

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

LÚCIO FLÁVIO DE ALBUQUERQUE CAMPOS

Classificação de Lesões em Mamografias Digitais Utilizando
Análise de Componentes Independentes e Perceptron
Multicamadas

SÃO LUÍS - MA

2006

LÚCIO FLÁVIO DE ALBUQUERQUE CAMPOS

Classificação de Lesões em Mamografias Digitais Utilizando
Análise de Componentes Independentes e Perceptron
Multicamadas

Dissertação apresentada ao Programa de Pós
Graduação em Engenharia de Eletricidade da
UFMA, como requisito para a obtenção do grau
de MESTRE em Engenharia de Eletricidade.

Orientador: Allan Kardec Duailibe Barros Filho

Universidade Federal do Maranhão

SÃO LUÍS - MA

2006

Campos, Lúcio Flávio de Albuquerque

Classificação de Lesões em Mamografias Digitais Utilizando
Análise de Componentes Independentes e Perceptron Multicamadas
/ Lúcio Flávio de Albuquerque Campos. - São Luís, 2006.

Dissertação (Mestrado) - Programa de Pós-Graduação em
Engenharia de Eletricidade, Universidade Federal do Maranhão.

1.Processamento de Imagens,2.Mamas-lesões-identificação
automática, 3.Mamografia I. Título.

CDU 004.932

LÚCIO FLÁVIO DE ALBUQUERQUE CAMPOS

Classificação de Lesões em Mamografias Digitais Utilizando
Análise de Componentes Independentes e Perceptron
Multicamadas

Dissertação apresentada ao Programa de Pós
Graduação em Engenharia de Eletricidade da
UFMA, como requisito para a obtenção parcial do
grau de MESTRE em Engenharia de Eletricidade.

Apresentado em 24 de março de 2006

BANCA EXAMINADORA

Allan Kardec Duailibe Barros Filho

Universidade Federal do Maranhão

Aristófanês Corrêa Silva

Universidade Federal do Maranhão

Alexandre César Muniz de Oliveira

Universidade Federal do Maranhão

Dráulio Barros de Araújo

Universidade de São Paulo

Dedico este trabalho à minha família

Resumo

Neste trabalho, propomos um método para discriminação e classificação de mamogramas, com diagnóstico maligno, benigno e normal, usando análise de componentes independentes e redes neurais. O método foi testado com mamogramas da MIAS database, e com redes perceptron multicamadas. O método obteve uma taxa de sucesso média de 97.83%, com 97.5% de especificidade, e 98% de sensibilidade.

Palavras-chaves:mamograma, câncer de mama, análise de componentes independentes, diagnóstico auxiliado por computador

Abstract

We propose a method for discrimination and classification of mammograms with benign, malignant and normal tissues using independent component analysis and neural networks. The method was tested for a mammogram set from MIAS database, and multilayer perceptron. The method obtained a success rate of 97.83% , with 97.5% of specificity and 98% of sensitivity.

Keywords: Mammogram, breast cancer, independent component analysis, neural networks, computer aided diagnosis.

Agradecimentos

À DEUS, por ter me dado perseverança e paciência

À toda a minha família, que sempre esteve presente, e teve infinita paciência comigo nestes últimos dois anos.

Ao meu tesouro, Jayne R. Campos, que não convive comigo, mais mora no meu coração.

Ao meu orientador, o Prof. Allan Kardec Barros, por ter depositado confiança no meu trabalho e na minha capacidade.

Ao Prof. Aristófanês Corrêa, por ter me apresentado o mundo do Processamento de Imagens, por ter me co-orientado nos meus artigos, e por ter tido uma enorme paciência na revisão deste trabalho.

A todos os meus amigos do laboratório de Processamento da Informação Biológica PIB, Carlos Magno, Denner Guilhon, Deusdete Brito, Diego, Ewaldo Éder, Glenda Raposo, Ivan Jr., Ranielma Machado, André Borges.

Aos meus amigos Jaderson, Mauro e Rycardo, do Laboratório de Instrumentação Eletrônica LIEA

À Cleidiane Gomes, que me acompanhou por mais de 4 anos da minha vida, e esteve presente, me dando apoio em grande parte deste mestrado.

Aos meus amigos Ricardo Robson, Raniere Machado e Alex A. Paz, pelos momentos de descontração e brincadeira nas sexta-feiras.

Ao meu amigo Alcides Neto, da secretaria da pós-graduação.

Ao meu amigo André Borges, por ter paciência em me ensinar conceitos básicos que não entravam na minha cabeça, e por ter quebrado todos os fones que eu comprava.

À minha amigona do peito Jaciani Pereira, por ter emprestado os ouvidos nas horas que mais precisava.

À minha vida, Ivana Souza, que re-apareceu do nada, e hoje se tornou simplesmente tudo.

*'Na realidade, as maiores bênçãos nos
chegam por intermédio da loucura,
quando é mandada como um presente
dos deuses.'*

Platão

Sumário

Lista de Figuras	9
Lista de Tabelas	11
1 Introdução	12
1.1 A Mamografia e o Câncer de Mama	14
1.2 Trabalhos Relacionados	20
1.3 Organização do Trabalho	22
2 Diagnóstico Auxiliado por Computador	23
2.1 Processamento Digital de Imagens	24
2.2 Análise de Textura	25
2.3 Análise de Componentes Independentes	25
2.3.1 Definições	27
2.3.2 Definição de Independência	28
2.3.3 Técnicas de Estimacão das Componentes Independentes	29
2.4 Seleção de Características mais Significantes	32
2.5 Redes Neurais como Classificadores	33
2.5.1 Redes Neurais Perceptron Multicamadas	35

3	O Método Proposto	39
3.1	ICA aplicado em Processamento de Imagens	39
4	Material	43
4.1	Base de dados MIAS	43
4.2	Softwares Utilizados	45
5	Resultados e Discussão	46
5.1	Extração de Parâmetros Utilizando ICA	46
5.2	Variáveis Seleccionadas	49
5.3	Configuração da Rede Neural MLP	50
5.4	Discussão	50
6	Conclusão e trabalhos futuros	53
	Referências Bibliográficas	55

Lista de Figuras

1.1	Incidência médio-lateral das mamas [SANTOS, 2002].	15
1.2	Incidência crânio-caudal das mamas. [SANTOS, 2002]	15
1.3	Mamografia com lesão circunscrita. Banco de dados MIAS, identificação Mdb005 fonte: MIAS [SUCKLING, 1994].	17
1.4	Mamografia com uma lesão nodular de bordas espiculadas. Banco de dados MIAS, identificação Mdb184 fonte: MIAS [SUCKLING, 1994].	17
1.5	Nódulo de alta densidade, lobulado, de contornos parcialmente definidos. fonte: www.fleury.com.br , acessado em 22/02/2006	18
1.6	Mamografia com uma lesão nodular de características benignas. Banco de dados MIAS, identificação Mdb021 fonte: MIAS [SUCKLING, 1994].	18
1.7	Mamografia com presença de microcalcificações. Banco de dados MIAS, identificação Mdb075 fonte: MIAS[SUCKLING, 1994].	19
2.1	Arquitetura de RNA com duas saídas.	34
2.2	Ilustração das direções de dois fluxos básicos de um sinal em uma rede MLP: Propagação direta dos sinais e a retro-propagação dos sinais de erro.	38
3.1	Diagrama do método proposto.	39
3.2	Modelo de separação linear de imagens.	41

4.1	Seleção das ROI's de tecidos benignos, malignos e normais.	44
5.1	Região de interesse com uma combinação linear de imagens bases mútua e estatisticamente independentes entre si.	46
5.2	Amostra de imagens bases obtidas de ROI's de tecidos normais, benignos e malignos.	49
5.3	Configuração de rede neural MLP, com 45 neurônios de entrada, 21 na camada intermediária, e 3 neurônios na camada de saída.	50

Lista de Tabelas

4.1	Exemplo do diagnóstico mostrado em algumas mamografias da base de dados MIAS	44
5.1	Classificação das ROI de tecidos Benignos, Malignos e Normais	51
5.2	Classificação das ROI's em normais e anormais	51

1 Introdução

O câncer de mama permanece como o segundo tipo de câncer mais freqüente no mundo e o primeiro entre as mulheres. De acordo com a organização mundial de saúde, mais de 1.2 milhões de pessoas foram diagnosticadas com câncer de mama em 2005. No Brasil a realidade é similar. Dos cerca de 467 mil novos casos de câncer esperados para 2006, 48.930 são de câncer de mama, representando um risco estimado de 52 casos a cada 100 mil mulheres, segundo dados fornecidos pelo Instituto Nacional do Câncer, situado no Rio de Janeiro. No Estado do Maranhão, dos 1740 casos estimados para 2006, 300 são de câncer de mama. Nas últimas duas décadas, observou-se um aumento considerável da taxa de mortalidade por câncer de mama entre mulheres. Esse número passou de 5,77/100.000 em 1979 a 9,74/100.000 em 2000, correspondendo a uma variação percentual relativa de +80,3% [INCa,2005].

O câncer de mama, assim como os demais, é resultado de alterações do DNA, que levam a uma proliferação celular desordenada. Quando há uma falha do mecanismo regulador que mantém o equilíbrio entre o crescimento celular e o bem estar do organismo, ocorre a formação de massas, denominadas tumores ou neoplasias, as quais são classificadas como malignas ou benignas. As neoplasias benignas têm crescimento organizado, em geral lento, e o tumor apresenta contorno bem nítido. Na neoplasia maligna, o crescimento é rápido, desordenado e infiltrativo. Nos tumores malignos, suas células têm capacidade de se desenvolver em outras partes do corpo, fenômeno este denominado metástase [BAUER,1980].

O câncer de mama se comporta menos previsivelmente do que outros cânceres, e embora algumas pacientes com doença não tratada possam sobreviver por longos períodos, outras com câncer mínimo de mama podem chegar a óbito muito rapidamente. Os aspectos importantes na predição da gravidade do câncer de mama incluem a magnitude da resposta do hospedeiro e as características do tumor. A sobrevida das pacientes é diretamente relacionada com o tamanho do tumor no diagnóstico inicial. Portanto, o diagnóstico precoce não apenas influencia o prognóstico, mas propicia cirurgia menos mutilante e com sobrevida comparáveis a intervenções cirúrgicas mais dramáticas e agressivas[BAUER,1980].

As causas para o aparecimento do câncer na mama ainda não são totalmente conhecidas. Sabe-se que a combinação de mutações genéticas, estilo de vida e influência ambiental podem levar, sem uma razão específica, a mudanças no funcionamento dos genes das células mamárias. Pode-se passar a vida inteira com uma mutação genética sem, necessariamente, desenvolver um tumor. O problema é quando ela ocorre em duas categorias específicas de genes: nos oncogenes, que são os causadores de câncer, ou nos genes supressores de tumor, responsáveis por prevenir qualquer alteração no DNA e conter o crescimento celular descontrolado.

Quanto mais precocemente for detectada a lesão, maiores as chances de cura da paciente. A forma mais eficaz para detecção precoce do câncer de mama é a mamografia, ou radiografia da mama, que é capaz de mostrar lesões em fase inicial, muito pequenas (de milímetros). É realizada em um aparelho de raio X apropriado, chamado mamógrafo. Nele, a mama é comprimida de forma a fornecer melhores imagens, e, portanto, melhor capacidade de diagnóstico.

A mamografia é um exame simples, barato, e de fácil acessibilidade, porém

a grande desvantagem da mamografia é seu erro de diagnóstico, que é considerado muito elevado, e está compreendido entre 12 a 54% [INCa,2005], erro esse que contribui consideravelmente para o aumento de número de diagnósticos falso-positivos, que podem levar a paciente a uma biopsia desnecessária, e o aumento de diagnósticos falso-negativos, que podem atrasar a descoberta de um câncer, que a maioria das vezes quando é descoberto, já é tardio.

As causas mais importantes de mamogramas de diagnóstico falso são os seguintes: o tamanho e a localização da lesão, a densidade do tecido mamário, a qualidade dos recursos técnicos e a habilidade de interpretação do radiologista [INCa,2005].

1.1 A Mamografia e o Câncer de Mama

A mamografia constitui uma forma particular de radiografia, que trabalha com níveis de tensões e correntes em intervalos específicos, destinada a registrar imagens da mama a fim de diagnosticar a presença ou ausência de estruturas que possam indicar patologias. Segundo o INCA [INCa,2005], em uma mamografia, duas incidências de cada mama são indispensáveis: uma visão lateral ou oblíqua, como mostra a Figura 1.1 e uma crânio-caudal, conforme a Figura 1.2.

No entanto, a incidência médio-lateral-oblíqua (MLO) é a mais eficaz, pois mostra uma quantidade maior de tecido mamário e inclui estruturas mais profundas do quadrante súpero-externo e do prolongamento axilar.

A incidência crânio-caudal (CC) tem como objetivo incluir todo o material pósteromedial, complementando a médio-lateral-oblíqua, que com frequência não está totalmente demonstrado na incidência MLO. Permite também mais compressão da mama,

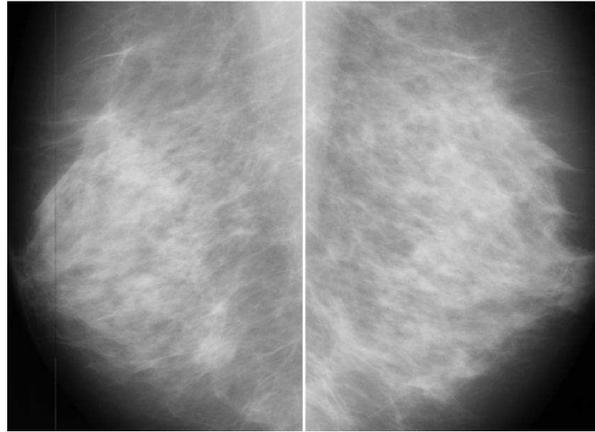


Figura 1.1: Incidência médio-lateral das mamas [SANTOS, 2002].

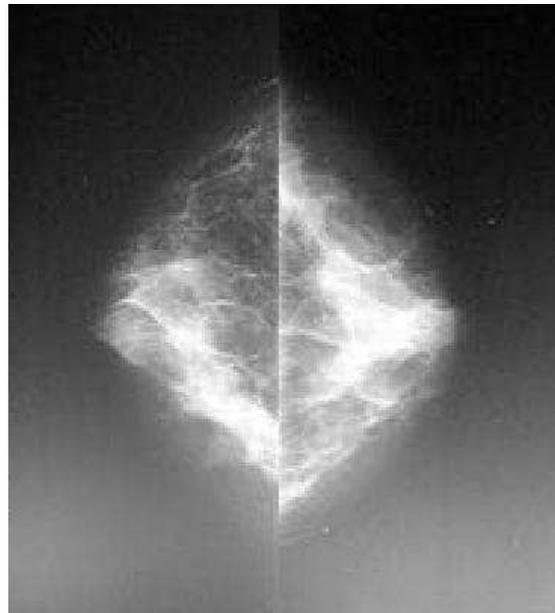


Figura 1.2: Incidência crânio-caudal das mamas. [SANTOS, 2002]

uma vez que não inclui a axila, resultando em uma definição superior da arquitetura mamária e de lesões. Os radiologistas estudam as incidências crânio-caudais e as médio-laterais aos pares de modo a permitir a comparação de regiões simétricas, pois qualquer assimetria pode ser indício de patologia.

A mamografia é um exame de alta sensibilidade, apesar de a maioria dos estudos evidenciarem perdas entre 10 a 15% [INCa,2005] dos casos de câncer com tumor detectável ao exame clínico. Esta sensibilidade, no entanto, está diretamente relacionada

à idade da mulher, sendo muito menor nas mulheres jovens, que apresentam uma alta densidade de tecido mamário, devido à predominância de tecidos fibroglandulares na sua composição.

O reconhecimento de estruturas que possam indicar a presença de câncer ocorre através da constatação de uma diferença de contraste entre os diversos tecidos envolvidos. A gordura, por exemplo, absorve uma menor quantidade de raios-X, aparecendo mais escura no mamograma, enquanto tecidos fibroglandulares apresentam densidade óptica maior e aparecem mais claros [BOYD,1995]. Geralmente microcalcificações e massas aparecem em tonalidades mais claras na imagem obtida após a revelação do filme mamográfico, mas essa diferenciação fica prejudicada em imagens de mamas densas. Por esse motivo, muitas vezes a descoberta do câncer de mama em mulheres com menos de 40 anos de idade acontece quando o tumor já apresenta um desenvolvimento avançado, o que dificulta o tratamento da doença. O diagnóstico de carcinomas não-palpáveis só é possível através da realização de mamografias minuciosas, em que cada detalhe é de extrema importância para evitar os diagnósticos falso-positivos e falso-negativos [BLAND,2000].

Em uma mamografia, o câncer mamário pode aparecer de muitas formas, tal a variedade de padrões com que se pode apresentar. Os padrões de lesão mais típicos são [SANTOS, 2002]:

- **Opacidade circunscrita:** Corresponde ao nódulo, sendo o achado mamográfico encontrado em 39% dos casos de câncer não palpáveis. Os nódulos devem ser analisados de acordo com o tamanho, densidade, contorno, como é observado na Figura 1.3.

- **Contorno espiculado ou estrelado:** corresponde ao chamado carcinoma cirroso. Nódulo de alta densidade com bordas espiculadas, geralmente é diagnosticado

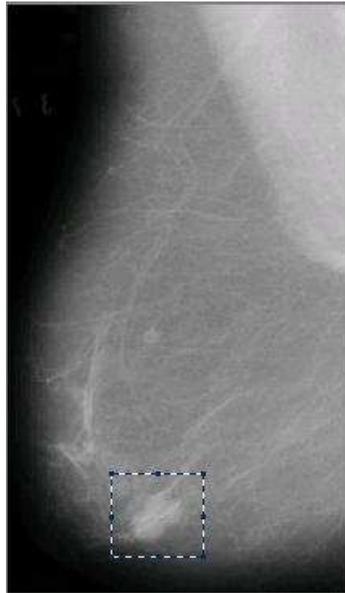


Figura 1.3: Mamografia com lesão circunscrita. Banco de dados MIAS, identificação Mdb005 fonte: MIAS [SUCKLING, 1994].

como carcinoma, como mostrado na Figura 1.4.

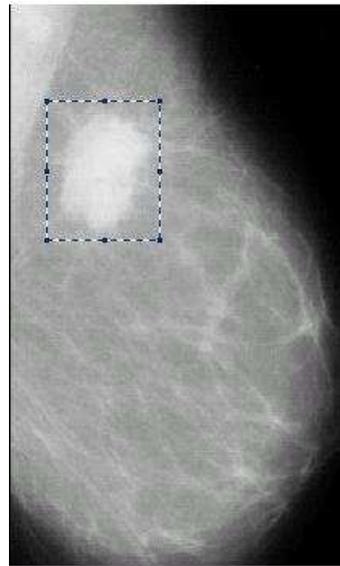


Figura 1.4: Mamografia com uma lesão nodular de bordas espiculadas. Banco de dados MIAS, identificação Mdb184 fonte: MIAS [SUCKLING, 1994].

- **Contorno lobulado:** O aspecto lobulado representa um crescimento tumoral e essa característica autoriza uma suspeição de malignidade, sobretudo se parte do contorno é mal definida ou atenuada, como mostrado na Figura 1.5.

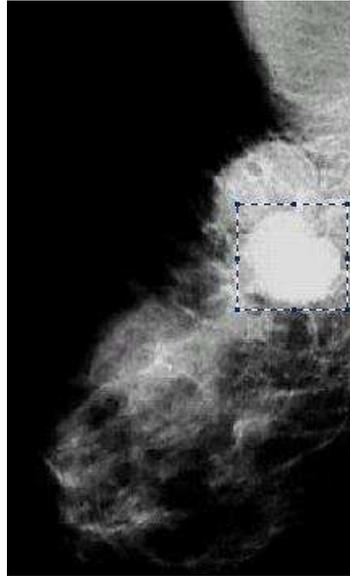


Figura 1.5: Nódulo de alta densidade, lobulado, de contornos parcialmente definidos. fonte: www.fleury.com.br, acessado em 22/02/2006

- **Contorno bem definido:** Característica de benignidade. Contudo, algumas lesões (medular, colóide e mucinoso) podem apresentar contorno bem definido, mas diferem em outras características, tais como textura, por exemplo, como visto na Figura 1.6.



Figura 1.6: Mamografia com uma lesão nodular de características benignas. Banco de dados MIAS, identificação Mdb021 fonte: MIAS [SUCKLING, 1994].

Além das estruturas citadas acima, ainda existem as calcificações, que são depósitos de sais de cálcio nos tecidos mamários que podem ter se desenvolvido devido a processos inflamatórios, alterações degenerativas cíclicas, processos tóxicos metabólicos, traumatismos ou mesmo como resultados de processos secretores ativos de células tumorais [AZEVEDO,1994], conforme a Figura 1.7.



Figura 1.7: Mamografia com presença de microcalcificações. Banco de dados MIAS, identificação Mdb075 fonte: MIAS[SUCKLING, 1994].

O objetivo deste trabalho é propor um método para classificar as regiões de interesse (ROI) de mamografias como normal, benigno e maligno, usando Análise de Componentes Independentes (ICA), para extrair características de textura. Em seguida, estas características são utilizadas como parâmetros de entrada para uma rede neural Perceptron Multicamadas, para efetuar a classificação final.

1.2 Trabalhos Relacionados

Nos últimos anos, vários pesquisadores no mundo têm desenvolvido trabalhos de Diagnóstico Auxiliado por Computador CAD.

Nicholas Petrick [PETRICK E BERKMAN, 2002] desenvolveu um trabalho baseado em *Density-Weighted Contrast Enhancement* (DWCE). Nesse trabalho, o filtro DWCE foi utilizado para acentuar estruturas da mamografia antes da detecção de bordas, por reconhecimento adaptativo local. Após a aplicação do filtro DWCE, a detecção de bordas, utilizando o algoritmo *k-means*, foi realizada para reconhecer as estruturas que indicam o aparecimento de uma lesão. O sistema obteve uma taxa de sucesso de 82.33% na classificação de mamogramas malignos e benignos.

No trabalho de Campanini *et al* [CAMPANINI *et al*, 2002], mamogramas foram classificados em malignos e benignos através de *Support Vector Machines* (SVM). Nesse trabalho, os autores fizeram um reescalonamento bilinear em cada imagem original, em seguida aplicaram a transformada de wavelet a partir das características extraídas de cada mamograma, através das suas wavelets, foi utilizado SVM para a classificação final. O sistema obteve uma sensibilidade de 84%.

Christoyianni *et al* [CHRISTOYIANNI *et al*, 2002] compararam três métodos: *Gray Level Histogram Moments* (GLHM), que utiliza medidas estatísticas, como média, desvio padrão, variância, assimetria, curtose, etc. Essas medidas estatísticas servem como parâmetros de entrada para um classificador baseado em redes neurais. *Spatial Gray Level Dependence Matrix* (SGLD), que é construída a partir da contagem do número pares de *pixels* iguais em uma dada janela. A matriz SGLD é construída através da probabilidade conjunta de níveis de cinza, e a que distância os pares de *pixels* ocorrem,

e Análise de Componentes Independentes (ICA), em que o autor utiliza previamente Análise de Componentes Principais (PCA), para reduzir os número de componentes, e logo após utiliza ICA, para extrair características significativas de cada imagem, para posteriormente classificá-las, com uma rede neural artificial. De acordo com os autores, ICA obteve a melhor performance, com 88% de sucesso, discriminando os mamogramas entre normais e anormais, e 79.31% discriminando entre normal, benigno e maligno.

Leite [LEITE *et al*, 2004] usou (ICA) para classificar mamogramas entre normais e anormais. Nesse trabalho, os autores utilizaram ICA para extrair características de cada imagem, e em seguida, utilizaram PCA para reduzir o vetor de características, que serviu de entrada para uma rede neural artificial. O método conseguiu um percentual de acerto de 82%, clasificando as mamografias em normais e anormais.

Verma e Zhang [VERMA e ZANGH, 2006], desenvolveram um método baseado em um algoritmo neural-génético para encontrar as características mais significantes em mamografias digitais. Cada população individual representou uma solução candidata para solucionar o problema da seleção do subconjunto mais significante. Segundo os autores, com a melhor combinação de características, o método obteve 85% de acerto.

Mousa et al [MOUSA *et al*, 2005] desenvolveram um sistema baseado em wavelets e redes neuro-fuzzy. Nesse trabalho, três técnicas são utilizadas para realçar as imagens: *Image pruning*, equalização de histograma thresholding em níveis de cinza. Em seguida, as características de cada imagem são extraídas através de suas wavelets, e essas características servem de entrada para uma rede neuro-fuzzy. Segundo os autores, o sistema desenvolvido apresentou uma taxa de acerto de 77%, classificando mamogramas entre normais e anormais.

Análise de Componentes Independentes é uma técnica amplamente utilizada em processamento de sinais de audio, eletrocardiograma, eletromagnetoencefalograma, etc. Neste trabalho, ICA é aplicado para a extração de características de textura em mamografias digitais, utilizando o algoritmo *FastICA*. O método obteve um desempenho superior, se comparado aos métodos de CAD desenvolvidos até o momento.

1.3 Organização do Trabalho

Neste trabalho, é apresentado o desenvolvimento de um método para classificar lesões ou regiões suspeitas de lesões em mamogramas. Este trabalho está organizado como segue:

No **Capítulo 2**, são descritos os conceitos de Diagnóstico Auxiliado por Computador, as técnicas que serão utilizadas nesse trabalho, tais como: Processamento de imagens, análise de textura, Análise de Componentes Independentes (ICA), o algoritmo de seleção de características *Forward-Selection*, e uma breve introdução sobre redes neurais Perceptron Multicamadas.

O método proposto, seu diagrama de blocos e a aplicação de análise de Componentes independentes em processamento de imagens será descrito no **Capítulo 3**

No **Capítulo 4** será descrito os materiais utilizados, a base de dados e os softwares utilizados

A aplicação das técnicas descritas e os resultados e discussão serão descritos no **Capítulo 5**, e finalmente no **Capítulo 6** será mostrado a conclusão e a proposta de trabalhos futuros.

2 Diagnóstico Auxiliado por Computador

A comunidade científica vêm reunindo esforços, no sentido de desenvolver sistemas de diagnóstico auxiliado por computador, baseado em técnicas computacionais, que possam ser utilizadas em conjunto com o exame da mamografia, para que esse erro, que é considerado elevado, possa ser diminuído.

Diagnóstico Auxiliado por Computador (CAD) são sistemas que podem ajudar o radiologista, fornecendo uma sugestão de diagnóstico, que pode ser usado nos estágios iniciais do tratamento. Os CADs contornam problemas que surgem da subjetividade de um laudo humano, tais como expectativas, pré-conceitos e cansaço que interferem negativamente no diagnóstico do radiologista. Os CADs buscam, assim, baixar as taxas de falsos positivos, e também baixar a taxa de falsos negativos.

Vários estudos revelam que algoritmos automáticos de detecção são capazes de proporcionar um aumento de mais de 20% no total de acertos do radiologista e, com isso, consegue-se evitar biópsias desnecessárias [ZHANG, 2002].

Para que isto ocorra, é importante desenvolver técnicas para detectar e reconhecer lesões e regiões suspeitas e discriminá-las. Vários métodos de diagnósticos de lesões em mamogramas tem sido desenvolvidos com técnicas variadas e apresentaram resultados satisfatórios. Algumas técnicas tem como objetivo maior encontrar um conjunto de características que possam diferenciar tecidos de classes diferentes, a chamada extração de parâmetros. Posteriormente, esse conjunto de parâmetros que diferenciam tecidos saudáveis de tecidos com lesão, por exemplo, servirão como entrada para um

classificador (uma rede neural, por exemplo), para que se possa fazer a classificação final.

Já outras técnicas têm como objetivo identificar tecidos com características suspeitas de lesão, através de algoritmos de agrupamento. Nos parágrafos seguintes, será descrita sucintamente cada técnica utilizada no nosso trabalho.

2.1 Processamento Digital de Imagens

O processamento digital de imagens pode ser definido como o conjunto de técnicas computacionais que transformam uma imagem digital de entrada em uma saída desejada que na maioria das vezes também é uma imagem digital. Dessa maneira, é possível melhorar o aspecto visual de certas feições estruturais para o observador humano e fornecer outros elementos para a interpretação visual da imagem, podendo inclusive gerar outros produtos que possam ser posteriormente submetidos a outros processamentos.

Ao longo das duas últimas décadas, a área de processamento de imagens digitais apresentou um rápido crescimento. A evolução da tecnologia de computação digital, bem como o desenvolvimento de novos algoritmos para lidar com sinais bidimensionais estão permitindo uma gama de aplicações cada vez maior. Na medicina, por exemplo, o processamento de imagens digitais permite a coloração ou a intensificação de contraste para facilitar a interpretação de radiografias e outras imagens biomédicas [GONZALES e WOODS, 1992].

2.2 Análise de Textura

A característica da textura em imagens é uma fonte importante de informações para o processo de análise e interpretação. Vários métodos de classificação de imagens baseados em análise texturais têm sido amplamente utilizados, possuindo aplicações na medicina, microscopia, sensoriamento remoto, controle de qualidade, etc.

A textura é definida como uniformidade, densidade, aspereza, regularidade, intensidade, entre outras características da imagem. A textura é uma das características mais importantes para classificar e reconhecer objetos e cenas, e pode ser caracterizada por variações locais em valores de *pixels* que se repetem de maneira regular ou aleatória ao longo do objeto ou imagem [GONZALES e WOODS, 1992].

Neste trabalho, a textura é utilizada através de abordagem estatística, ou seja, é definida por um conjunto de medidas locais extraídas de um padrão. Medidas estatísticas comuns incluem contraste, energia, entropia, correlação, homogeneidade, momento, variância, dentre outras. A análise de imagens e extração de suas características através da abordagem estatística propicia uma descrição da mesma, através de regras estatísticas que definem a distribuição e a relação entre seus níveis de cinza. A distribuição dos níveis de cinza dos pixels pode ser descrita por estatísticas de primeira, segunda ou alta ordem.

2.3 Análise de Componentes Independentes

A Análise de Componentes Independentes (Independent Component Analysis-ICA) é um método que é visto como uma extensão da Análise de Componentes Principais

(Principal Component Analysis-PCA). A ICA foi desenvolvida no contexto de separação cega de fontes (*Blind Source Separation*-BSS), em que o problema é definido na estimação da saída de uma fonte conhecida, quando esta fonte recebe vários sinais misturados e desconhecidos. ICA tem sido aplicada em diversas áreas, como por exemplo: áudio, radar, instrumentação médica, comunicação móvel, engenharia biomédica, e outras.

ICA é utilizado em BSS porque consegue recuperar as fontes "não-observáveis" de uma mistura de diversas fontes. O termo *blind* refere-se ao fato de que existem fontes não observáveis no sinal e nada ou pouca informação se tem sobre a mesma. Uma aplicação interessante de BSS é o problema *cocktail party*, em que separam-se as fontes originais de um sinal misturado, sem o conhecimento prévio dos coeficientes de mistura, nem a provável distribuição do sinal, usando apenas independência estatística como critério de separação de fontes [GIROLAMI e FYFE, 1997].

Jutten e Héroult [JUTTEN e HERAULT, 1991] desenvolveram o primeiro algoritmo de aprendizado para BSS. Bell e Sejnowski [BELL e SEJNOWSKI, 1995] desenvolveram uma rede neural capaz de aprender regras que minimizam a informação mútua dos nós de saída. Karhunen e Joutsensalo [KARHUNEN e JOUTSENSALO, 1997] propuseram algumas variações não-lineares de PCA, e demonstraram a utilidade destes algoritmos para estimação de frequência sinusoidais.

Blind Source Separation representa um grande problema na engenharia, pois a técnica mais utilizada anteriormente era PCA, que utiliza apenas estatística de segunda ordem, o suficiente apenas para descorrelacionar um conjunto de dados, mas não necessário para independência, que requer estatística de alta ordem. Por esta razão, a ICA é vista como um método mais "robusto" que PCA, pois se PCA consegue descorrelacionar as fontes não observáveis, ICA consegue deixá-los mútua e estatisticamente independentes

entre sí.

2.3.1 Definições

Considere que sejam observadas n misturas lineares x_1, \dots, x_n de n componentes independentes

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n \quad j = 1, \dots, n \quad (2.1)$$

e que cada mistura x_j , assim como cada componente independente s_k seja uma variável aleatória, e o a_j os coeficientes (pesos) da mistura linear.

Sem perda de generalidade, assume-se que tanto as variáveis da mistura quanto aquelas das componentes independentes têm média zero. Por conveniência, será usada a notação vetorial em vez de somas, como aquelas vistas na Equação (2.1), utilizando letras minúsculas e maiúsculas, para representar, respectivamente, vetores e matrizes. Dessa maneira, podemos reescrever a Equação (2.1) da seguinte forma:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (2.2)$$

O objetivo da técnica é recuperar as fontes \mathbf{s} , através de \mathbf{x} , sem nenhuma informação sobre as propriedades de \mathbf{A} .

O modelo estatístico definido na Equação (2.2) é chamado de modelo de Análise de Componentes Independentes. Esse modelo descreve os dados observados pelo processo de mistura das componentes independentes s_i , que não podem ser observadas diretamente. É preciso estimar tanto \mathbf{s} quanto a matriz de mistura \mathbf{A} , que também é desconhecida, pois tudo o que se observa é o vetor aleatório \mathbf{x} .

O problema do modelo de dados de ICA é estimar a matriz \mathbf{A} usando apenas a

informação contida na matriz \mathbf{x} . Para tanto, é preciso fazer suposições tão gerais quanto possível [HYVÄRINEN, 2001]. Portanto, supõe-se que:

- a) As componentes s_i são estatisticamente independentes;
- b) As componentes têm distribuições não-gaussianas;
- c) Por motivos de simplicidade, a matriz \mathbf{A} seja quadrada.

2.3.2 Definição de Independência

Sejam y_1 e y_2 duas variáveis aleatórias. Tais variáveis são ditas independentes se a ocorrência ou não ocorrência de y_1 não influenciar na ocorrência ou não ocorrência de y_2 , e vice-versa. Independência estatística é definida em termos de densidade de probabilidade. Seja $p(y_1, y_2)$ a função densidade de probabilidade (pdf) conjunta de y_1 e y_2 . Então, $p_1(y_1)$ denota a pdf de y_1 :

$$p_1(y_1) = \int p(y_1, y_2) d_{y_2} \quad (2.3)$$

e similarmente para y_2 . Duas variáveis aleatórias são estatisticamente independentes se e somente se a pdf conjunta for:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2) \quad (2.4)$$

Podemos definir descorrelação de duas variáveis aleatórias y_1 e y_2 , com covariância zero, como:

$$E(y_1, y_2) = E(y_1)E(y_2) = 0 \quad (2.5)$$

Se duas variáveis são independentes, também são descorrelacionadas, mas o contrário não é verdadeiro.

2.3.3 Técnicas de Estimação das Componentes Independentes

Dadas n misturas lineares x_1, \dots, x_n de n componentes independentes. Para estimar as componentes deve-se encontrar a inversa da matriz \mathbf{A} , que é chamada de W . A solução para a estimação das componentes independentes pode ser descrita da forma:

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} \rightarrow \text{DPs} \quad (2.6)$$

Sendo D uma matriz diagonal não-singular, e P uma matriz de permutação qualquer. Observa-se imediatamente que y_i é uma combinação linear de s_i . Baseado no teorema do limite central, y_i é mais gaussiano do que qualquer s_i e torna-se menos gaussiano, quando de fato é igual a uma componente de \mathbf{s} .

Para finalmente estimar as componentes independentes, deve-se encontrar a matriz \mathbf{W} que minimiza a não gaussianidade de $\mathbf{W}\mathbf{x}$.

A não-gaussianidade é um elemento chave para a estimação do modelo de ICA, pois a matriz \mathbf{A} não é idêntica quando as componentes independentes têm distribuição gaussiana. Consideremos que o vetor \mathbf{x} é distribuído de acordo com o modelo de ICA na Equação 2.1, e que todas as componentes independentes têm distribuições iguais. Para estimar as componentes independentes, basta encontrar as combinações lineares corretas das variáveis da mistura x_i , de modo a

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} \quad (2.7)$$

Assim, pode-se expressar uma combinação linear de x_i por

$$\mathbf{y} = \mathbf{b}^T \mathbf{x} \quad (2.8a)$$

$$= \sum_i b_i x_i \quad (2.8b)$$

$$= b^T \mathbf{A} \mathbf{s} \quad (2.8c)$$

em que b deve ser determinado. A partir da Equação (2.8c) pode-se observar que y é uma combinação linear de s_i , com coeficientes dados por $q = b^T A$. Logo obtêm-se

$$y = q^T s \quad (2.9a)$$

$$= \sum_i q_i s_i \quad (2.9b)$$

Se \mathbf{b} corresponder a uma das linhas da inversa de \mathbf{A} , então \mathbf{y} será uma das componentes independentes e, nesse caso, apenas um dos elementos de \mathbf{q} será igual a 1, enquanto todos os outros serão iguais a zero. Não é possível determinar \mathbf{b} exatamente, mas podemos estimar seu valor com boa aproximação.

Uma forma de determinar \mathbf{b} é variar os coeficientes em \mathbf{q} e então verificar como a distribuição de $\mathbf{y} = \mathbf{q}^T \mathbf{s}$ muda. Já que, conforme o Teorema do Limite Central [PAPOULIS, 2002], a soma de duas variáveis aleatórias independentes é mais gaussiana que as variáveis originais, $\mathbf{y} = \mathbf{q}^T \mathbf{s}$ normalmente é mais gaussiana que qualquer uma das s_i e menos gaussiana quando se iguala a uma das s_i . Nesse caso, apenas um dos elementos q_i de \mathbf{q} é diferente de zero [HYVÄRINEN, 2001].

Como, na prática, os valores de \mathbf{q} são desconhecidos e sabe-se que, através das Equações (2.8a) e (2.9a), que:

$$\mathbf{b}^T \mathbf{x} = \mathbf{q}^T \mathbf{s} \quad (2.10)$$

podemos variar \mathbf{b} e observar a distribuição de $\mathbf{b}^T \mathbf{x}$. Portanto, pode-se tomar, como \mathbf{b} , um vetor que maximiza a não-gaussianidade de $\mathbf{b}^T \mathbf{x}$, sendo que esse vetor necessariamente

corresponde a $\mathbf{q} = \mathbf{A}^T \mathbf{s}$, vetor esse que possui apenas uma de suas componentes diferente de zero. Isso significa que \mathbf{y} na Equação (2.8a) é igual a uma das componentes independentes. Logo, a maximização da não-gaussianidade de $\mathbf{b}^T \mathbf{x}$ permite encontrar uma das componentes.

Negentropia como Medida de Não-Gaussianidade

Uma medida importante de não-gaussianidade é a negentropia, que é baseada na entropia. Tomando um vetor aleatório y cuja função densidade de probabilidade é $f(y)$, tem-se a entropia diferencial dada por

$$H(y) = - \int f(y) \log f(y) \quad (2.11)$$

Como um dos resultados fundamentais da Teoria da Informação, sabe-se que uma variável gaussiana tem a maior entropia entre todas as variáveis aleatórias de igual variância [HYVÄRINEN, 2001, PAPOULIS, 2002]. Isso quer dizer que uma versão modificada da entropia diferencial pode ser usada como medida de não-gaussianidade. Essa medida é chamada negentropia, definida por

$$J(\mathbf{y}) = H(y_{gauss}) - H(y) \quad (2.12)$$

em que y_{gauss} é uma variável aleatória de mesma matriz de covariância que y . A negentropia é sempre não-negativa, tem valor igual a zero, se e somente se y tem distribuição gaussiana e é invariante para transformações lineares inversíveis.

Em contraste às suas qualidades como medida de não-gaussianidade, a negentropia é de difícil estimação. Por isso, é necessária a utilização de aproximações usando, por exemplo, momentos de alta ordem. Logo:

$$J(\mathbf{y}) \approx \frac{1}{12} E \{y^3\}^2 + \frac{1}{48} \text{kurt}(y)^2 \quad (2.13)$$

sendo $\text{kurt}(y)$, a curtose de y , é definida como o momento de quarta ordem da variável aleatória y , expresso por

$$\text{kurt}(y) = E \{y^4\} - 3(E \{y^2\})^2 \quad (2.14)$$

no entanto, essa aproximação usa a curtose, que é uma medida de não-gaussianidade. Dessa forma, é mais conveniente utilizar outras abordagens, que inclusive substituem os momentos polinomiais y_3 e y_4 por outra função G . O método propõe a aproximação da negentropia baseado em expectativas E . [HYVÄRINEN, 2001]

$$J(y) \approx k_1(E\{G_1(y)\})^2 + k_2(E\{G_2(y)\} - E\{G_2(v)\})^2 \quad (2.15)$$

sendo k_1 e k_2 constantes positivas, v uma variável gaussiana de média zero, e $G_1\{y\} = y^3$ e $G_2\{y\} = y^4$

2.4 Seleção de Características mais Significantes

Forward-Selection é um método para encontrar a melhor combinação de características (variáveis) iniciando com uma única característica, e incrementando mais características, passo a passo [ORR, 1999]. Dessa forma, cada característica é adicionada no modelo de cada vez. A cada passo, cada característica que ainda não pertence ao modelo é testada para ser incluída. As características mais significativas, ou seja, que mostram um maior decréscimo na função de erro descrita acima são adicionados ao modelo, até que se consiga um subconjunto p , menor que o conjunto P selecionado.

Na técnica *Forward-Selection*, cada etapa envolve o crescimento da rede pela adição de uma função base (ou seja, uma característica). Adicionar uma função base nova

é uma das operações incrementais. A equação fundamental da técnica é:

$$P_{m+1} = P_m - \frac{P_m \cdot f_j \cdot f_j^T \cdot P_m}{f_j^T \cdot P_m \cdot f_j} \quad (2.16)$$

que expressa a relação entre P_m , a matriz de projeção de m camadas escondidas do subconjunto corrente e P_{m+1} , a projeção sucedente do j -ésimo membro do último conjunto de características adicionado. Os vetores $\{f_j\}_{j=1}^M$ são colunas de um conjunto selecionado de funções-bases

$$F = [f_1, f_2, f_M] \quad (2.17)$$

sendo $M \gg m$

A escolha das funções bases é baseada em encontrar erro-médio-quadrático.

Das regras de atualização para a matriz de projeção e da equação para o erro

$$T_m - T_{m+1} = \frac{(y^T \cdot P_m \cdot f_j)^2}{f_j^T \cdot P_m \cdot f_j} \quad (2.18)$$

sendo T_m o conjunto anterior e T_{m+1} o conjunto atual.

2.5 Redes Neurais como Classificadores

As Redes Neurais Artificiais (RNA's) foram desenvolvidas, originalmente, na década de 40, pelo neurofisiologista Warren McCulloch e pelo matemático Walter Pitts, que dentro do espírito cibernético, fizeram uma analogia entre células nervosas vivas e o processo eletrônico num trabalho publicado sobre "neurônios formais". O trabalho consistia num modelo de resistores variáveis e amplificadores representando conexões

sinápticas de um neurônio biológico.

Desde então, mais enfaticamente a partir da década 80, diversos modelos de redes neurais artificiais têm surgido com o propósito de aperfeiçoar e aplicar esta tecnologia. A Figura 2.1 apresenta a arquitetura de uma RNA.

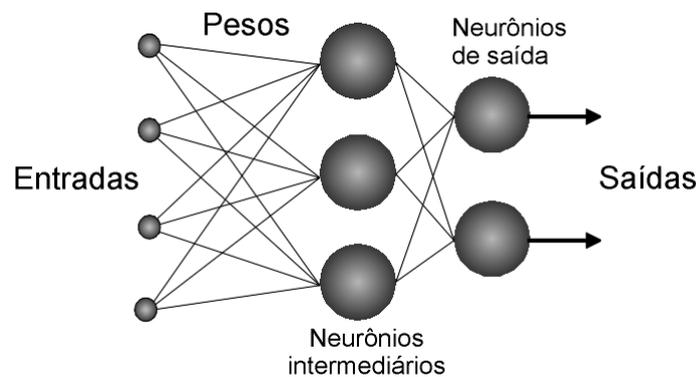


Figura 2.1: Arquitetura de RNA com duas saídas.

As variantes de uma rede neural são muitas, e combinando-as, pode-se mudar a arquitetura conforme a necessidade da aplicação. Basicamente, os itens que compõem uma rede neural e, portanto, sujeito a modificações, são os seguintes:

- Conexões entre camadas;
- Camadas intermediárias;
- Quantidade de neurônios;
- Função de transferência;
- Algoritmo de aprendizado;

Todas as RNA's possuem alguma regra de treinamento, em que os pesos de suas conexões são ajustados de acordo com os padrões apresentados, de tal forma que pode-se dizer que uma importante propriedade é a capacidade do aprendizado. Isso é feito através de um processo iterativo de ajustes aplicados aos pesos das conexões ao

qual denomina-se treinamento. O aprendizado ocorre quando a RNA atinge uma solução generalizada para uma classe de problemas.

Um conjunto de regras bem definidas para que a RNA possa aprender a solução de um problema é chamado de algoritmo de aprendizado. Existem muitos algoritmos de aprendizado específicos para determinados modelos de redes neurais, estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados. Outro ponto importante é a categorização das situações de aprendizado das RNA's. Neste contexto pode-se citar as seguintes classes de aprendizado [BRAGA, 2000], [KROSE e SMAGT, 1996]:

- Aprendizado Supervisionado, quando é utilizado um agente externo que indica à rede a resposta desejada para o padrão de entrada;
- Aprendizado Não Supervisionado, quando não existe um agente externo indicando a resposta desejada para os padrões de entrada.

2.5.1 Redes Neurais Perceptron Multicamadas

As Redes Neurais Perceptron Multicamadas (MLP) são as redes neurais mais frequentemente usadas em reconhecimento de padrões [DUDA e HART, 1973], [BISHOP, 1999]. Uma RNA do tipo MLP é constituída por um conjunto de nós fonte, os quais formam a camada da entrada da rede (*input layer*), uma ou mais camadas escondidas (*hidden layers*), e uma camada de saída (*output layer*), que extraem informações durante o aprendizado, e atribuindo coeficientes (pesos) às camadas de entrada.

O número de nós fonte na entrada da rede é determinado pela dimensionalidade do espaço de observação, que é responsável pela geração dos sinais de entrada. O número

de neurônios na camada de saída é determinado pela dimensionalidade requerida na resposta desejada. Assim, o projeto de uma rede MLP requer a consideração de três aspectos:

- I) A determinação do número de camadas escondidas;
- II) A determinação do número de neurônios em cada uma das camadas escondidas;
- III) A especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede.

Os aspectos I e II determinam a complexidade do modelo da RNA escolhida, e infelizmente, não há regras determinadas para tal especificação. A função das camadas escondidas em uma RNA é a de influir na relação entrada-saída da rede de uma forma ampla. Uma RNA com uma ou mais camadas escondidas é apta a extrair as características de ordem superior de algum desconhecido processo aleatório subjacente, responsável pelo comportamento dos dados de entrada, processo sobre o qual a rede está tentando adquirir conhecimento. A RNA adquire uma perspectiva global do processo aleatório, apesar de sua conectividade local, em virtude do conjunto adicional de pesos sinápticos e da dimensão adicional de interações neurais proporcionada pelas camadas escondidas.

O aspecto III envolve a utilização de algoritmos de treinamento supervisionado. As RNA's MLP's têm sido aplicadas na solução de diversos e difíceis problemas através da utilização de tais algoritmos [?]. O algoritmo de treino geralmente utilizado é o algoritmo de retropropagação, popularmente conhecido como *backpropagation*. O algoritmo *backpropagation* baseia-se na heurística do aprendizado por correção de erro. Este algoritmo pode ser visto como uma generalização do algoritmo LMS (*Least Mean*

Square), desenvolvido por Bernard Widrow [WIDROW, 1985]. Basicamente, o algoritmo *backpropagation* consiste de dois passos através das diferentes camadas do MLP: um passo direto e um passo reverso.

- No passo direto, um padrão de atividade do processo a ser aprendido (ou vetor de entrada) é aplicado ao nós de entrada do MLP e o seu efeito se propaga através da rede, camada por camada, produzindo na camada de saída a resposta do MLP a excitação aplicada (vetor de saída). Durante o passo direto os pesos sinápticos são todos fixos.

- No passo reverso, os pesos sinápticos são todos ajustados de acordo com a regra de aprendizado por correção de erro. Especificamente, a resposta do MLP à excitação é subtraída de um padrão de resposta desejado para aquela excitação aplicada, de forma a produzir um sinal de erro, de forma semelhante ao algoritmo LMS, como mostra a Figura 2.2. Este sinal de erro é, então, propagado de volta aos mesmos neurônios utilizados no passo direto, mas no caminho contrário do fluxo de sinal nas conexões sinápticas- daí o nome *Backpropagation*. Os pesos sinápticos são, então, ajustados de forma que a resposta obtida da MLP aproxime-se mais do padrão de resposta desejado, de tal forma que

$$e_j(n) = d_j(n) - y_j(n) \quad (2.19)$$

em que e_j é o sinal de erro, d_j o sinal desejado e y_j o sinal direto

O modelo de cada neurônio MLP inclui uma função de ativação não-linear. É importante salientar que esta não linearidade é suave, ao contrário da função utilizada no modelo Perceptron de Rosemblat (função Sigmóide). Uma forma comumente utilizada

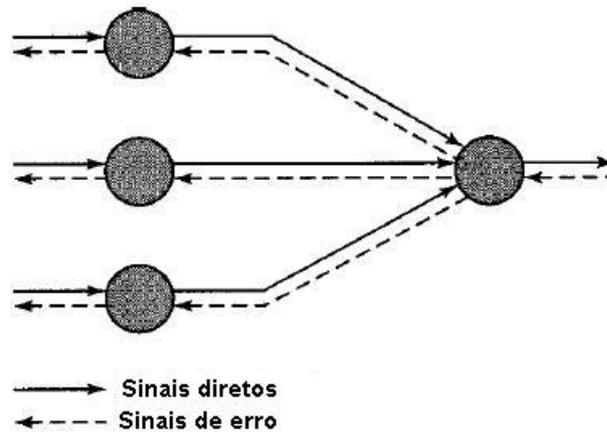


Figura 2.2: Ilustração das direções de dois fluxos básicos de um sinal em uma rede MLP: Propagação direta dos sinais e a retro-propagação dos sinais de erro.

de não linearidade que satisfaz este requisito é a não linearidade sigmodal definida pela função logística:

$$y_j = \frac{1}{1 + \exp(-v_j)} \quad (2.20)$$

sendo v_j o potencial de ativação (isto é, a soma ponderada de todas as entradas sinápticas mais a polarização), do neurônio e y_j é a saída do neurônio.

Durante o treinamento, As MLP's controem um espaço multi-dimensional definido pela ativação dos nós das camadas escondidas, de modos que as classes sejam mais separáveis possível. A superfície de separação se adapta aos dados.

3 O Método Proposto

Todas as técnicas descritas anteriormente foram aplicadas no método desenvolvido. A partir das mamografias utilizadas no trabalho, foram selecionadas as regiões de interesse (ROI). O diagrama do método proposto é mostrado na Figura 3.1, que consiste na seleção das regiões de interesse, na extração das características estatísticas da imagem usando ICA, a seleção de características mais significantes, usando a técnica *Forward-Selection* e a decisão final do diagnóstico, usando redes neurais *Multilayer Perceptron*.

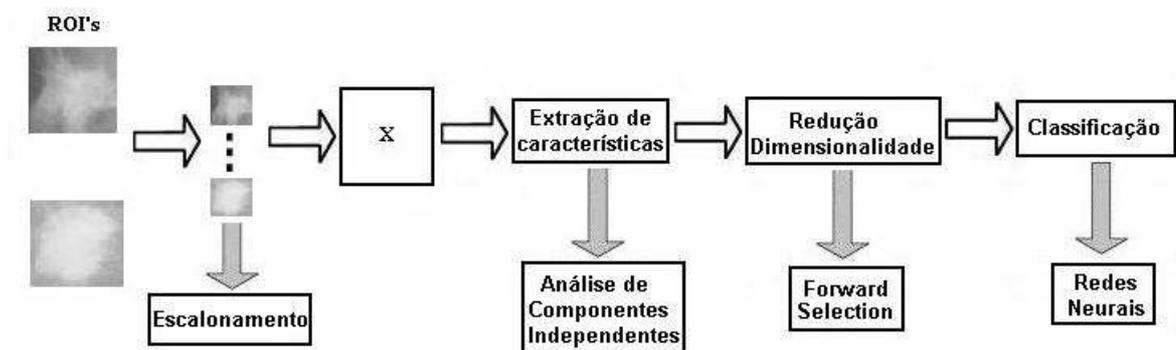


Figura 3.1: Diagrama do método proposto.

3.1 ICA aplicado em Processamento de Imagens

Generalizando para processamento de imagens, ICA é capaz de encontrar imagens-bases que originaram uma determinada mistura, a partir de uma imagem observada, supondo independência estatística entre elas, ou seja, tendo observado um conjunto de dados, e supondo que eles tenham sido gerados a partir de uma mistura de

outros dados. A técnica é capaz de estimar imagens-bases que originaram a mistura, e os pesos de cada fonte na mistura, a partir da análise estatística da imagem observada.

Hubel e Wiesel [HUBEL e WIESEL, 1968] realizaram experimentos clássicos em neurônios do córtex visual e chegaram a conclusão que o córtex visual funciona com um detector de características. Barlow [BARLOW, 1989] conduziu assim a proposta que nosso córtex visual, detector de características, pode trabalhar com processos de redução de redundância [BARLOW, 1989], em que a ativação de cada detector de características, é estatisticamente independente dos outros detectores o quanto possível. Tal codificação envolve a dependência estatística de todas as ordens, mas a grande maioria dos estudos usam estatísticas de segunda ordem, PCA por exemplo, em que a estatística de segunda ordem é suficiente apenas para descorrelacionar um conjunto de saídas de um detector de características, e não remover a independência estatística entre os detectores. Por essa razão, PCA não é suficiente para prever a formação de detectores de características. O sistema perceptual é exposto para uma série de pequenas janelas, como mostra a Figura 3.2, extraídas de uma ou mais imagens maiores. Considere que sejam observadas n janelas aleatórias x_1, x_2, \dots, x_n de uma imagem, modelada por uma combinação linear de n funções-bases:

$$x_n = a_1 \cdot s_{j1} + a_2 \cdot s_{j2} + a_3 \cdot s_{j3} + \dots + a_n \cdot s_{jn} \quad (3.1)$$

em que cada janela x_n , assim como cada componente independente s_n , que no nosso caso chama-se de imagem-base, seja uma variável aleatória. Sem perda de generalidade, pode-se reescrever a equação acima da mesma da Equação 2.2

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (3.2)$$

É preciso estimar tanto a matriz de imagens-bases \mathbf{s} , quanto a matriz de componentes independentes \mathbf{A} , que também é desconhecida, pois temos apenas o conjunto de janelas X .

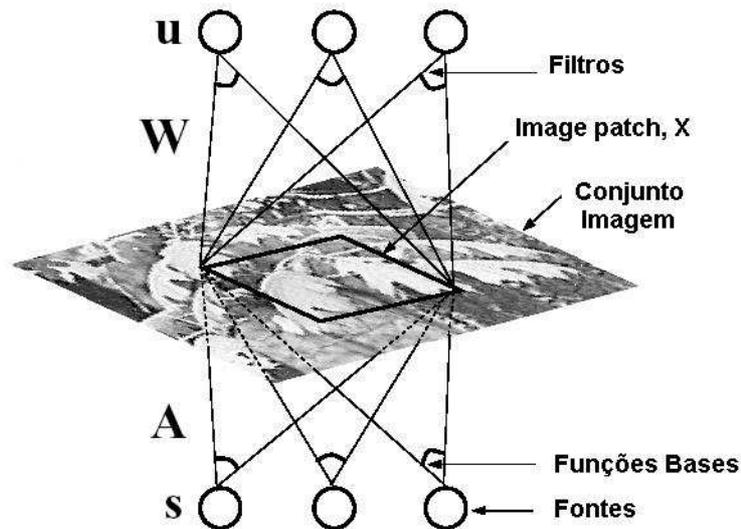


Figura 3.2: Modelo de separação linear de imagens.

Cada componente deste vetor tem sua própria função base associada, e representa uma fonte não-observada da imagem (imagem-bases).

O objetivo do sistema perceptual é transformar as janelas x , linearmente, com a matriz de filtros W , resultando então em um vetor:

$$\mathbf{u} = \mathbf{W} \cdot \mathbf{x} \quad (3.3)$$

recuperando assim as imagens-bases não-observadas \mathbf{s} , possivelmente de ordem diferente.

Representando por \mathbf{P} , uma matriz de permutação arbitrária, e por \mathbf{S} , uma matriz de escalonamento arbitrária, tal sistema converge para:

$$\mathbf{u} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{P}\mathbf{S}\mathbf{s} \quad (3.4)$$

O escalonamento e a permutação das fontes \mathbf{s} é arbitrário. Assim, considera-se as fontes a serem definidas, tal que $\mathbf{P}\mathbf{S} = \mathbf{I}$ (sendo \mathbf{I} a matriz identidade). As funções bases (colunas de \mathbf{A}) e os filtros que recuperam as fontes (linhas de \mathbf{W}), tem a seguinte relação:

$$\mathbf{W} = \mathbf{A}^{-1} \quad (3.5)$$

Tudo que falta agora é definir um algoritmo de aprendizado para encontrar \mathbf{W} (e também \mathbf{A}), para enfim estimarmos as fontes \mathbf{s} , o que foi visto na seção 2.3.3.

4 Material

Neste capítulo será descrita a base de dados e os *softwares* utilizados.

4.1 Base de dados MIAS

A base de dados utilizada neste trabalho foi cedida pelo *Mammographic Institute Analysis Society* [SUCKLING, 1994]. Esta base de dados é composta por 322 mamogramas de mamas esquerda e direita, de 161 pacientes, sendo 53 mamogramas com diagnóstico maligno, 69 com diagnóstico benigno, e 206 de diagnóstico normal. As mamografias têm um tamanho de 1024x1024 pixels, e resolução de 200 micron. As anormalidades são classificadas de acordo com a classe encontrada (calcificação, massa circunscrita, distorções arquiteturais, assimetrias, e outras massas, sem formas definidas).

Esta base de dados contém também um arquivo, explicando detalhadamente cada mamografia, como por exemplo, a classe da anormalidade, as coordenadas xy do centro da anormalidade, e o raio aproximado da anormalidade, em pixels. A Tabela 4.1 mostra detalhes de algumas mamografias utilizadas.

Desta base de dados, selecionamos 50 mamografias com diagnóstico benigno, 50 com diagnóstico maligno, e 200 com diagnóstico normal, somando no total 300 mamografias. De cada mamograma, as regiões de interesse (ROI) foram selecionadas manualmente, contendo a lesão, no caso de mamografias com diagnóstico maligno e benigno. as ROIs foram selecionadas através das coordenadas xy da imagem, do centro

Tabela 4.1: Exemplo do diagnóstico mostrado em algumas mamografias da base de dados MIAS

Número	Classe	Anormalidade	Diagnóstico	Coordenadas	Raio (<i>pixels</i>)
mdb001	G	CIRC	B	535 , 425	197
mdb002	G	CIRC	B	522 , 280	69
mdb003	D		N		
mdb004	D		N		
mdb005	F	CIRC	B	477 , 133	30
mdb005	F	CIRC	B	500 , 168	26
mdb006	F		N		
mdb007	G		N		

da anormalidade, informação essa obtida da própria base de dados. Para mamogramas normais, foi selecionada aleatoriamente uma ROI. Somente o músculo peitoral não foi considerado ROI, mas tecido granular e tecido gorduroso foi considerado. Cada ROI tinha tamanhos diferentes, então foi preciso reescalonar cada ROI. Sendo assim, cada região de interesse ficou com um tamanho de 24x24 pixels. A Figura 4.1 exemplifica a seleção das ROI's de mamogramas de diagnóstico benigno, maligno e normal, respectivamente.

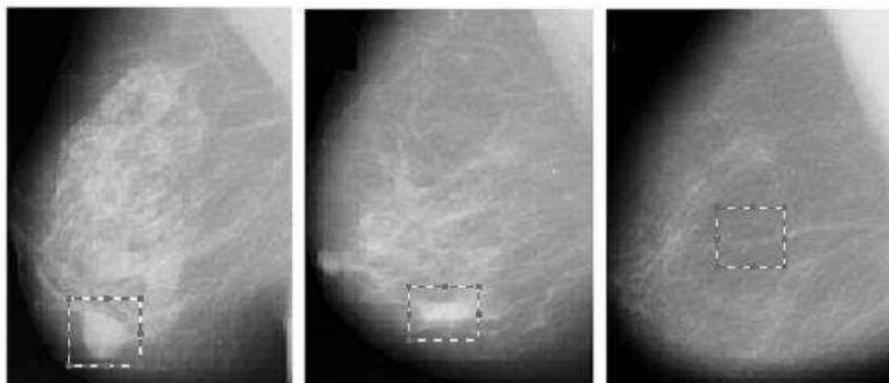


Figura 4.1: Seleção das ROI's de tecidos benignos, malignos e normais.

4.2 Softwares Utilizados

Após a seleção das regiões de interesse (ROI) de cada mamograma, utilizamos o software **MatLab**, para gerar a matriz de coeficientes **A**, através do algoritmo *FastIca*. Logo após, foi necessário fazer uma seleção de características encontradas de cada ROI, já que o vetor de entradas para rede neural ficaria muito grande, o que poderia comprometer o desempenho da rede. A seleção de características foi realizada através do algoritmo *Foward-Selection*. Depois de escolhidas as características mais significantes de cada ROI, tais características serviram de entrada para a rede neural artificial Perceptron Multicamadas, para que fosse dada a decisão final de diagnóstico. Tanto a seleção de características, como a classificação através da rede neural foram realizadas pelo programa **Trajan Neural Networks**.

5 Resultados e Discussão

Neste capítulo, será descrito a aplicação das técnicas utilizadas no método proposto, os resultados encontrados e a discussão do resultados.

5.1 Extração de Parâmetros Utilizando ICA

Considera-se que uma imagem é formada pela soma de imagens bases s_1, \dots, s_n , estatisticamente independentes entre s_1 . A imagem então é formada por uma combinação linear das imagens s , n imagens-bases modeladas por uma combinação linear de n variáveis a_1, a_n , como observado na Figura 5.1: [CHRISTOYIANNI *et al*, 2002],[HYVÄRINEN, 2001].

$$x_i = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{in} \cdot s_n \quad (5.1)$$

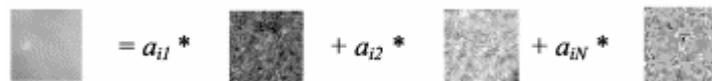


Figura 5.1: Região de interesse com uma combinação linear de imagens bases mútua e estatisticamente independentes entre si.

No modelo acima, apenas as variáveis x_i são conhecidas, e a partir delas serão estimados os coeficientes da mistura a_{ij} e as componentes independentes s_j , ou seja:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (5.2)$$

Sendo \mathbf{x} a matriz de mistura, \mathbf{A} as funções-bases e \mathbf{s} as fontes que compõem a imagem.

Na equação, \mathbf{x} foi representada usando as ROI's escolhidas de cada mamografia. As imagens com as ROI's foram reescaladas para um tamanho de 24×24 *pixels*, e transformadas em um vetor unidimensional, de tamanho:

$$P = P_x \cdot P_y \quad (5.3)$$

sendo P_x as linhas, e P_y as colunas de P , e P tendo um tamanho de 1×576 .

Cada amostra representa uma linha da matriz de mistura. A matriz \mathbf{x} é representada por amostras na dimensão de P , ou seja, 1×576 . Então, cada linha da matriz \mathbf{A} corresponde a uma ROI, e cada coluna corresponde a um peso atribuído para a imagem base, ou seja, um parâmetro de entrada para a rede neural [CHRISTOYIANNI *et al*, 2002].

O algoritmo utilizado pra fazer a extração de parâmetros através de ICA foi o *FastICA*, que é um algoritmo comumente utilizado para resolver problemas de separação cega de fontes (BSS), onde queremos estimar as funções bases a partir da matriz de mistura \mathbf{X} .

Este algoritmo é baseado em iterações de ponto fixo [HYVÄRINEN e OJA, 1997], [HYVÄRINEN, 1997], [HYVÄRINEN, 1999]. Em Hyvarinen [HYVÄRINEN e OJA, 1997], o algoritmo de ponto fixo foi introduzido usando curtose, e em [HYVÄRINEN, 1997], [HYVÄRINEN, 1999], o algoritmo *FastIca* foi generalizado para outras funções. Para dados pré-processados por branqueamento (*whitening*), o algoritmo *FastIca* tem a seguinte forma:

$$w(k) = E\{x.g(w(k-1)^T x)\} - E\{g'(w(k-1)^T x)\}w(k-1) \quad (5.4)$$

em que o vetor de pesos w é então normalizado para a norma unitária após cada iteração, e

$$\mathbf{w}^{-1} = \mathbf{A} \quad (5.5)$$

A função g é derivada da função G usada nas funções de custo em geral, ou seja:

$$J_{G(y)} = |E_y\{G(v)\}|^p \quad (5.6)$$

sendo v uma variável aleatória, y normalizado para variância unitária, e o expoente $p=1,2$.

Uma forma básica do algoritmo *FastICA* é dado como segue:

1. Escolha um vetor inicial aleatório w ;
2. Faça $w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w$;
3. Faça $w = \frac{w^+}{\|w^+\|}$;
4. Se não convergir, volta ao passo 2.

As esperanças são estimadas, na prática, usando a média das amostras sobre uma quantidade suficiente grande dos dados de entrada.

Dentre as vantagens do algoritmo *FastICA*, podemos citar a rápida convergência, a simplicidade do algoritmo, e seu processamento paralelo, distribuído, computacionalmente simples, que requer pouco espaço de memória.

Usando o algoritmo *FastICA* e a matriz \mathbf{x} , obtemos a matriz de funções bases \mathbf{A} , que contém as características de cada amostra.

A Figura 5.2 exemplifica as imagens bases encontradas usando as funções bases das ROI's malignas, benignas e normais, respectivamente. Podemos observar claramente a diferença entre as imagens bases de cada classe. Então, pode-se observar que as funções bases dos tecidos benignos são diferentes das funções bases dos tecidos malignos.

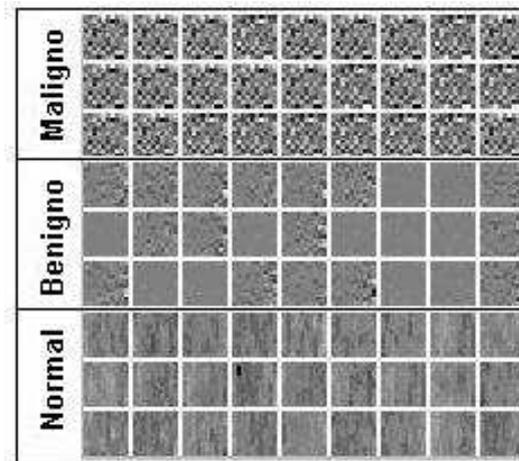


Figura 5.2: Amostra de imagens bases obtidas de ROI's de tecidos normais, benignos e malignos.

5.2 Variáveis Seleccionadas

As 300 amostras da base de dados foram divididas em 150 para treino da rede, e 150 para teste. Sendo assim, dividiu-se a matriz \mathbf{X} , de dimensão 300×576 , em x_{treino} , e em x_{teste} , ambas de dimensão 150×576 . Como o algoritmo *FastIca* gera a matriz de características \mathbf{A} quadrada, então cada ROI era composta de um vetor de características de dimensão 1×150 . Das 150 características de cada ROI, o algoritmo *Forward-Selection* selecionou apenas 45, sendo as mais significantes do conjunto.

5.3 Configuração da Rede Neural MLP

O programa **Trajan Neural Networks** selecionou as cinco melhores configurações de redes neurais MLP. O número de neurônios na camada escondida foi incrementado até 50, porém a rede com 21 neurônios apresentou melhor desempenho. O erro encontrado no treinamento foi de 0,008, que, segundo o programa utilizado, indicou que a rede neural escolhida já estava especializada. Sendo assim, a melhor rede neural MLP encontrada ficou com a seguinte configuração: 45 neurônios na camada de entrada, 21 na camada escondida e 3 na camada de saída, conforme mostra a Figura 5.3:

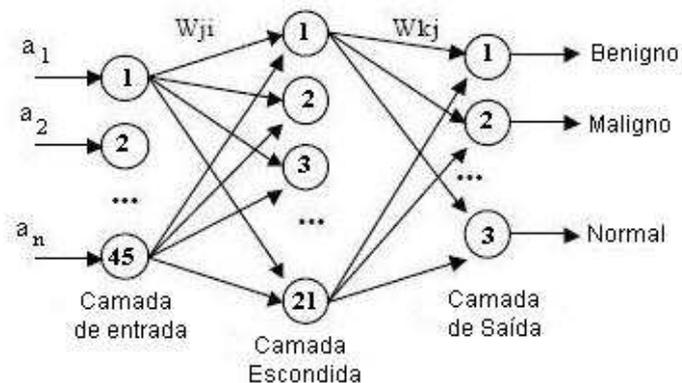


Figura 5.3: Configuração de rede neural MLP, com 45 neurônios de entrada, 21 na camada intermediária, e 3 neurônios na camada de saída.

A Tabela 5.1 mostra o desempenho da aplicação da técnica de ICA com redes neurais MLP para a classificação de cada ROI em malignas, benignas e normais.

5.4 Discussão

Baseado na Tabela 5.1, o método obteve uma taxa de sucesso média de 97,83%, discriminando ROI de tecidos malignos, benignos e normais.

Considerando que toda mamografia onde foi encontrada algum tipo de lesão,

Tabela 5.1: Classificação das ROI de tecidos Benignos, Malignos e Normais

	Benigno	Maligno	Normal
Benigno	48	0	5
Maligno	0	50	0
Normal	2	0	195
Sucesso (%)	96	100	97,5

seja benigna ou maligna, foi considerado anormal, e que todo mamograma sadio foi considerado normal, a Tabela 5.2 mostra a classificação das ROI's em normais e anormais.

Tabela 5.2: Classificação das ROI's em normais e anormais

	Normal	Anormal
Normal	195	2
Anormal	5	98
Sucesso (%)	97.5	98

Sensibilidade e especificidade são as medidas mais utilizadas para descrever um sistema de diagnóstico. Sensibilidade (**S**) é a proporção de verdadeiros positivos que são corretamente identificados pelo teste, e é definida por $S = VP/(VP+FN)$. Especificidade (**E**) é proporção de verdadeiros negativos que são corretamente identificados no teste, e é dada por $E = VN/(VN+FP)$. Sendo FN Falso Negativo, FP Falso Positivo, VN Verdadeiro Negativo e VP Verdadeiro Positivo.

Baseado na Tabela 5.2, a Especificidade encontrada foi de 97,5% e a sensibilidade de 98,00%. O método obteve 98 diagnósticos verdadeiro-positivos, 195 verdadeiro-negativos, 5 falso-positivos e 2 falso negativos.

O erro encontrado, levando em consideração as três classes (normal, benigna e

maligna), foi de apenas 2,17%, que além de pequeno pode ser justificado. Na seleção das ROI's, as mesmas foram capturadas como um quadrado de dimensão XY , mas todas as lesões tinham formas diferente de um quadrado. Assim, além de capturar a lesão, no caso de mamografias com diagnóstico anormal, também foi capturado tecido sadio da mama.

O programa *Trajan Neural Networks* indicou que, com o erro de 0,008, a rede já estava especializada, porém não indicou se todas as 150 amostras que foram reservadas para o teste foram utilizadas, ou seja, não se soube se para chegar a especialização da rede foram necessárias todas as 150 amostras.

6 Conclusão e trabalhos futuros

Este trabalho propõe um método de classificação de mamogramas digitais utilizando Análise de Componentes (ICA) e Redes Neurais Perceptron Multicamadas. O método proposto utilizou o algoritmo *FastICA* para extrair um conjunto de características em conjunto com a técnica *Foward Selection*, para selecionar as características mais significantes para que cada ROI pudesse ser melhor representada. O conjunto destas técnicas permitiu que o resultado encontrado fosse melhor que em métodos que utilizam-se de outras técnicas.

O presente resultado demonstrou que a técnicas utilizadas neste trabalho conseguiram classificar tecidos normais, tecidos de lesões benignas e malignas. Observou-se também, que a rede neural MLP obteve uma taxa de sucesso média de 97.83%, com um erro de apenas 2,7%. Com um erro menor, pode-se obter um menor número de biópsias desnecessárias, e também um menor número de casos de câncer descobertos tardiamente.

Relacionando com trabalhos anteriores, o método proposto utilizou o algoritmo *FastIca* em conjunto com algoritmo de seleção de características *Foward-Selection*, ao passo que no método descrito em [CHRISTOYIANNI *et al*, 2002], que obteve com resultado 79.31% discriminando mamogramas de diagnóstico normais, malignos e benignos, utilizou o algoritmo *Infomax*, em conjunto com redes neurais com Função de Ativação de Base Radial.

A partir do proposto no presente trabalho, futuramente, pretende-se:

- Testar a eficácia do método com uma base de dados regional, ainda em fase

de construção.

- Elaborar um programa, que seja de baixo custo computacional e simples de ser manuseado, para ser utilizado por médicos e radiologistas.

- Adaptar o método para classificação de outras estruturas, que utilizem como entrada raio-x, por exemplo: Nódulo pulmonar.

Este trabalho deu origem aos seguintes artigos:

- Lúcio F. A. Campos, Aristófanés C. Silva, Allan Kardec Barros: "Independent Component Analysis and Neural Networks Applied for Classification of Malignant, Benign and Normal Tissue in Digital Mammography", V International Workshop on Biosignal Interpretation, Japan, May 2005

- Lúcio F. A. Campos, Aristófanés C. Silva, Allan Kardec Barros: "Diagnosis on Breast Cancer Using Independent Component Analysis and Neural Networks ", X CIARP - LNCS 3773 - Springer- Verlag Ed.

- Lúcio F. A. Campos, Aristófanés C. Silva, Allan Kardec Barros: "Independent Component Analysis and Neural Networks Applied for Classification of Malignant, Benign and Normal Tissue in Digital Mammography", Escolhido como Special Issue do "V International Workshop on Biosignal Interpretation, Japan" a ser publicado no Journal of Methods of Information in Medicine, no final de 2006.

Referências

- [AZEVEDO,1994] AZEVEDO, C. M. , Radiologia da Mama. Rio de Janeiro, INCa, 1994.
- [BAUER,1980] BAUER, W.; IGOT, J.P.; LE, G.Y.: Chronologic du cancer mammaire Utilisant un Modele de Croissance de Gompertz. Ann Anat Pathol 25:39-56, 1980
- [BOYD,1995] BOYD, N.F.; BYNG, J.W.; JONG, R.A.; FISHELL, E.K.; LITTLE, L.E.; MILLER, A.B.; LOCKWOOD, G.A.; TRITCHLER, D.L.; YAFFE, M.J. Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study. Journal of the National Cancer Institute, v. 87, p. 670-675, 1995.
- [BLAND,2000] BLAND, K.I.; COPELAND, E.M.: A Mama: Tratamento Compreensivo das Doenças Benignas e Malignas, Ed., Manole Ltda
- [BARLOW, 1989] BARLOW H.B.1989. Unsupervised learning . Neural Computation 1. 295-311.
- [BELL e SEJNOWSKI, 1995] BELL A. J. , SEJNOWSKI T. J, 1995. An information maximization approach to blind separation and blind deconvolution. Neural Computation, 7, 1129,1159.
- [BRAGA, 2000] BRAGA, A.P . Redes neurais Artificiais, teoria e aplicações, LTC. 2000.
- [BISHOP, 1999] BISHOP, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, New York (1999)
- [CAVALCANTE, 2006] CAVALCANTE, André B.; MANDIC, Danilo P.; RUTKOWSKI, Tomasz; BARROS, Allan K. Speech Enhancement Based on the Response Features of Facilitated EI Neurons. Proceedings of 6th International Conference on Independent Component Analysis and Blind Source Separation, Março 2006, Charleston, South Carolina, USA.
- [CAMPANINI *et al*, 2002] CAMPANINI Renato, BAZZANI Armando, et al. A novel approach to mass detection in digital mammography based on Support Vector Machines (SVM). In proceedings of th the 6 International workshop in digital Mammography (IWDM), pages 399-401, Bremem, Germany, 2002, Springer Verlag.
- [CASTRO] CASTRO, F.C., CASTRO, M.C.: Multilayer Perceptron. Notas de aula do Mestrado em Engenharia Elétrica da PUCRS.2004
- [CHRISTOYIANNI *et al*, 2002] CHRISTOYIANNI I., KOUTRAS A., KOKKINAKIS G., "Computer aided diagnosis of breast cancer in digitized mammograms", Comp. Med. Imag. e Graph., 26:309-319, 2002.

- [COMON, 1994] COMON P. 1994. Independent component analysis, a new concept? *Signal Processing*, 36, 287-314.
- [COVER e THOMAS, 1990] COVER T.M. , THOMAS J.A. *Elements of information theory* . John Wiley.
- [DUDA e HART, 1973] DUDA, R.O., HART, P.E.: *Pattern Classification and Scene Analysis*. Wiley- Interscience Publication, New York (1973)
- [GONZALES e WOODS, 1992] GONZALES, R.C.; WOODS, R. E. *Digital image processing*. Addison Wesley, 1992.
- [GIROLAMI e FYFE, 1997] GIROLAMI M., FYFE C.: An extended exploratory pursuit network with linear and non-linear anti-hebbian lateral connections applied to the cocktail party problem. *Neural Networks*, 10:1607-1678,1997
- [HANCOCK e BADDELEY, 1992] HANCOCK P.J.B., BADDELEY R.J. , Smith L.S.1992. The principal components of natural images. *Network*, 3, 61-72.
- [HUBEL e WIESEL, 1968] HUBEL D.H., WIESEL T.N.. Receptive fields and functional architecture of monkey striate cortex. *Journal Physiology.*, 195:215 244.
- [HYVÄRINEN e OJA, 1997] HYVÄRINEN and E. OJA. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483-1492, 1997.
- [HYVÄRINEN, 1997] HYVÄRINEN, Aapo. A family of fixed-point algorithms for independent component analysis. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'97)*, pages 3917-3920, Munich, Germany, 1997.
- [HYVÄRINEN, 1999] HYVÄRINEN, Aapo . Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 1999.
- [HYVÄRINEN, 2000] HYVÄRINEN A. OJA, E. *Independent Component Analysis: Algorithms and Applications*. *Neural Networks*, 13(4-5):411-430, 2000.
- [HYVÄRINEN, 2001] HYVÄRINEN A., KARHUNEN J., OJA E. *Independent Component Analysis*. Nova York: John Wiley & Sons. 2001. 481p.
- [INCa,2005] INCa, Instituto Nacional do Câncer. Endereço na internet: <http://www.inca.gov.br/acessado> em 04/05/2005.
- [JUTTEN e HERAULT, 1991] JUTTEN C., HERAULT J. Blind separation of sources. *signal processing*, 24:1-10,1991.
- [KARHUNEN e JOUTSENSALO, 1997] KARHUNEN J., JOUTSENSALO J. Representation and separation of signals using nonlinear PCA type learning. *Neural Networks*, 8:486-504,1997.
- [KROSE e SMAGT, 1996] KROSE B. ,SMAGT P. *An Introduction to neural Networks*. University of Amsterdã. 1996
- [LEITE *et al*, 2004] LEITE L. ,SILVA E. A. B. da , e PEREIRA W. C. A.(2004).Análise de Componentes Independentes Aplicada à identificação de regiões lesionadas em mamogramas.In *proceedings do III CLAEB, João Pessoa 2004*.

- [MARR e HILDRETH, 1968] MARR D. , HILDRETH E. Theory of edge detection
 Proceedins R. Society London. 207.188-217.1968
- [MOUSA *et al*, 2005] MOUSA Rafayah , MUNIB Qutaishat, and MOUSSA Abdallah.
 Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. Expert
 Systems with Applications, Volume 28, Issue 4, May 2005, Pages 713-723
- [PETRICK E BERKMAN, 2002] NICHOLAS Petrick, BERKMAN Sahiner,
 HEANGPING Chan, MARK A. Helvie, SOPHIE aquerault, and LUBOMIR M.
 Hadjiiski. Breast Cancer Detection: Evaluation of a mass-Detection Algorithm for
 Computer-aided Diagnosis. Experience in 263 patients Radiology 224:217-224.2002
- [ORR, 1999] ORR, Mark, Introduction to Radial Basis Function Networks disponível
 em:<http://www.anc.ed.ac.uk/rbf/intro/node51>. acessado em 23/02/2006
- [PAPOULIS, 2002] PAPOULIS, Athanasios; PILLAI, S. Unnikrishna. Probability,
 Random Variables and Stochastic Processes. 4 e.d. Nova York: McGraw-Hill. 2002.
 852 p.
- [SANTOS, 2002] SANTOS V. T.: Segmentação de Imagens Mamográficas para Detecção
 de Nódulos de Mamas Densas. Dissertação. São Carlos, 2002.
- [SUCKLING, 1994] SUCKLING J. et al (1994): The Mammographic Image Analysis
 Society Digital Mammogram Database Exerpta Medica. International Congress
 Series 1069 pp375-378.
- [VERMA e ZANGH, 2006] VERMA Brijesh, ZANGH Ping. A novel neural-genetic
 algorithm to find the most significlant combination of features in digital
 mammograms. ARTICLE Applied Soft Computing, In Press, Corrected Proof,
 Available online 24 February 2006. Science direct
- [VON ZUBEN, 2001] VON ZUBEN F. J., CASTRO L. N. : Redes Neurais com Funções
 de Ativação de Base Radial. DCA/FEEC/Unicamp. 2001
- [WIDROW, 1985] B. Widrow and Samuel D. Stearns. "Adaptive Signal Processing".
 Prentice-Hall signal processing series, 1985.
- [ZHANG, 2002] ZHANG, L., SANKAR, R., QIAN, W., "Advances in micro-calcification
 clusters detection in mammography", Comp. in Biology e Medicine, v. 32, pp. 515-
 528, 2002.