	50

7.2 - Resultado obtido com treinamento "leave-one-out"

	VP (%)	FN (%)	VN (%)	FP (%)	Acerto Total (%)
Teste15	42,5	7,5	40,0	10,0	82,5

Nas tabelas acima, as siglas VP, FN, VN e FP correspondem às possibilidades de resposta do sistema ou observador: verdadeiro-positivo (VP) para respostas verdadeiras em relação aos casos positivos, falso-negativo (FN) para respostas falsas em relação aos casos positivos, verdadeiro-negativo (VN) para respostas verdadeiras em relação aos casos negativos e falso-positivo (FP) para respostas falsas em relação aos casos negativos. Estes valores informam acertos e erros no diagnóstico, tendo sido obtidos a partir da porcentagem em relação ao total de exemplos utilizados na fase de testes.

A soma dos erros mínimos quadrados considerados na fase de treinamento para os resultados mostrados na tabela 7.1 e 7.2 foi estipulada em 0,01.

Na figura 7.1 a seguir, fica clara a influência do número de neurônios no desempenho da rede na classificação. Nesse experimento variamos de cinco em cinco o número de neurônios na camada escondida da rede. Embora o desempenho do classificador com número maior de neurônios na camada escondida tenha sido insatisfatório, isso não significa que se aumentássemos para 100 o número de neurônios o desempenho seria pior, pelo contrário, poderia acontecer de melhorar o desempenho da rede na classificação. Conforme a literatura, não existe relação direta entre o número neurônios e a performance do sistema, embora existam pesquisadores

estudando maneiras de estabelecer o número ideal de neurônios nas camadas escondidas [BRASIL - 99].

Figura 7.1 Desempenho das redes em função do número de neurônios.

Pelos resultados obtidos nessa simulação foi possível perceber a eficiência da RNA quando usada para classificação, embora nesta fase contássemos com uma base de dados ainda restrita. Além disso, não trabalhamos com exemplos comprovados histologicamente, o que dificultava ainda mais a separação das duas classes.

Contudo, 82,5% de acerto total na predição dos "clusters", em suspeitos e não-suspeitos, é um bom índice quando comparado com os resultados encontrados na literatura, como em [TAO-98] com 72%, [SHEN-93] com 87% para os casos malignos e 94% para os casos benignos, [BAKER-95] com sensibilidade igual a 95% e especificidade de 62%, [NAGEL-98] com 83% de sensibilidade, entre outros; e também dadas as taxas de erros dos diagnósticos médicos enviados para a biópsia, em que mais de 60% dos casos resultam negativos.

7.2 Resultados do Grupo de Imagens I

Nos testes referentes a Segunda etapa, para consecução do trabalho, além de porcentagem, também usamos curvas ROC para análise, o que facilitou a comparação da performance do nosso classificador com outros sistemas CAD da literatura, conforme apresentados no capítulo 5.

Aqui, os resultados foram obtidos com a RNA implementada em *Delphi 4.0* e usando vários atributos como entrada para a rede, conforme mencionado no capítulo anterior.

Os testes foram feitos considerando quatro diferentes grupos de atributos: apenas atributos de descritores de forma, descritores de forma e a idade da paciente, descritores de forma e a "densificação" e descritores de forma mais a idade e mais a "densificação".

Primeiro foram analisados resultados da RNA com 15 neurônios na primeira camada escondida e 5 na segunda camada escondida, depois os resultados da RNA com 30 neurônios na primeira camada escondida e 10 neurônios na segunda camada escondida.

As tabelas 7.3 e 7.4 mostram alguns dos resultados obtidos dos testes feitos com o *Grupo de Imagens I*.

Tabela 7.3 - Índices obtidos na análise da curva ROC para o *Grupo de Imagens I*

	NÚMERO DE NEURÔNIOS	VP	FP	VN	FN	ERRO	Az
Descritores de forma	15 – 5	0,833	0,12	0,88	0,167	0,1435	0,92299
Descritores de forma	30 – 10	0,833	0,277	0,723	0,167	0,222	0,84483
Descritores de forma + Idade	15 – 5	0,833	0,22	0,78	0,167	0,1935	0,87662
Descritores de forma + Idade	30 – 10	0,77	0,277	0,723	0,23	0,2535	0,81703
Descritores de forma + Densificação	15 – 5	0,833	0,22	0,78	0,167	0,1935	0,87662

Descritores de forma + Densificação	30 – 10	0,77	0,33	0,67	0,23	0,28	0,77904
Descritores de forma + Idade + Densificação	15 – 5	0,833	0,5	0,5	0,167	0,3335	0,73491
Descritores de forma + Idade + Densificação	30 – 10	0,833	0,38	0,62	0,167	0,2735	0,78613

Tabela 7.4 - Índices obtidos na análise com porcentagens para o Grupo de Imagens I

	NÚMERO DE NEURÔNIOS	VP(%)	FP(%)	VN(%)	FN(%)	ACERTO TOTAL(%)
Descritores de forma	15 – 5	41,65	6	44	8,35	85,65
Descritores de forma	30 – 10	41,65	13,85	36,15	8,35	77,8
Descritores de forma + Idade	15 – 5	41,65	11	39	8,35	80,65
Descritores de forma + Idade	30 – 10	38,5	13,85	36,15	11,5	74,65
Descritores de forma + Densificação	15 – 5	41,65	11	39	8,35	80,65
Descritores de forma + Densificação	30 – 10	38,5	16,5	33,5	11,5	72
Descritores de forma + Idade + Densificação	15 – 5	41,65	25	25	8,35	66,65
Descritores de forma + Idade + Densificação	30 – 10	42	19	31	8,35	72,65

A sensibilidade é um parâmetro que indica acertos em casos positivos e especificidade indica acertos em caso negativos. Estes parâmetros representam uma das maneiras de quantizar os diagnósticos VP e VN respectivamente. O cálculo destes parâmetros foi apresentado no capítulo 3.

A área sob a curva ROC representa o índice de acerto do 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, o que justifica uma área maior e valor próximo de 1. Pela análise, quando o verdadeiro-positivo é alto, o

falso-negativo deve ser baixo, pois o sistema ideal deve ter probabilidade de falso-positivo e falso-negativo igual a zero; por isso no sistema real, obviamente deseja-se que estes índices sejam os menores possíveis.

A medida de precisão dá uma informação de confiabilidade do sistema. Ela é definida pela razão entre o número de decisões corretas e o número de casos analisados.

Os valores mostrados na tabela 7.3 foram obtidos a partir da construção das curvas ROC. Esses valores podem variar de acordo com o critério adotado, como já citado no item 3.2.1 do capítulo 3. Os gráficos mostrados nas figuras a seguir ilustram melhor os resultados da tabela 7.3.

(a)

(b)

Figura 7.2 - Curvas ROC do classificador considerando apenas descritores de forma como atributo: a) treinado com 20 neurônios e b) treinado com 40 neurônios.

(a)

(b)

Figura 7.3 - Curvas ROC do classificador considerando descritores de forma e idade como atributos: a) treinado com 20 neurônios e b) treinado com 40 neurônios.

(a)

(b)

Figura 7.4 - Curvas ROC do classificador considerando descritores de forma e "densificação" como atributos: a) treinado com 20 neurônios e b) treinado com 40 neurônios.

(a)

(b)

Figura 7.5 - Curvas ROC do classificador considerando descritores de forma, idade e "densificação" como atributos: a) treinado com 20 neurônios e b) treinado com 40 neurônios.

Pelos resultados apresentados nas figuras 7.2 a 7.5 é possível perceber que os melhores desempenhos conseguidos com o *Grupo de Imagens I* foram aqueles em que utilizamos como entrada para a RNA apenas os atributos dos descritores de forma. Por um lado, isto justifica a eficiência dos descritores de forma, momentos, e mais ainda dos Invariantes de Hu. Por outro lado, um atributo relevante analisado na

prática médica, como a idade, para este conjunto de exemplos, não influenciou no desempenho do classificador de maneira significativa. Isso também é verdade para o parâmetro de "densificação", embora uma melhor discretização deste parâmetro deva ser melhor estudada, de maneira a quantizar esta informação baseando em características obtidas pela variação dos níveis de cinza em um mamograma ([GRAM-97]), o que pode indicar que a densificação seja um atributo de maior peso para a classificação.

Mesmo assim, a segunda maior A_z que obtivemos para os resultados acima foi combinando os atributos: os descritores de forma e a idade e o mesmo resultado foi obtido combinando descritores de forma e a "densificação".

As curvas ROC construídas com os resultados obtidos para o *Grupo de Imagens I* mostraram os diferentes comportamentos do classificador para cada conjunto de atributos.

7.3 Resultados do Grupo de Imagens II

Para o *Grupo de Imagens II*, os resultados também foram analisados a partir de curvas ROC e porcentagem. Nas tabelas 7.5 e 7.6 estão representados os valores obtidos da análise. O conjunto de testes neste grupo de imagens era composto por 48% de casos "suspeitos" e 52% de casos não suspeitos.

Primeiro foram analisados resultados da RNA com 25 neurônios na primeira camada escondida e 10 na segunda camada escondida, depois os resultados da RNA com 35 neurônios na primeira camada escondida e 10 neurônios na segunda camada escondida.

Tabela 7.5 - Índices obtidos na análise da curva ROC para o Grupo de Imagens II.

	Número de neurônios	VP	FP	VN	FN	Erro	A_z
Descritores de forma	25 – 10	1	0,1538	0,8462	0	0,19968	0,96915

Descritores de forma	35 - 10	1	0,1538	0,8462	0	0,12	0,96915
Descritores de forma + Idade	25 - 10	1	0,30769	0,69231	0	0,159999	0,93626
Descritores de forma + Idade	35 - 10	1	0,3846	0,6154	0	0,199992	0,91993
Descritores de forma + Densificação	25 - 10	1	0,384	0,616	0	0,079976	0,92072
Descritores de forma + Densificação	35 - 10	1	0,23076	0,76923	0	0,079976	0,94856
Descritores de forma + Idade + Densificação	25 - 10	1	0,69	0,31	0	0,3588	0,84069
Descritores de forma + Idade + Densificação	35 - 10	1	0,53	0,47	0	0,2756	0,8771

Tabela 7.6 - Índices obtidos na análise com porcentagens para o Grupo de Imagens II

	Número de neurônios	VP(%)	FP(%)	VN(%)	FN(%)	Acerto Total(%)
Descritores de forma	25 - 10	48	7,9976	44,0024	0	92
Descritores de forma	35 - 10	48	7,9976	44,0024	0	92
Descritores de forma + Idade	25 - 10	48	15,9998	36,0001	0	84
Descritores de forma + Idade	35 - 10	48	19,9992	32,0008	0	80
Descritores de forma + Densificação	25 - 10	48	19,968	32,032	0	80
Descritores de forma + Densificação	35 - 10	48	11,9999	40	0	88
Descritores de forma + Idade + Densificação	25 - 10	48	35,88	16,12	0	64,1
Descritores de forma + Idade + Densificação	45	48	27,56	24,44	0	72,4

Os valores acima tabelados mostram as diferenças nos desempenhos das redes treinadas com diferentes números de neurônios e diferentes combinações de atributos de entrada para o classificador. As curvas ROC representadas nas figuras 7.6 a 7.9 ilustram estas diferenças. Isso também pode ser visto nos gráficos construídos para os testes com o *Grupo de Imagens II*.

(a)

(b)

Figura 7.6 - Curvas ROC do classificador considerando apenas descritores de forma como atributos: a) treinado com 35 neurônios e b) treinado com 45 neurônios.

(a)

(b)

Figura 7.7 - Curvas ROC do classificador considerando descritores de forma e idade como atributos: a) treinado com 40 neurônios e b) treinado com 50 neurônios.

(a)

(b)

Figura 7.8 - Curvas ROC do classificador considerando descritores de forma emais densificação como atributos: a) treinado com 35 neurônios e b) treinado com 45 neurônios.

(a)

(b)

Figura 7.9 - Curvas ROC do classificador considerando descritores de forma mais idade e mais densificação como atributos: a) treinado com 40 neurônios e b) treinado com 50 neurônios.

A partir dos resultados obtidos para o *Grupo de Imagens II*, foi possível verificar a influência da qualidade da imagem na performance do classificador. Embora a resolução espacial deste grupo de imagens seja inferior à do *Grupo de Imagens I*, a técnica de digitalização é bastante superior com a varredura a laser, de

modo que visualmente a resolução de contraste obtida nesta técnica é melhor. Estes resultados nos levam a crer que a resolução de contraste seja um parâmetro de alta relevância, influenciando na fase de processamento da imagem e compensando eventual piora de resolução espacial.

Um aspecto que deve ser discutido é a especificidade igual a 100%. Isso significa que o classificador não retornou nenhum falso-negativo (FN). Entretanto, classificar corretamente todos os casos positivos tem um custo que reflete nas respostas falso-positivas (FP), aumentando este índice em alguns casos. Isso é bastante raro de se conseguir tanto em esquemas CAD quanto no diagnóstico dado pelo próprio médico. Esse resultado bastante satisfatório se deve à base de dados representar bem o problema, pois, como já citado anteriormente, os "clusters", pelo simples fato de estarem presentes em uma imagem, já representam uma suspeita maior sugerindo melhor investigação.

Os diagnósticos FP e FN são indesejáveis não só pelos custos que eles representam, mas também pela relação custo-benefício do próprio paciente, tanto nos traumas físicos e psicológicos dos diagnósticos FP quanto na falta de tratamento e aumento do índice de mortalidade que pode ocorrer para os diagnósticos FN.

Para os resultados obtidos com o *Grupo de Imagens II*, a performance do classificador com os testes usando apenas os descritores de forma foi bastante superior, mostrando mais uma vez a eficiência dos Invariantes de Momentos de Hu.

7.4 Precisão e Erro obtidos com análise das curvas ROC

A precisão do classificador é dada pela fórmula 7.1 a baixo. Esta medida foi extraída para todos os testes.

$$P = \text{sensibilidade} \times [\text{fração da população que é realmente positiva}] + \text{especificidade} \times [\text{fração da população que é realmente negativa}] \quad (7.1)$$

A figura 7.10 mostra as áreas sob as curvas ROC obtidas em função das respectivas precisões.

Figura 7.10 - Precisão dos resultados obtidos com o *Grupo de Imagens I* representado pelos pontos azuis e dos resultados obtido com o *Grupo de Imagens II* representado pelos pontos vermelhos.

A partir do gráfico apresentado na figura anterior observamos que o classificador apresenta maior precisão nos testes realizados com o *Grupo de Imagens II*, sendo que para o teste com maior área sob a curva ROC ($A_z = 0,96915$) a precisão ficou próxima de 95%. Enquanto com os testes realizados com o *Grupo de Imagens I*, a maior precisão não atingiu 90%.

Na figura 7.11 é possível observar os erros do sistema. A margem de erro do classificador para cada curva ROC foi calculada segundo está descrito no capítulo 3.

Figura 7.11 - Erros em função de A_z , representados pelos pontos azuis os resultados obtidos com o *Grupo de Imagens I* e pelos pontos vermelhos os resultados obtidos com *Grupo de Imagens II*.

O menor erro verificado também foi para o *Grupo de Imagens II* (para os testes usando apenas descritores de forma, cujo erro foi menor que 8×10^{-3} para a melhor precisão). Para o *Grupo de Imagens I*, o menor erro ficou próximo de 0,12.

Nos gráficos de barras nas figuras a seguir é possível observar a influência da variação do número de neurônios nas camadas escondidas no desempenho do classificador, bem como o melhor desempenho para os exemplos do *Grupo de Imagens II*.

Figura 7.12 - Gráfico de desempenho do classificador para os quatro conjuntos de atributos com variações da quantidade de neurônios para os exemplos do *Grupo de Imagens I* e *Grupo de Imagens II*.

A figura 7.12 mostra o desempenho superior do classificador quando consideramos apenas os atributos dos descritores de forma para os testes feitos com os exemplos dos dois grupos de imagens considerados.

As barras em tons de azul representam os resultados para o *Grupo de Imagens I* e as barras em tons de vermelho os resultados para o *Grupo de Imagens II*. A superioridade do classificador é clara para os resultados do *Grupo de Imagens II* para todas as combinações de atributos e para as variações dos números de neurônios na camada escondida da rede também.

A baixa performance para os testes com o conjunto de atributos em que consideramos os descritores de forma, mais a idade e mais a "densificação" foi notada com os dois grupos de imagens.

Os atributos idade e "densificação", quando combinados um a um com os descritores de forma, produziram bons resultados, principalmente com o *Grupo de Imagens II*, embora não tenham sido melhores que os testes apenas com os descritores de forma.

Com o *Grupo de Imagens I*, a baixa performance nos testes utilizando "densificação" como atributo pode ser justificada pela característica da base de dados

de poucos casos de mama densa. Já o *Grupo de Imagens II*, a base de dados é mais homogênea.

Com o gráfico apresentado na figura 7.12, fica evidente mais uma vez a influência do número de neurônios da camada escondida da RNA.

8

Conclusões

8.1 Atributos Extraídos dos Agrupamentos

Geralmente os trabalhos de classificação de achados mamográficos que encontramos na literatura enfatizam a classificação de microcalcificações e não de "clusters", ou então classificam o "cluster" a partir de atributos extraídos de cada microcalcificação contida nele.

A partir da teoria dos momentos, vimos que a descrição dos agrupamentos de microcalcificações de maneira global foi bastante eficiente, pois possibilitou extrair características que descrevem exatamente o que é analisado na prática médica, como distribuição, distância, área, forma e ainda irregularidade.

A descrição dos agrupamentos com momentos e os outros descritores de forma serviram para mostrar que, mesmo não extraindo características diretamente de cada microcalcificação pertencente ao "cluster", conseguimos representar bem os casos da base de dados e mostraram a relevância de uma análise detalhada destes achados antes do encaminhamento para o exame histológico. Além disso, comprovamos que formas dos "clusters" estão relacionadas com a questão da malignidade, como sugerido por [EGAN-80].

A idade, bem como a "densificação" são atributos colhidos do histórico e laudo da paciente e podem ou não ser relevantes, dependendo do conjunto de exemplos selecionados para a classificação.

8.2 Classificação por RNA

As RNAs mostram eficiência quando aplicadas na classificação almejada nesse trabalho. O algoritmo "Backpropagation", além de eficiente, é relativamente simples de implementar. Para uma base de dados mais complexa, outros métodos devem ser investigados.

Com os resultados colhidos, ficou clara a importância da definição da base de dados para treinamento e testes do classificador. Isso pôde ser visto quando realizamos experimentos usando diferentes conjuntos de atributos como entrada para a rede, tanto no treinamento quanto na fase de testes, além de grupos de imagens com diferentes características de digitalização.

Os agrupamentos de microcalcificações são achados mamográficos que requerem atenção especial por parte dos especialistas. O reconhecimento e classificação destes achados não são tarefas simples, uma vez considerados os problemas de contraste das imagens mamográficas e as pequenas dimensões destes achados. Isso leva aos erros comumente ocorridos no diagnóstico do câncer de mama. Mais de 60% dos casos que os especialistas consideram suspeitos e são enviados para biópsia são casos negativos. Por outro lado, diagnósticos falsos-negativos também ocorrem, deixando a paciente sem os cuidados necessários.

Os esquemas CAD têm como objetivo oferecer ao especialista uma segunda opinião ajudando o médico no momento de decidir-se quanto à suspeitabilidade do caso e as providências a serem tomadas. Com o esquema desenvolvido nesta pesquisa, optamos por uma nomenclatura mais sutil para as classes do nosso classificador - predição "suspeita" e "não-suspeita" - pois o diagnóstico maligno e benigno é uma tarefa complexa e exige mais do que análise visual de mamogramas.

Os resultados do esquema classificador mostraram taxas baixíssimas de predições FN para quase todos os testes. Além disso, para os testes com o *Grupo de imagens II* as predições FN foram eliminadas e os índices obtidos com análise das curvas ROC mostraram altas performances.

8.3 Conclusões Finais

Através das características patológicas e mamográficas discutidas no capítulo 2 foi possível extrair atributos dos mamogramas digitalizados que representassem as caracterizações dos achados "suspeitos" e "não-suspeitos" através de descritores matemáticos. Além disso, informações de laudo e histórico da paciente foram consideradas de acordo com o que vimos na literatura.

Os procedimentos de reconhecimento de padrões discutidos no capítulo 3 foram implementados e a relevância de cada etapa pôde ser avaliada. As Redes neurais artificiais serviram como técnica de classificação e mostraram eficiência na generalização dos padrões em cada classe. Isso se torna evidente com os percentuais mostrados no capítulo anterior, já que a maioria dos resultados obtidos para o *Grupo de Imagens II* mostrou entre 80% e 92% de acerto total.

A classificação usando características globais de "clusters" permitiu uma otimização na fase de extração de atributos, bem como uma representação diferente do convencional dos agrupamentos "suspeitos" e "não-suspeitos", que, na literatura, são avaliados principalmente a partir de características das microcalcificações individuais.

8.4 Sugestões para Trabalhos Futuros

- Fazer uma análise mais detalhada de achados mamográficos, extraindo mais atributos além dos já testados neste trabalho, a partir de um banco de imagens que ofereça informações mais detalhadas de cada caso;
- Fazer uma análise do parâmetro de "densificação".
- Usar técnica de seleção de atributos, como PCA (Principal Components Analysis);
- Aumentar o número de classes do classificador seguindo a metodologia do BI-RADS;
- Testar outras técnicas de classificação;
- Desenvolver um esquema classificador híbrido, que combine as melhores técnicas de classificação para obtenção de respostas mais precisas e detalhadas.

9

Bibliografia

[ANDERSEN-85]

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-95]

ANDERSON, J. A. *An Introduction to Neural Network*. The MIT Press, Massachusetts Institute of Technology, Massachusetts, Cambridge 1995

[BALLARD-82]

BALLARD, D. H.; BROWN, C. M. *Computer Vision*. Prentice-Hall Inc, Englewood Cliffs, New Jersey, 1982

[BEABLE-90]

BEABLE, R.; JACKSON, T. *Neural Computing: An Introduction*. Adam Hilger, Bristol, 1990

[BRASIL-99]

BRASIL, L.M. “Proposta de arquitetura para sistema especialista híbrido e a correspondente metodologia de aquisição do conhecimento”. Tese de Doutorado, Universidade Federal de Santa Catarina, UFSC, Brasil, 1999

[CAMERON-78]

CAMERON, J.R.; SKOFRONICK, J.G.: *Medical Physics*. New York: Wiley -1978.

[CANELLA-99]

CANELLA, E. Microcalcificações. *Jornal da Imagem*, 2º Caderno, n° 245, jan. 1999

[CHAKRABORTY-90]

CHAKRABORTY, P. D.; WINTER, L. H.; Free-Response Methodology: Alternate Analysis and A New Observer-Performance Experiment . *Radiology*, v. 174, p. 873-881, 1990

[CHAN-97]

CHAN, H-P.; SAHINER, B.; PETRICK, N.; HELVIE, M. A.; LAM, K. L.; ALDER, D. D.; GOODSIT, M.; Computerized classification of malignant and benign microcalcifications on mammograms: Texture analysis using artificial Neural network. *Phys. Med. Biol.* V.42 (1997) 549-567.

[CHITRE-94]

CHITRE, Y.; DHAWAN, A. P.; MOSKOWITZ, M.. Artificial Neural Network based Classification of Mammographic Microcalcifications Using Image Structure Feature. *Annual-International-Conference of the IEEE Engineering in Medicine and Biology Society Proceedings*, v. 16, n pt 1, IEEE, Piscataway, NJ, USA, p. 592-593, 1994

[CURRYIII-90]

CURRY III, T.S.; DOWDEY, J.E.; MURRY JR., R.C. *Christensen's Physics of Diagnostic Radiology*. Lea & Febiger, Philadelphia, 4th Edition 1990

[DHAWAN-93]

DHAWAN, A.P.; CHITRE, Y.; MOSKOWITZ, M.; *Artificial neural network based classification of mammographic microcalcifications using image structure features*. Proceedings of SPIE v.1905, p.820-831, 1993

[DOI-95]

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-96]

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, 1994

[DUDA-73]

DUDA, R.O; HART, P.E.H. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1973.

[DUDA-79]

DUDA, R.O; NITZAN, D.; BARRET, P. *Uses of Range and Reflectance Data to Find Surface Regions*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, n° 3, July, pp. 259-271, 1979.

[EGAN-80]

EGAN, R. L.; McSWEENEY, M.B.; SEWELL, C.W. Intramammary calcifications without na associated mass in benign and malignant diseases. *Radiology*, v.137(1), p. 1-7, 1980.

[ELLIOTT - 13]

ELLIOTT, E.B. – *An introduction to the Algebra of Quantics*, Chelsea Publishing Company – Bronx – New York, 1913.

[EVANS-81]

EVANS, A LI. *The Evolution Medical Images*. Medical Physics Handbooks 10, 1981

[FRANÇA-99]

FRANÇA, C. A. *Avaliação da Qualidade de Placas de Madeira através de um Sistema de Inferência Nebuloso Baseado em Redes Adaptativas*. São Carlos. 208p. Tese (Doutorado) - Instituto de Física de São Carlos. Universidade de São Paulo.

[FREEMAN-88]

FREEMAN, M.O & SALEH, B.E.A – *Moments invariants in the space and frequency domains* – Journal of Optical Society of America A, V.5, p. 1073 -1988

[FREDERICK-98]

FREDERICK, E.D.; FLOYD, C.E. Analysis of mammographic findings and patient history data with genetic algorithms for the prediction of breast cancer biopsy outcome. *Proceedings of SPIE* v.3338, p.241-245. 1998

[FU-97]

FU, LIMIN; *Neural Networks in Computer Intelligence*. McGRAW-Hill Series in Computer Science, University of Florida, USA, 1997

[FUKUNAGA-90]

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*, Second Edition , 1990.

[GIULIANO- 61]

GIULIANO, V.E., JONES, P.E., KIMBALLI, G.E., MEYER, R.F., STEIN, B.A.: *Automatic Pattern Recognition by a Gestalt Method. Information and Control*, Vol.4, (1961) 332-345.

[GONZALES-87]

GONZALEZ, R.C.; WINTZ, P. *Digital Image Processing*. Addison-Wesley Publishing Company, United States of America, 1987

[GOLDSTEIN-98]

GOLDSTEIN, E. B. – “Sensation and Perception” – Wadsworth Publishing company – 1989.

[GRAM-97]

GRAM, I.T.; FUNKHOUSER, E.; TABÁR, L. The Tabár classification of mammographic parenchymal patterns. *European Journal of Radiology* v.24, p. 131-136. 1997

[GUREVICH - 64]

GUREVICH, G.B – *Foundations of the Theory of Algebraic Invariants* – Nordoff, Groningen, The Netherlands – 1964.

[HALL - 79]

HALL, E. L. - *Computer Image Processing and Recognition* – Academic Press – NY - 1979

[HARALICK-79]

HARALICK, R.M. Statical and Approaches to Texture. *Proceedings of the IEEE*, v.67, nº 5, May, p.786-804, 1979

[HERTZ-91]

HERTZ, J.; KROGH, A.; PALMER, R. G.; *Introduction to the Theory of Neural Computation*. Addison_Wesley Publishing Company, June, 1991

[HU-61]

HU, M. K. - Pattern Recognition by Moments Invariants – Proc. of the IRE, V. 49, N.9 - p. 1428, 1961

[HU - 62]

HU, M. K. - Visual Pattern Recognition by Moments Invariants - IRE Trans. Inf. Theory, V. 8 - pp. 179-187 , 1962.

[JAIN - 95]

JAIN, R. – *Machine Vision* – MIT Press & McGraw-Hill Inc. – 1995.

[JIANG-96]

JIANG, Y.; NISHIKAWA, R.M.; WOLVERTON, D.E.; METZ, C.E.; GIGER, M.L.; SCHMIDT, R.A.; VYBORNÝ, C.J.; DOI, K. Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology*, 198, p. 671-679, 1996

[JIANG-98]

JIANG, Y.; NISHIKAWA, R.M.; PAPAIOANNOU, J; Requirement of microcalcification for computerized classification of malignant and benign clustered microcalcifications. *Proceedings of SPIE*, v.3338, p.313-317, 1998.

[KHOTANZAD - 90]

KHOTANZAD, A.; HONG, Y.H. Invariant Image Recognition by Zernike Moments – *IEEE Trans. On Pattern Analysis and machine intelligence*, V.12, N.5, May 1990.

[KOVACS-96]

KOVACS, Z. L.; *Redes Neurais Artificiais : Fundamentos e Aplicações*. 2. Ed. Acadêmica, São Paulo, 1996

[LANYI-86]

LANYI, M.; *Diagnosis and Differential Diagnosis of Breast Calcifications*. Röntgeninstitut Kaiserstrabe, German, 1986

[LE GAL-84]

LE GAL, M.; CHAVANNE, G.; PELLIER, D. Valeur diagnostique des microcalcifications groupées découvertes par mammographies. *BULL Cancer* v.71, p.57-64 1984

[LIN - 94]

LIN, W-G & WANG, S-S – A note on the calculation of moments – *Pattern Recognition Letters*, V.15, pp. 1065-1070, 1994.

[LoJY-97]

LO, JY.; BAKER, J.A; KORGUTH, P.J.; INGLEHARD, J.D.; FLOYD, C. Predicting Breast Cancer Invasion with Artificial Neural Networks on the Basis of Mammographic Feature. *Radiology* No. 203, pp. 159-163, 1997

[LoJY-98]

LO, JY.; BAKER, J.A; KORGUTH, P.J.; FLOYD, C. Artificial Neural Networks: Improving the Quality of Breast Cancer Biopsy Recommendations.
Radiology No. 216, pp. 131-135, 1998

[LoJY-99]

LO, JY.; BAKER, J.A; KORGUTH, P.J., FLOYD, C. E. Effect of Patient History Data on the Prediction of Cancer from Mammography Findings with Artificial Neural Networks.
Academic Radiology vol. 6 No. 1, pp. 10-15, 1999

[MARAR-97]

MARAR, J. F., Polinômios Potência de Sigmóide (PPS): Uma Nova Técnica para Aproximação de Funções, Construções de Wavenets e suas Aplicações em Processamento de Imagens e Sinais, 180p. Tese (Doutorado) – Unoversidade Federal de Pernambuco. CCEN. Informática, Recife 1997.

[MARR]

MARR, D. – “Vision” – Freeman, São Francisco, (1982).

[MARTÍ-98]

MARTÍ, J.; GUFÍ, X.; REGINCÓS, J.; ESPAÑOL, J.; PONT, J.; BARCELÓ, C.
Shape-based Features Selections for Microcalcification Evaluation *Proceedings of SPIE*, v.3338, p.1215-1224, 1998.

[MASCARENHAS-89]

MASCARENHAS, N. D.A.; VELASCO, F.R. D. *Processamento Digital de Imagens*
Editorial Kapelusz S.A., Buenos Aires, Argentina, 1989.

[MASTERS-94]

MASTERS, T. *Signal and Image Processing with Neural Networks a C++ Sourcebook*. JohnWiley & Sons, Inc., New York, 1994

[Mc-CULLOCH-43]

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-86]

METZ, C.E ROC Methodology in Radiologic Imaging. *Investigate Radiology*, v.21
n.9, p. 720-33. Sep. 1986

[MINSKY-69]

MINSKY, M. L. and PAPERT, S. A *Perceptrons*. MIT Press, Cambrige, Massachets, 1969.

[MITCHELL-88]

MITCHELL JR., G. W.; BASSET, L. W. *Mastologia Prática*. Revinter, 1988.

[NAGEL-98]

NAGEL, R. H.; NISHIKAWA, R.M.; PAPAIOANNOU, J.; DOI, K. Analysis of methods for reducing false positive in automated detection of clustered microcalcifications in mammograms. *Medical Physics*, v.25, n.8, p.1502-1506, 1995

[NÄPPI-99]

NÄPPI, J.; DEAN, P.B.; NEVALAINEN, O.; TOIKKANEN, S. Computerized Diagnosis of Breast Calcification using Specimen Radiography and Simulated Calcifications. *Proceedings of SPIE*, v.3661, p.1351-1360, 1999.

[NEVES-95]

NEVES, E. M. A. - Reconhecimento Óptico de Caracteres Isolados baseados em suas Características Fundamentais - Dissertação de Mestrado - Escola de Engenharia de São Carlos - 1995.

[NISHIKAWA-93]

NISHIKAWA, R.M.; GIGER, M.L.; DOI, K.; VYBORNÝ, C.J.; SCHMIDT, R.A. Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. *State of the Art in Digital Mammographic Image Analysis* World Scientific Publishing Co.1993

[NISHIKAWA-95]

NISHIKAWA, R.M.; GIGER, M.L.; DOI, K.; VYBORNÝ, C.J.; SCHMIDT, R. A. Computer-Aided Detection of the Clustered Microcalcifications on Digital Mammograms. *Medical Biol. Eng. Comput.*, v. 33, p. 174-178, Mar. 1995

[NUNES-97]

NUNES, F. L. S. *Esquema de diagnóstico auxiliado por computador para detecção de agrupamentos de microcalcificações por processamento de imagens mamográficas*. São Carlos. 208p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos - Universidade de São Paulo.

[PATROCINIO-2000]

PATROCINIO, A.C., SCHIABEL, H.; BENATTI, R.H.; GOES, C.E.; NUNES, F.L.S., Investigation of clustered microcalcification features for an automated classifier as part of a mammography CAD Scheme. *Cd Rom Proc. of The World Congress On Medical Physics And Biomedical Engineering*, Chicago(USA), 4 pp., July 23-28, 2000.r

[PATTERSON-96]

PATTERSON, D.W. *Artificial Neural Networks - Theory and Applications*. Prentice Hall, Singapore, 1996

[PARK- 86]

PARK, K.T.; HALL, E.L.; Form recognition using moment invariants for three dimensional perspective transformation – Science on Form: Proc. of the First International Symposium for Science on Form, S. Ishikawa (Ed.), pp. 255-264, 1986, KTK Scientific Publishers, Tokyo.

[PRATT-91]

Pratt, W. K. *Digital Image Processing*. A Wiley-Interscience Publication, 1991

[PROKOP - 92]

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.

[RICH-88]

RICH, E. *Inteligência Artificial*. McGRAW-HILL, 1988

[ROBBINS-51]

ROBBINS, H.; MONRO, S. A stochastic approximation method. *Annals. Of Mathematical Statistic*, p. 25: 737-744, 1951

[ROGOVA-99]

ROGOVA, G. L.; STOMPER, P. C.; KE, C. C. Microcalcification Texture Analysis in a Hybrid System for Computer Aided Mammography *Proceedings of SPIE* v.3661, p.1426-1433 1999

[ROSEMBLATT-58]

ROSEMBLATT, F. The Perceptron: A probabilistic model for information storage and organization in brain. *Psychological Review*, pages 65:386-408, 1958

[ROSEMBLATT-60]

ROSEMBLATT, F. Perceptron simulation experiments. *Proceedings of the Inst. of Radio Eng.*, p. 48: 301-309, 1960

[RUMELHART-86]

RUMELHART, D. E.; WILLIAMS, R.J. Learning internal representations by error propagation. *Parallel distributed processing*, v 1, p.318-62, 1986

[SCHALKOFF-89]

SCHALKOFF, R. J.; *Digital Image Processing*. John Wiley & Sons, Inc., New York, 1989

[SHEN-93]

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

[SHEN-94]

SHEN, L.; RANGAYYAN, R.M.; DESAUTELS, J.E. L. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging* v.13(2), p.263-274. 1994

[SICKLES-82]

SICKLES, E.A. Mammographic detectability of breast microcalcifications. *Am. J. Radiol.* v.139, p.913-918 1982

[SICKLES-86]

SICKLES, E. A. Breast calcifications: mammographic evaluation. *Radiology* v.160, p.289-293 1986

[STRACHAN - 90]

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-98]

TAO, E.; ORNES, C.; SKLANSKY, J. Automatic detection of microcalcifications in digital mammography. *Proceedings of SPIE* v.3338, p.1450-1458.

[TEAGUE - 80]

TEAGUE, M.R. Image Analysis via the general theory of moments - *J. Opt. Soc. Am.*, - V.70, N. 8, pp. 920-930., 1980.

[VELDKAMP-99]

VELDKAMP, W. J. H.; KARSSEMEIJER, N. An Improved Method for Detection of Microcalcification Clusters in Digital Mammograms. *Proceedings of SPIE* v.3661, p.512-522 1999

[VILLELA-96]

VILLELA, R.L. Estudo comparativo das técnicas de segmentação de microcalcificações em imagens mamográficas. *Anais do III Fórum Nacional de Ciência e Tecnologia em Saúde*, v.2, p. 575-576, 1996

[VYBORNÝ-94]

VYBORNÝ, C.J.; GIGER, M.L. Computer vision and artificial intelligence mammography. *AJR* v. 162, p. 699-708 1994.

[WIDROW-60]

WIDROW, B.; Hoff, M. E.. Adaptive switching circuits. *IRE WESCON Conversation Record*, p. 96-104, 1960.

[WINTER-88]

WINTER, B. S.; DENKER, J. S. Strategies for teaching Layered networks Classifications Tasks, *Neural Informations Processing Systems*. American Institute of Physics- New York, 1988.

[WOLBASRST-93]

Wolbasrst, Anthony Brinton; *Physics of Radiology*, Appleton & Lange, 1993.

[WOOD - 96]

Wood, J. – Invariant Pattern Recognition: A review – *Pattern Recognition*, V.29, N.1, pp. 1- 17, 1996.

[WU-93a]

WU, Y.; DOI, K.; METZ, C. E.; ASADA, N.; GIGER, M. L. Simulations Studies of Data Classification by Artificial Neural Networks: Potential Applications in Medical Images and Decision Making. *Med. Physics* vol.6 No. 2 , 1993.

[WU-93b]

WU, Y.; DOI, K.; GIGER, M. L.; NISHIKAWA, R. M. Computerized Detection of Clustered Microcalcifications in Digital Mammograms: Applications of Artificial Neural Networks. *Med. Physics* vol.19 No. 3 , 1993.

[YOUNG-86]

YOUNG, T.Y. & FU, K. S.; *Handbook of Pattern Recognition and Image Processing*. Academic Press, USA, 1986.

[YU-98]

YU, S.; GUAN, L.; BROWN, S. Automatic detection of clustered microcalcifications in digital mammograms based on wavelet features and neural network classification. *Proceedings of SPIE* v.3338, p.1540-1546 1998

[ZHANG-94]

ZHANG, W.; DOI, K.; GINGER, M. L.; WU, Y.; NISHIKAWA R.M.; SCHMIDT, R. A. Computerized detection of clustered microcalcification in digital mammograms using a shift-invariant artificial neural network. *Medical Physics*, 21(4) April 1994