

UnB - UNIVERSIDADE DE BRASÍLIA
FGA - FACULDADE GAMA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
BIOMÉDICA

SISTEMA DE AUXÍLIO NA AVALIAÇÃO DE CALCIFICAÇÕES
MAMÁRIAS POR PROCESSAMENTO DIGITAL DE IMAGENS
E INTELIGÊNCIA ARTIFICIAL

FÁTIMA GISELE GOMES ELPÍDIO

ORIENTADORA: Dra. LOURDES MATTOS BRASIL

CO-ORIENTADORA: Dra. JANICE MAGALHÃES LAMAS

DISSERTAÇÃO DE PÓS-GRADUAÇÃO STRICTO SENSU EM
ENGENHARIA BIOMÉDICA

PUBLICAÇÃO: 003/2012

BRASÍLIA/DF: AGOSTO – 2012

UnB - UNIVERSIDADE DE BRASÍLIA
FGA - FACULDADE GAMA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
BIOMÉDICA

SISTEMA DE AUXÍLIO NA AVALIAÇÃO DE CALCIFICAÇÕES
MAMÁRIAS POR PROCESSAMENTO DIGITAL DE IMAGENS E
INTELIGÊNCIA ARTIFICIAL

FÁTIMA GISELE GOMES ELPÍDIO

DISSERTAÇÃO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO DA
FACULDADE GAMA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM
ENGENHARIA BIOMÉDICA.

APROVADA POR:

Profa. Dra. Lourdes Mattos Brasil
(Orientadora)

Dra. Janice Magalhães Lamas
(Co-orientadora)

Prof. Dr. Cristiano Jacques Miosso
(Examinador Interno)

Prof. Dr. Renato da Veiga Guadagnin
(Examinador Externo)

BRASÍLIA/DF, 31 DE AGOSTO DE 2012.

FICHA CATALOGRÁFICA

ELPÍDIO, FÁTIMA GISELE GOMES

Sistema de auxílio na avaliação de calcificações mamárias por processamento digital de imagens e inteligência artificial, [Distrito Federal] 2012.

156p., 210 x 297 mm (FGA/UnB Gama, Pós-Graduação *Stricto Sensu*, Engenharia Biomédica, 2012). Dissertação de Pós-Graduação *Stricto Sensu* em Engenharia Biomédica, Faculdade Gama, Programa de Pós-Graduação.

- | | | |
|---------------------------|-------------------------------------|--------------------------------|
| 1. Calcificações mamárias | 2. BI-RADS | 3. Raciocínio Baseado em Casos |
| 4. Rede Neural Artificial | 5. Processamento Digital de Imagens | |
| I. FGA UnB Gama/ UnB. | | |

REFERÊNCIA BIBLIOGRÁFICA

ELPÍDIO, F. G. G. (2012). Sistema de auxílio na avaliação de calcificações mamárias por processamento digital de imagens e inteligência artificial. Dissertação de Mestrado em Engenharia Biomédica, Publicação, Publicação 003/2012, Programa de Pós-Graduação em Engenharia Biomédica, Faculdade Gama, Universidade de Brasília, Brasília, DF, 156p.

CESSÃO DE DIREITOS

AUTOR: Fátima Gisele Gomes Elpídio.

TÍTULO: Sistema de auxílio na avaliação de calcificações mamárias por processamento digital de imagens e inteligência artificial.

GRAU: Pós-Graduação *Stricto Sensu* em Engenharia Biomédica. ANO: 2012.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de pós-graduação e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação pode ser reproduzida sem a autorização por escrito do autor.

Fátima Gisele Gomes Elpídio

Quadra 05, Lote 3, Casa 2 – COND. RESID. Athos I, Jardim dos Ipês

CEP 72874670 Valparaíso de Goiás, GO – Brasil.

"Se enxerguei longe, foi porque me apoiei nos ombros de gigantes."

Isaac Newton

AGRADECIMENTOS

Inicialmente agradeço a Deus e a minha família por serem minha base e motivação para realizar meus ideais.

À Dra. Lourdes Mattos Brasil, minha orientadora, por seu apoio e amizade incondicional, por acreditar e investir no meu potencial, guiando-me e fornecendo-me os subsídios necessários para concretização desse trabalho.

À Dra. Janice Lamas, minha co-orientadora, pela gentileza de compartilhar comigo seu conhecimento, experiência, material e momentos preciosos do seu tempo. Sua participação foi fundamental para a concretização dessa pesquisa. Nesse mesmo contexto, agradeço também aos profissionais da Clínica Janice Lamas, em especial à Tâmara Almeida de Moraes, técnica em mamografia, pelo apoio na realização das reuniões com a Dra. Janice e auxílio na disponibilização dos materiais empregados nessa pesquisa.

Ao Prof. Cristiano Miosso, cujo apoio e orientações foram essenciais no decorrer das pesquisas realizadas no âmbito desse trabalho.

À CAPES pelo apoio financeiro concedido para a realização dessa pesquisa.

Por fim, agradeço a todos que contribuíram direta ou indiretamente na concretização desse trabalho.

RESUMO

SISTEMA DE AUXÍLIO NA AVALIAÇÃO DE CALCIFICAÇÕES MAMÁRIAS POR PROCESSAMENTO DIGITAL DE IMAGENS E INTELIGÊNCIA ARTIFICIAL

Autor: Fátima Gisele Gomes Elpídio

Orientadora: Profa. Dra. Lourdes Mattos Brasil

Co-orientadora: Dra. Janice Magalhães Lamas

Programa de Pós-Graduação em Engenharia Biomédica

Brasília, Agosto de 2012.

A melhor forma de prevenção e controle do câncer de mama, o segundo tipo de câncer com maior incidência mundial, continua sendo o diagnóstico precoce. A mamografia é considerada o principal método de detecção precoce do câncer de mama. As calcificações (acúmulo de cálcio em regiões da mama) são frequentemente encontradas em exames de rastreamento da mama. A distinção entre calcificações tipicamente benignas e calcificações suspeitas de malignidade é uma tarefa complexa. O sistema *Breast Image Reporting and Data System* (BI-RADS) normatiza a descrição e o manejo dos achados mamográficos entre os especialistas da área médica. Este trabalho fundamenta-se em técnicas de Processamento Digital de Imagens para realizar a extração de informações morfológicas das calcificações e uma Rede Neural Artificial (RNA) *Perceptron* Múltiplas Camadas (PMC) para classificar as calcificações analisadas de acordo com a classificação morfológica definida pelo BI-RADS. O desempenho da RNA PMC foi avaliado em um teste controlado com calcificações sintéticas, geradas para simular calcificações reais, e validado com calcificações reais, extraídas de mamografias cedidas pela clínica de estudo. Adicionalmente foi desenvolvido um sistema de Raciocínio Baseado em Casos para indicar a classificação BI-RADS final da mamografia conforme análise das calcificações mamárias. A RNA PMC adotada foi testada usando o algoritmo Backpropagation cujos melhores resultados demonstraram uma convergência rápida da RNA e uma boa generalização do conhecimento, permitindo uma classificação com até 98% de acerto. A avaliação da técnica de RBC foi realizada com testes com a especialista da área e técnicas de validação preditiva. Os resultados obtidos validam que a presente proposta disponibiliza um processo consistente de análise das calcificações mamárias através do uso das técnicas de IA e PDI.

Palavras-chaves: Calcificações mamárias, *BI-RADS*, Rede Neural Artificial, Raciocínio Baseado em Casos, Processamento Digital de Imagens.

ABSTRACT

AID SYSTEM TO EVALUATION OF BREAST CALCIFICATION BY DIGITAL PROCESSING OF IMAGES AND ARTIFICIAL INTELLIGENCE

Author: Fátima Gisele Gomes Elpídio

Supervisor: Dra. Lourdes Mattos Brasil

Co-supervisor: Dra. Janice Magalhães Lamas

Post-Graduation Program in Biomedical Engineering

Brasília, August of 2012.

Early diagnosis still represents the best approach to prevent and control breast cancer, the second most frequent form of cancer worldwide. In this context, mamography has been largely used as major method for early disease detection. Calcifications (calcium build-ups) in breast regions are frequently found in breast screening examinations. The distinction between clusters of benign and malignant calcification is a complex task. The BI-RADS system (Breast Image Reporting and Data System) standardizes the description and management of mammographic findings among medical experts. This work is based on techniques Digital Image Processing (DIP) to perform the extraction of morphological information of calcifications and an Artificial Neural Network (ANN) Multi-Layer Perceptron (MLP) to classify calcifications analyzed according to the morphological classification defined by BI-RADS. The performance of MLP ANN was evaluated in a controlled test with synthetic calcifications, generated to simulate real calcifications and validated with real calcifications, extracted from mammograms provided by associated clinical. Additionally it was developed a system of Case Based Reasoning (CBR) to indicate the BI-RADS mammography of breast calcifications according to BI-RADS category most applicable. The ANN MPL adopted was tested using the Backpropagation algorithm whose best results demonstrated a rapid convergence of ANN and a good generalization of knowledge, allowing a rating up to 98% accuracy. The evaluation of CBR technique was performed with tests of medical expert and techniques for predictive validity. The results validate that this proposal provides a consistent process for analyzing breast calcifications through the use of AI techniques and DIP.

Key-words: Breast Calcification, BI-RADS, Artificial Neural Network, Case-Based Reasoning, Processing Digital Images.

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	Contextualização e Formulação do Problema	15
1.2	Objetivos	18
1.2.1	Objetivo geral	18
1.2.2	Objetivos específicos	18
1.3	Revisão da Literatura	18
1.4	Organização do Trabalho.....	22
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Análise da mama.....	23
2.2	Calcificações	25
2.3	Fatores pré-dispositivos do câncer de mama.....	36
2.4	A mamografia e a classificação BI-RADS dos achados	39
2.5	Processamento digital de imagens - PDI	40
2.5.1	Segmentação.....	43
2.5.2	Morfologia Matemática na extração de características	43
2.5.3	Transformada de Wavelet.....	45
2.6	O uso de RBC como Sistema de Apoio à Decisão	47
2.6.1	Etapas de desenvolvimento do sistema RBC	47
2.6.2	Vantagens e aplicações dos sistemas RBC	53
2.6.3	Validação do sistema RBC na área médica	54
2.7	Reconhecimento de padrões em imagens com RNA PMC.....	55
2.7.1	A aquisição do conhecimento das RNAs.....	58
2.7.2	Algoritmo de aprendizagem <i>Backpropagation</i>	59
3	METODOLOGIA	62
3.1	Definição da pesquisa	62
3.2	Delimitação do estudo	63
3.3	Coleta de dados para RNA.....	63
3.3.1	Formação e Aquisição de ROIS.....	63
3.3.2	Extração de atributos de ROIs (pré-processamento da RNA PMC)	65
3.3.3	Geração de base de treinamento e teste da RNA	66
3.3.4	Modelo de RNA proposto	66
3.4	Aquisição de conhecimento e modelagem do sistema RBC.....	67

3.4.1	Levantamento de requisitos.....	68
3.4.2	Definição do problema e representação dos casos com RBC	69
3.4.3	Atribuição dos pesos às características dos casos.....	71
3.4.4	Modelagem da solução proposta	72
4	RESULTADOS.....	75
4.1	Pré-processamento de ROIs	75
4.2	Análise de atributos morfológicos extraídos das ROIs.....	78
4.3	Treinamento de RNA PMC para classificar calcificações mamárias.....	80
4.4	Funcionamento do sistema RBC	84
5	DISCUSSÃO E CONCLUSÃO	90
6	TRABALHOS FUTUROS.....	92
	REFERÊNCIAS BIBLIOGRÁFICAS.....	95
	ANEXOS – MODELOS DE DOCUMENTOS.....	104
	ANEXO 1: Artigo publicado no anais do CBEB XXII	105
	ANEXO 2: Artigo publicado no anais do ECT2010.....	110
	ANEXO 3: Artigo publicado no anais do PAHCE-2011	114
	ANEXO 4: Artigo publicado no anais do ISEM-2011	116
	ANEXO 5: Artigo publicado no anais do WC2012	121
	ANEXO 6: Artigo publicado no anais do PAHCE-2012	125
	ANEXO 7: Artigo publicado no anais do CBEB2012	131
	ANEXO 8: Modelo de questionário para aquisição de conhecimento do RBC.....	138
	ANEXO 9: Modelo de proposição de heurísticas para aquisição e modelagem do conhecimento no sistema de RBC	143

LISTA DE TABELAS

Tabela 1 - Padrões de densidade mamária estabelecidos pelo BI-RADS (ACR, 2005).....	25
Tabela 2 – Comparação de caso de consulta com casos da base do SACM.....	86

LISTA DE FIGURAS

Figura 1: Anatomia da mama (NETTER, 2008).....	24
Figura 2: Classificação de calcificações mamárias em tipos morfológicos (LE GAL, 1984).....	26
Figura 3: Calcificações cutâneas (Imagem concedida pela CJLR).....	27
Figura 4: Calcificações vasculares (Imagem concedida pela CJLR).....	28
Figura 5: Calcificações em forma de pipoca (Imagem concedida pela CJLR).....	28
Figura 6: Calcificações em forma de bastonetes (Imagem concedida pela CJLR).....	29
Figura 7: Calcificações redondas (Imagem concedida pela CJLR).....	29
Figura 8: Calcificações em forma de ovo (Imagem concedida pela CJLR).....	30
Figura 9: Calcificações centro radiotransparente (Imagem concedida pela CJLR).....	30
Figura 10: Calcificações em leite de cálcio (Imagem concedida pela CJLR).....	31
Figura 11: Calcificação distrófica (Imagem concedida pela CJLR).....	31
Figura 12: Calcificações em fio de sutura (Imagem concedida pela CJLR).....	32
Figura 13: Calcificações puntiformes (Imagem concedida pela CJLR).....	33
Figura 14: Calcificações heterogêneas (Imagem concedida pela CJLR).....	33
Figura 15: Conjunto de calcificações ramificadas (Imagem concedida pela CJLR).....	34
Figura 16: Características morfológicas e de distribuição das calcificações mamárias. (CENTRO DE CIÊNCIAS TECNOLÓGICAS).....	35
Figura 17: Etapas do PDI (Adaptado de GOMES, 2001).....	41
Figura 18: Etapas do processamento frequencial de imagens (GONZALEZ E WOODS, 2008).....	43
Figura 19: Fundamentos de atuação dos operadores morfológicos (GONZALEZ E WOODS, 2008).....	45
Figura 20: Árvore de decomposição Wavelet, contendo 3 níveis. Em cada nível são utilizados filtros passa-baixas obtendo as Aproximações (gn) e filtros passa-altas obtendo os Detalhes (hn) (PEREIRA, 2003).....	46
Figura 21 Ciclo RBC (AAMODT e PLAZA, 1994).....	48
Figura 22: Exemplo de uma RNA PMC típica com uma camada oculta (AZEVEDO et al., 2000).....	56
Figura 23: Etapas aplicadas na metodologia.....	62
Figura 24: Diagrama de atividades para extração de informações morfológicas de ROIs.....	64
Figura 25: Arquitetura da RNA PMC adotada.....	67
Figura 26: Método VORD utilizado na etapa de levantamento de requisitos do RBC (SOMMERVILLE, 2003).....	68
Figura 27: Estrutura de representação dos casos.....	70
Figura 28: Estrutura de raciocínio empregada na pesquisa (Adaptado de SILVA, 2005).....	74
Figura 29: Procedimento para extração de ROIs (Adaptado de imagem original concedida pela CJLR).....	75
Figura 30: Exemplo de ROIs de calcificações reais adotadas no estudo em comparação com calcificações fictícias geradas para simular as reais (Adaptado de imagens originais concedidas pela CJLR).....	76
Figura 31: Exemplo de ROIs de calcificações extraídas de mamografias em paralelo com as calcificações fictícias geradas.....	77
Figura 32: Distribuição de classes por parâmetros extraídos das ROIs de calcificações sintéticas.....	79

Figura 33: Distribuição de classes por parâmetros extraídos das ROIs de calcificações reais.....	79
Figura 34: Distribuição de classes calcificação em função dos parâmetros área e perímetro.	80
Figura 35: Arquitetura da RNA PMC adotada.	81
Figura 36: Distribuição dos parâmetros extraídos de ROIs sintéticas utilizadas no treinamento da RNA.	81
Figura 37: Variação do EMQ da RNA PMC empregada para distinguir calcificações redondas de puntiformes.....	82
Figura 38: Disposição das classes no arquivo de treinamento utilizado como teste da RNA PMC treinada.....	83
Figura 39: Desempenho da RNA PMC testada durante a classificação de um arquivo teste com as ROIs reais.....	83
Figura 40: Diagrama Entidade Relacionamento da aplicação.	85
Figura 41: Tela de acesso a consulta de casos no SACM.	87
Figura 42: Tela de consulta de caso na base do sistema SACM.....	88
Figura 43: Tela apresentando sugestão de classificação BI-RADS conforme consulta por similaridade entre as características do caso de consulta e os casos presentes na base do SACM. ..	89
Figura 44: Teste para detectar automaticamente as regiões de calcificação em imagem de mamografia.....	93
Figura 45: Processo de segmentação de imagem por histograma.....	94

LISTA DE SÍMBOLOS, NOMENCLATURAS E ABREVIACÕES

ACM – *Association for Computing Machinery*

ACR – *American College of Radiology*

ANS - Agência Nacional de Saúde Suplementar

BI-RADS – *Breast Imaging Reporting and Data System*

CADe – *Computer-Aided Detection*

CADx – *Computer-Aided Diagnosis*

CC - Crânio Caudal

CJLR – Clínica Janice Lamas Radiologia

CBEB – Congresso Brasileiro de Engenharia Biomédica

EMQ – Erro Médio Quadrático

FDA – *Food and Drug Administration*

IBICT – Instituto Brasileiro de Informações em Ciência e Tecnologia

IEEE – *Institute of Electrical and Electronics Engineers*

INCA - Instituto Nacional de Câncer

ISEM - *International Symposium on Applied Electromagnetics and Mechanics*

LNCC/RJ - Laboratório Nacional de Computação Científica (unidade do Rio de Janeiro)

MLO - Médio Lateral Oblíqua

MRI - *Magnetic Resonance Imaging*

MVS - Máquina de Vetores de Suporte

NCBI – *National Center for Biotechnology Information*

NLM - *National Library of Medicine*

PAHCE - *Pan American Health Care Exchanges*

PDI - Processamento Digital de Imagens

PMC - *Perceptron* Múltiplas Camadas

RBC – Raciocínio Baseado em Casos

RBF - *Radial Basis Function*

RNA - Rede Neural Artificial

ROI - *Region of Interest*

SACM - Sistema de Avaliação de Calcificações Mamárias

TW - Transformada de *Wavelet*

TDF - Transformada Discreta de Fourier

UnB-GAMA – Universidade de Brasília – Faculdade Gama

USP – Universidade de São Paulo

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO E FORMULAÇÃO DO PROBLEMA

O câncer de mama é um tipo de neoplasia heterogênea, com variações histopatológicas, variável agressividade tumoral e potencial metastático, cuja ocorrência é mais freqüente na fase pré e pós menopáusicas femininas apresentando diferentes respostas clínicas (FREITAS *et al.*, 2006).

Considerado o segundo tipo de câncer com maior incidência mundial, o câncer de mama é uma enfermidade que atinge primordialmente a população feminina. Entretanto, pode afetar, embora em menor percentual, a população masculina, com resultados igualmente preocupantes (INCA, 2012). De acordo com estimativas pela Organização Mundial de Saúde (OMS), nas décadas de 60 e 70 houve um aumento de 10 vezes nas taxas de incidência de registros de câncer por faixa etária da base populacional de diversos continentes. Esse número tem aumentado nas últimas décadas. Estima-se que as incidências aumentem de 10 milhões para 15 milhões em 2020 (WHO, 2012). No Brasil, conforme registros do Instituto Nacional de Câncer (INCA), a quantidade de registros de mortes resultantes de câncer de mama continua elevada, principalmente porque o diagnóstico da doença é feito em estágio avançado.

A melhor forma de prevenção e controle do câncer de mama continua sendo o diagnóstico precoce (WHO, 2012). A mamografia é considerada o principal método de detecção precoce do câncer de mama (MAJID A. *et al.*, 2003; CHAKRABARTI *et al.*, 2006). A eficácia do exame foi comprovada por ensaios aleatórios controlados nos quais foi possível atingir uma redução da mortalidade absoluta, graças a capacidade da mamografia de achar carcinomas *ductais in situ* e cânceres infiltrantes de um tamanho menor e em estágio mais precoce do que em grupos de controle que não passaram pelo exame (MORRISON, 1992; ACR, 2005).

As calcificações (acúmulo de cálcio em regiões da mama) (NAPPI *et al.*, 1999) são freqüentemente encontradas em exames de rastreamento da mama, apesar da maioria caracterizar alterações benignas (LEUNG e SICKLES, 2007), existe uma grande parcela de carcinomas que iniciam-se com aglomerados de calcificações (LIBERTMAN *et al.*, 1998). Estima-se que 90% dos cânceres intraductais não palpáveis sejam detectados

morfologicamente pela presença de calcificações, as quais podem estar associadas a aproximadamente 40% de todos os tumores (LAMAS, 1998).

A distinção entre calcificações tipicamente benignas e calcificações suspeitas de malignidade é uma tarefa complexa (DINKEL, GASSEL e TSCHAMMLER 2000; PODHOLA e URMISKA, 2004), muitas vezes exigindo uma biópsia (retirada de um fragmento do tecido para análise microscópica) para uma conclusão efetiva. Segundo Kouskos *et al.* (2006) metade das biopsias são realizadas para investigar calcificações mamárias suspeitas. Apesar da alta probabilidade de detecção de câncer nas regiões com calcificações suspeitas de malignidade, as pacientes que são submetidas a biopsia sem necessidade enfrentam uma série de problemas (BIFFI, 2003) (psicológicos, financeiros, familiar, social, etc.) que poderiam ser evitados caso houvesse uma maior precisão na análise do exame antes da intervenção cirúrgica (MARTINS, BARRA e LUCENA, 2010).

O acompanhamento mamográfico em menor espaço de tempo (6 meses) das calcificações provavelmente benignas, seguido de controles mamográficos anuais, bilaterais, por pelo menos 2 anos é uma alternativa á biópsia (SICKLES, 1995; LAMAS e SICKLES, 2004). Todavia atualmente existe uma alta taxa de exames falso-positivos, com aumento do custo total do rastreamento em virtude da necessidade de investigações histopatológicas e, conseqüentemente, cirurgias desnecessárias, as quais poderiam ser minimizadas através de técnicas menos invasivas para ampliar a eficiência da análise precoce das calcificações antes da submissão do paciente a biópsias (MARTINS, BARRA e LUCENA, 2010).

O sistema de Laudos e Registros de Dados de Imagens da Mama – BI-RADS (em inglês, *Breast Imaging Reporting and Data System*) foi desenvolvido em 1993 para padronizar os relatórios mamográficos, visando diminuir divergências na interpretação subjetiva das imagens mamográficas, além de facilitar o controle dos resultados (ACR, 2005).

A detecção precoce e o diagnóstico correto do câncer mamário são um processo complexo, que depende dentre outros fatores, do raciocínio e experiência do profissional da área médica. Nesse contexto, o uso de técnicas computacionais no diagnóstico auxiliado por computador (do inglês CAD - *Computer-Aided Diagnosis*) na área médica é largamente difundido na literatura (TURGAY *et al.*, 2010; CHENG *et al.*, 2009). Por se tratar de uma ferramenta computacional os CADs não recebem a influência da fadiga e

pré-conceito, fatores intrínsecos ao ser humano que podem interferir no resultado (ZHENG *et al.*, 2004; MAJID *et al.*, 2003; PAQUERAULT *et al.*, 2009).

A Inteligência Artificial (IA) destaca-se como um paradigma computacional, vastamente defendido e empregado na literatura (PATROCINIO e SCHIABEL, 2001; SAMPAIO, 2009; JASMINE *et al.*, 2009; LASHKARI, 2010; etc.), em soluções computacionais onde a programação convencional não é suficiente (HAYKIN, 1999). Nesse âmbito, entre as técnicas computacionais utilizadas na retenção, recuperação e reutilização de conhecimento para auxílio à decisão destacam-se, as Redes Neurais e o Raciocínio Baseado em Casos (RBC). As Redes Neurais Artificiais (RNA) têm sido freqüentemente utilizadas em problemas de reconhecimento e classificação de padrões devido a sua capacidade de aprender por meio de exemplos (HAYKIN, 1999; BRAGA, CARVALHO e LUDERMIR; AZEVEDO, LIMÃO e BRASIL, 2000). Já o RBC é uma técnica útil na resolução de problemas baseado na absorção e utilização do conhecimento tácito, armazenado previamente (WANGENHEIM, 2003; SILVA, 2005).

A proposta dessa pesquisa consiste no emprego de técnicas de Processamento Digital de Imagens (PDI) (GONZALEZ e WOODS, 2008) para reconhecer, destacar e extrair informações de regiões de interesse (ROI - *Region of Interest*) de imagens mamográficas relacionadas a calcificações e classificá-las com a categoria BI-RADS mais aplicável utilizando um sistema RBC vinculado a uma RNA *Perceptron* Múltiplas Camadas (PMC) (HAYKIN, 1999; AZEVEDO, LIMÃO e BRASIL, 2000). Esse trabalho é fruto da parceria da UnB-GAMA com a Clínica Janice Lamas Radiologia (CJLR) e integra o projeto intitulado “Atlas Anatômico 3D Aplicado à Mama” desenvolvido com o Laboratório Nacional de Computação Científica do Ministério de Ciência e Tecnologia (LNCC/MCT) e Instituto Nacional de Ciência e Tecnologia em Medicina Assistida por Computação Científica (MACC), contribuindo no desenvolvimento de um ambiente educacional *web* para o aprendizado prático-morfológico das estruturas da mama.

1.2 OBJETIVOS

1.2.1 Objetivo geral

Este trabalho visa disponibilizar um sistema de auxílio ao diagnóstico por imagem para classificar calcificações mamárias, segundo categorias de risco da classificação BI-RADS.

1.2.2 Objetivos específicos

Especificamente são objetivos da pesquisa:

- Implementar, treinar e avaliar uma RNA para classificação de calcificações mamárias em concordância com a nomenclatura e os critérios adotados pelo BI-RADS;
- Desenvolver um sistema de um RBC para auxiliar na classificação BI-RADS final de mamografias, considerando as características morfológicas das calcificações detectadas na mamografia, segundo os critérios adotados pela classificação BI-RADS.

1.3 REVISÃO DA LITERATURA

A pesquisa bibliográfica realizada neste trabalho adotou livros, monografias, teses, e artigos de fontes especializadas tais como: *Institute of Electrical and Electronics Engineers* (IEEE), *PubMed-ACM (Association for Computing Machinery)*, *National Center for Biotechnology Information* (NCBI), Universidade de São Paulo (USP), e Instituto Brasileiro de Informações em Ciência e Tecnologia (IBICT).

O IEEE, uma organização profissional sem fins lucrativos, mantém uma série de materiais com o intuito de difundir conhecimento no ramo da engenharia, elétrica, eletrônica, computação e afins. Na base do IEEE foram localizados 10 resultados para a pesquisa por trabalhos relacionados ao tema: “*Digital Image Processing and Artificial Neural Networks for detection Calcifications in Mammography*”. Um dos trabalhos de maior correlação com o tema foi (PATROCINIO e SCHIABEL, 2001) o qual aborda o uso de uma RNA supervisionada com algoritmo de aprendizagem “*Backpropagation*” (HAYKIN, 1999; AZEVEDO, LIMÃO e BRASIL, 2000) para classificar agrupamentos de calcificações e nódulos mamários em “suspeitos” e “não-suspeitos”. A pesquisa utiliza 112

ROIs previamente segmentadas e recortadas de mamogramas digitalizados. Para treinar a RNA na classificação foram empregadas informações extraídas da ROI tais como: momentos de primeira, segunda e terceira ordens (MASTERS, 1994), invariantes momentos de Hu (HU, 1962), compacidade, irregularidade, raio de giração¹, perímetro² e área, além de características do histórico da paciente (idade e densidade da mama). No estudo realizado por (REZAI-RAD et al., 2005; JASMINE et al., 2009) foi constatado que as microcalcificações estavam presentes nas componentes das altas frequências da imagem. Partindo dessa premissa, realizou-se a decomposição da imagem mamográfica em diferentes bandas de frequência, suprimindo as baixas frequências. A imagem final é reconstruída empregando-se apenas as altas frequências, nas quais as calcificações estão presentes. A detecção das calcificações é realizada através da associação de RNA e a transformada de Wavelet. De forma similar (LASHKARI, 2010), aborda o uso de uma RNA associada a Gabor wavelets³ (SHEN e BAI, 2006), momento geométrico invariante, energia, entropia, contraste e outras características estatísticas (média, mediana, variância, correlação e valores máximos e mínimo de intensidade) para analisar e classificar anormalidades do tecido mamário.

O PubMed, criado pelo NCBI e mantido pela *National Library of Medicine* (NLM), disponibiliza material publicado em revistas de grande circulação da área médica. Foram localizados 9 registros para a busca por “*Digital Image Processing and Artificial Neural Networks for detection Calcifications in Mammography*” e 20 alterando o argumento de pesquisa para “*Digital Image Processing and Artificial Neural Networks in Mammography*”. Na pesquisa merecem destaque os trabalhos (HALKIOTIS, TAXIARCHIS e RANGOUSI, 2007; RANGAYYAN, BANIK, e DESAUTELS, 2010; NUNES, SCHIABEL e GOES, 2006). O estudo de (HALKIOTIS, TAXIARCHIS e RANGOUSI, 2007) emprega morfologia matemática para detectar e evidenciar aglomerados de calcificações mamárias e duas RNA PMC associadas a duas RNAs de funções de base radiais (do inglês RBF - *Radial Basis Function*) (HAYKIN, 1999) como

¹ **Raio de giração** – representa a distribuição espacial de pixels significantes da imagem (MASTERS, 1994).

² **Perímetro** – é medido pelo somatório das distâncias entre pontos consecutivos (CHENG, H. D. et al., 2009).

³ **Transformada de Gabor** - similar a transformada de Fourier, a transformada de Wavelet analisa variações espectrais em resoluções tempo-frequência diversas. A transformada de Gabor difere, introduzindo um parâmetro que é inversamente proporcional à frequência fazendo que a distribuição de frequência situe-se em um intervalo positivo (COX, 2004).

classificador de ROIs. Os trabalhos de (TURGAY, 2010; HALKIOTIS, TAXIARCHIS e RANGOUSI, 2007; RANGAYYAN, BANIK, e DESAUTELS, 2010; NUNES, SCHIABEL e GOES, 2006) versam sobre o uso de CADs no apoio à decisão médica sobre ROIs. (TURGAY, 2010) realiza uma análise dos modelos de CADs no apoio ao diagnóstico médico em mamografias, ultrassonografia⁴ e interpretação *Magnetic Resonance Imaging* (MRI). O estudo sumariza os tipos de modelos computacionais, tamanho dos dados de entrada, características dos métodos de seleção, características dos tipos de entrada, padrão de referência e medição de desempenho para cada estudo. Os trabalhos desenvolvidos por (HALKIOTIS, TAXIARCHIS e RANGOUSI, 2007; RANGAYYAN, BANIK, e DESAUTELS, 2010) são abordagens práticas, onde no primeiro caso é desenvolvido um CAD para detecção e classificação de imagens resultantes de ultra-sonografia da mama e o segundo propõe um CAD para analisar distorções arquiteturais⁵. Em (NUNES, SCHIABEL e GOES, 2006) foram combinadas três técnicas (aumento de contraste através da transformação dos níveis de cinza do histograma, realce de contraste pela atenuação de coeficientes e realce por disposição das características da curva do filme radiográfico) para detectar aglomerados de microcalcificações em mamografias de mamas densas, com pouco contraste.

Nas bases bibliográficas nacionais consultadas foram localizados 30 resultados na USP para o tema “Calcificações mamárias via processamento digital de imagens e redes neurais artificiais” e no IBICT 2 trabalhos relacionados ao tema mamografia e 99 sobre Processamento Digital de Imagens e RNA. A maioria dos trabalhos localizados na base do IBICT é direcionada a área da saúde ou não apresentam grande relevância com o tema deste documento, sendo Sampaio (2009) o trabalho mais próximo do tema abordado pela presente pesquisa, o qual utiliza RNA Celulares (do inglês *Cellular Neural Network* - CNN) para segmentar e extrair informações de ROIs, em seguida emprega descritores de geometria (excentricidade, circularidade, densidade, desproporção circular e densidade circular) para analisar a textura da ROI e Máquina de Vetores de Suporte para classificá-

⁴ **Ultra-sonografia mamária** - é um método auxiliar da mamografia, com maior aplicação na distinção de lesão cística de sólida e no estudo de mama radiologicamente densa (TURGAY, 2010).

⁵ **Distorção arquitetural** – estrutura que apresenta linhas finas ou espiculadas irradiando-se de um ponto e retração ou distorção focal da margem do parênquima. Pode estar relacionada a nódulo, assimetria ou calcificações (ACR, 2005).

las. As RNAs CNN, propostas por Chua e Yang (1988), são inspiradas nos autômatos⁶ e são caracterizadas pelas interconexões entre os seus neurônios limitarem-se a uma determinada vizinhança entre estes. A Máquina de Vetores de Suporte (MVS) corresponde a um sistema de aprendizado supervisionado cujo objetivo é classificar os dados de entrada em duas classes distintas. Segundo Sampaio (2009) a MVS atua como um hiperplano caracterizando uma superfície de decisão para permitir uma máxima margem de separação entre classes.

Na base da USP foi possível localizar bastante material afim ao objeto de estudo dessa pesquisa, dentre os quais merecem destaque (SANTOS, 2002; ANGELO, 2007; SILVA JÚNIOR, 2009 e GUIMARÃES, 2009). Trata-se de uma série de pesquisas envolvendo o desenvolvimento de CADs para auxiliar no diagnóstico médico de calcificações e nódulos mamários. SANTOS (2002) desenvolveu um esquema de processamento para detecção de nódulos em mamas densas de imagens mamográficas digitalizadas, baseado na transformada de Watershed. A transformada de Watershed, desenvolvida por Digabel e Lantuéjoul (1978), pode ser abstratamente definida como uma técnica de segmentação de imagens que utiliza um gradiente com níveis de cinza correspondentes a superfície com bacias de captação. As linhas de segmentação buscam detectar as linhas de divisão de águas que ficam no encontro das bacias de captação, onde uma gota d'água pode escorrer para dois mínimos regionais diferentes.

O trabalho de (ANGELO, 2007) é uma continuidade do trabalho descrito anteriormente, além de prover a disponibilização do sistema *online*. Na seqüência, (SILVA JÚNIOR, 2009) tem como objetivo principal a detecção e classificação de cluster de microcalcificações para processar integralmente imagens mamografias. E na última pesquisa da série (GUIMARÃES, 2009) é realizado um estudo de mamas densas, as quais geralmente apresentam baixo contraste em virtude do tecido fibroglandular de alta densidade. Diante desse contexto, foi proposto a subtração digital de imagens obtidas em diferentes níveis de energia do feixe de raios X, com o intuito de detectar lesões malignas, as quais em geral são camufladas, quando superpostas pelo tecido de absorção semelhante, no método tradicional.

⁶ **Autômatos Celulares** - segundo Sampaio (2009) são sistemas dinâmicos discretos no tempo e espaço, localmente acoplados, cujos estados também são discretos.

Os trabalhos relacionados nesse item denotam a relevância do estudo de calcificações mamárias e a necessidade de detectá-las previamente. Nesse intuito diversas técnicas de PDI e IA são empregadas, entretanto a presente proposta consiste em uma iniciativa pioneira na adoção de RNA em conjunto com RBC para prover a análise e classificação BI-RADS de calcificações mamárias.

1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em seis capítulos, incluindo este capítulo.

No capítulo dois, é apresentada uma visão geral do referencial teórico, objetivando a compreensão das tecnologias, conceitos, e padrões utilizados no escopo dessa pesquisa.

O capítulo três detalha a metodologia utilizada no estudo.

O capítulo quatro descreve os resultados obtidos e a estrutura do processo de trabalho proposto para disponibilizar o sistema de avaliação de calcificações mamárias.

O capítulo cinco discute os pontos de maior importância envolvendo o tema deste estudo e apresenta as conclusões finais do trabalho.

Por fim, o capítulo seis apresenta os trabalhos futuros que podem ser derivados das idéias apresentadas nessa pesquisa.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 ANÁLISE DA MAMA

A análise e a classificação de sutis desvios da normalidade do tecido mamário requerem conhecimento da estrutura anatômica da glândula mamária, bem como a sua correlação com métodos de imagem (KIRBY, BLAND e COPELAND, 1994). A existência de pessoas com mamas estruturalmente idênticas é praticamente nulo, visto que a mama sofre alterações com o decorrer do tempo (dentre outros fatores: modificações fisiológicas naturais de acordo com a idade, durante ou após uma gestação, com uso de terapia de reposição hormonal, alteração de peso, etc.) (BRITO, 2008).

A glândula mamária tem uma origem dérmica e localiza-se inserida na pele através de uma fáscia, a qual possui uma estrutura semelhante a favo de mel, contendo compartimentos volumosos e desiguais denominados lóculos. Os lóculos envolvem-se separadamente e em conjunto, originando os lobos mamários. Os lobos são compostos por ductos, lóbulos e tecido conectivo. O seio lactífero, localizado aproximadamente na região retroareolar, é constituído pelos ductos galactóforos com dilatação segmentar. O último segmento ductal é o ducto terminal extra-lobular, o qual é denominado Unidade Ductolobular Terminal (UDLT) ao penetrar no lóbulo (LAMI, 1991). A Figura 1 apresenta as principais estruturas que compõem a glândula mamária.

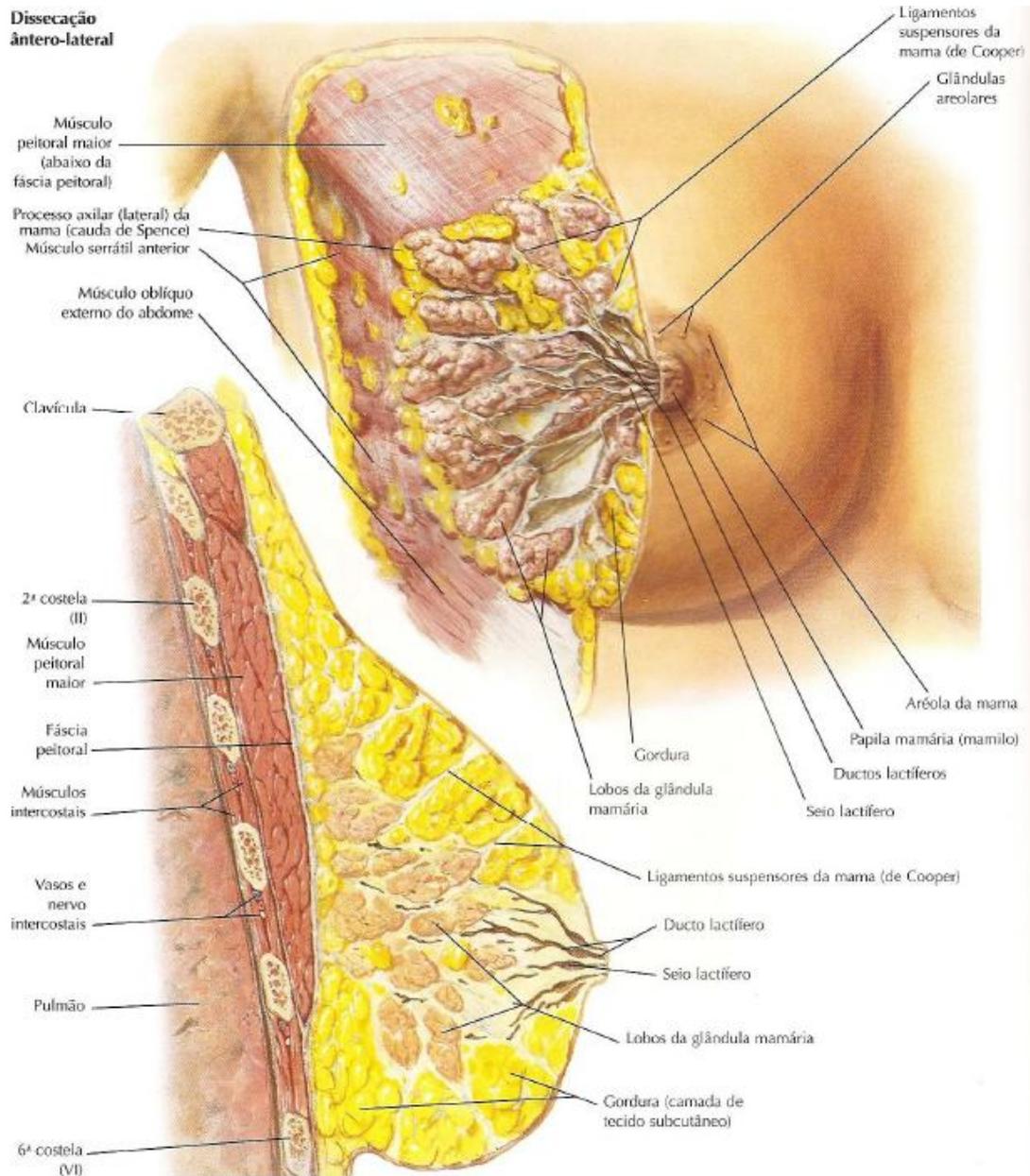


Figura 1: Anatomia da mama (NETTER, 2008).

O parênquima mamário apresenta variações naturais de densidade. A Tabela 1 apresenta o padrão de densidade mamário instituído pelo BI-RADS (ACR, 2005).

Tabela 1 - Padrões de densidade mamária estabelecidos pelo BI-RADS (ACR, 2005).

Densidade mamográfica	Descrição
Mama liposubstituída	Tecido glandular < 25%
Mama fibroglandular	Tecido glandular entre 25% e 50%
Mama heterogeneamente densa	Tecido glandular aproximadamente entre 51% e 75%
Mama acentuadamente densa	Tecido glandular > 75%

Diversos estudos (WOLF, 1976; BOYD, 1998; ACR, 2005; MACHETTI e MARANA, 2007) relatam a importância da análise da densidade mamária, por ser um fator que pode dificultar a detecção do câncer de mama e ampliar as chances de desenvolvimento da doença, por essa razão a sensibilidade de detecção tumoral da mamografia varia conforme o tipo de densidade mamária. Em geral foi observado pelos pesquisadores que as mamas acentuadamente densas têm um alto risco de ocorrência de câncer de mama em comparação com mamas de baixa densidade.

2.2 CALCIFICAÇÕES

Um achado muito frequente na análise de imagens mamográficas, as calcificações podem ser caracterizadas como pequenos depósitos radiopacos com alta sensibilidade aos raios X (LEMACKS et al, 2002; MOURÃO, 2009). Geralmente, elas são associadas a processos benignos, embora caracterizem cerca de 30 a 50% das lesões não detectáveis clinicamente (HIGHNAM e BRADY, 1999). As calcificações são estruturas resultantes do depósito de secreções granulosas ricas em cálcio que se solidificam com a especialização do tecido mamário durante a produção de leite (fonte de cálcio). As calcificações suspeitas de malignidade geralmente surgem em aglomerados (CHENG *et al.*, 2003). A importância da análise e diagnóstico prévio das calcificações suspeitas tem sido relatada na literatura desde a sua descrição inicial, por Leborgne em 1951 até os tempos atuais (QUEIROZ e HUMMEL, 1995; FEIG, 1995; SICKLES, 1995; ACR, 2005; MARTINS, BARRA e LUCENA, 2010).

Le Gal (1984) foi um dos pioneiros na análise morfológica das calcificações mamográficas, ao publicar uma proposta de diferenciação das calcificações em 5 tipos, baseado na histopatologia da Região de Interesse. A Figura 2 apresenta os 5 tipos de classificação adotados por *Le Gal*.

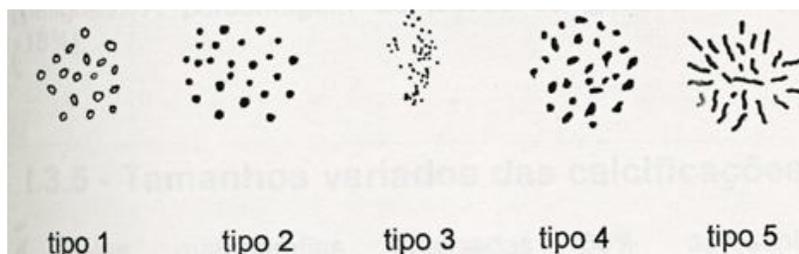


Figura 2: Classificação de calcificações mamárias em tipos morfológicos (*LE GAL*, 1984).

Lanyi (1986) foi outro autor que se destacou no estudo histopatológico e classificação das calcificações mamárias. Entretanto a nomenclatura e critérios mais empregados na atualidade são os do BI-RADS (2005), proposto inicialmente em 1993, o qual adotou outros critérios, além da morfologia, para discriminar as calcificações mamárias, tais como a distribuição, a localização, a estabilidade e os achados associados.

Atualmente sabe-se que o nível de suspeita de malignidade das calcificações mamárias pode aumentar ou diminuir de acordo com as características delas tais como o tipo, tamanho, a quantidade, a densidade e a distribuição (*KETTRITZ, MORACK e DECKER, 2005*). Geralmente são consideradas calcificações suspeitas de malignidade as que são agrupadas e apresentam aspecto irregular ou linear (*LE GAL, 1984; KETTRITZ, MORACK e DECKER, 2005*). Sendo que a maioria dos autores considera que os carcinomas são associados aos agrupamentos que possuem mais de 9 calcificações e que ocupam uma faixa maior que 3mm da mama (*DINKEL, GASSEL e TSCHAMMLER 2000; HOLLAND e HENDRIKS, 1994; STOMPER et. al., 2003*).

Grupos de Classificações

As calcificações são classificadas em dois grandes grupos:

- I. **Calcificações associadas a processos benignos** – geralmente são maiores que as malignas, muitas vezes apresentam forma grosseira e são facilmente detectáveis na

mamografia (KOPANS, 2000; ACR, 2005). De acordo com ACR (2005), elas podem ser classificadas como:

- *Cutâneas* – apresentam uma configuração geométrica com centro radiotransparente. Geralmente são detectadas na região parasternal, próximo a prega inframamária, axila, auréola, ou em menor escala através de incidências mamográficas tangenciais a pele que reveste a mama. A Figura 3 exemplifica esse tipo de calcificação.

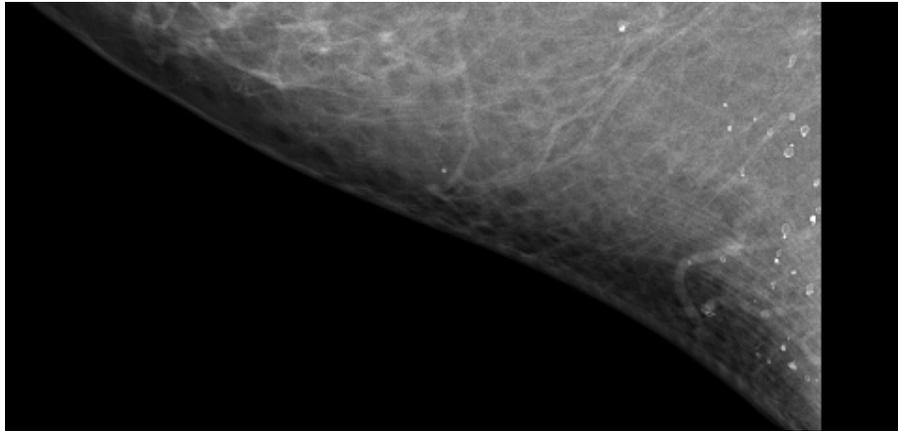


Figura 3: Calcificações cutâneas (Imagem concedida pela CJLR).

- *Vasculares* – caracterizadas por trilhas paralelas ou calcificações lineares, claramente associadas com estruturas tubulares. São localizadas ao longo das paredes dos vasos em aspecto serpentiforme. A Figura 4 exemplifica esse tipo de calcificação.

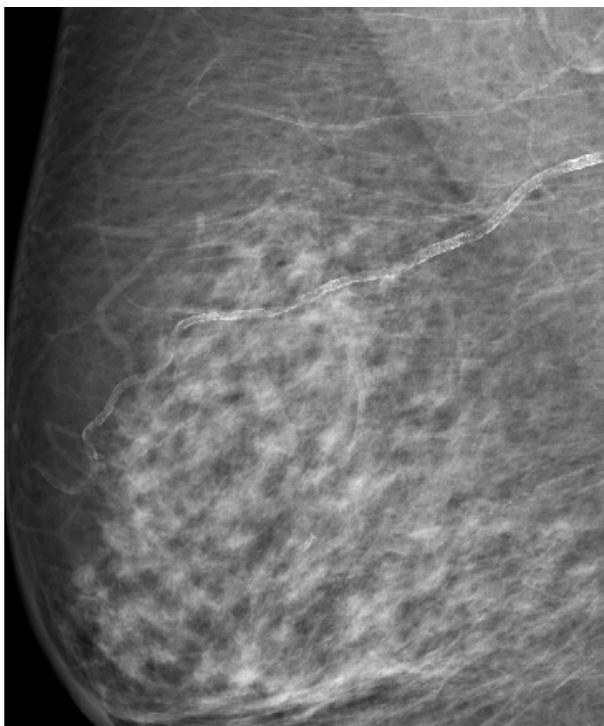


Figura 4: Calcificações vasculares (Imagem concedida pela CJLR).

- Grosseiras ou semelhantes a “pipoca” – são calcificações grandes, com diâmetro superior a 2-3mm, típicas de fibroadenomas em involução. A Figura 5 exemplifica esse tipo de calcificação.

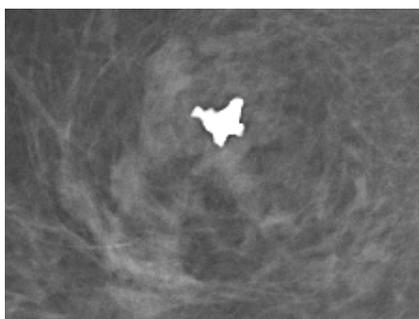


Figura 5: Calcificações em forma de pipoca (Imagem concedida pela CJLR).

- “*Em forma de bastonetes*” – são resultantes de patologia secretória ductal e podem formar bastonetes lineares, descontínuos ou sólidos frequentemente com diâmetro igual ou superior a 1mm. Geralmente são bilaterais e possuem uma distribuição

ductal que irradia-se em direção a papila com possíveis ramificações. A Figura 6 exemplifica esse tipo de calcificação.

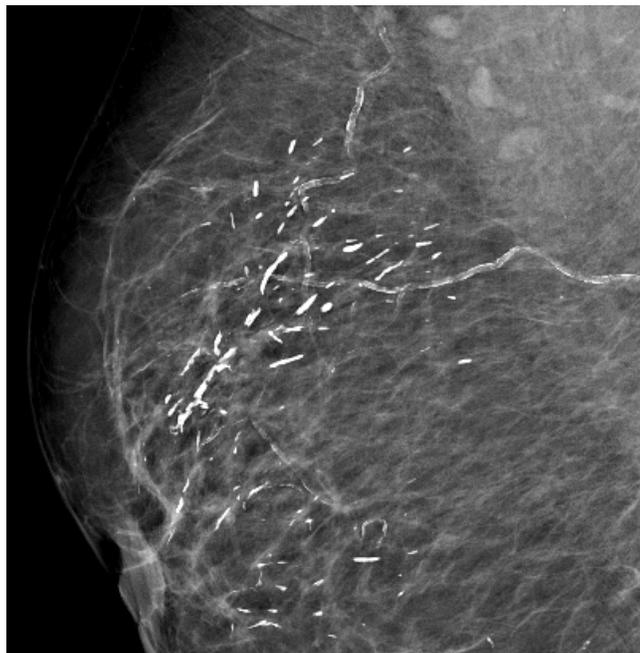


Figura 6: Calcificações em forma de bastonetes (Imagem concedida pela CJLR).

- *Redondas* – apresentam formato esférico, podendo variar em tamanho quando são múltiplas, porém com um tamanho geralmente igual ou superior a 0,5mm de diâmetro. A Figura 7 exemplifica esse tipo de calcificação.



Figura 7: Calcificações redondas (Imagem concedida pela CJLR).

- “*Casca de ovo*” ou “em anel” – geralmente são extremamente finas e curvilíneas as quais são visualizadas como depósito de cálcio sobre a superfície de uma esfera. As calcificações mais comuns desse tipo são necrose gordurosa e calcificações nas paredes de cisto. A Figura 8 exemplifica esse tipo de calcificação.

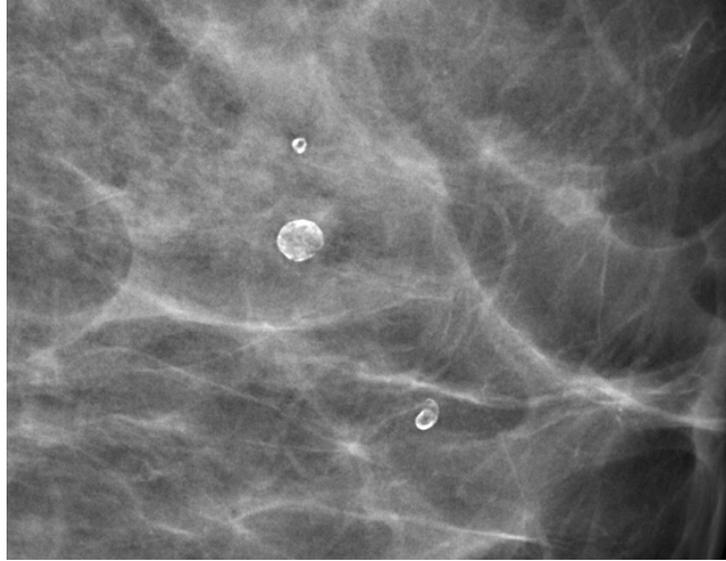


Figura 8: Calcificações em forma de ovo (Imagem concedida pela CJLR).

- *Centro radiotransparente* – caracterizam-se como depósitos redondos ou ovóides, com superfícies lisas e têm centro radiotransparente. A "parede" formada é mais espessa que no tipo de calcificação em "anel" ou "casca de ovo". Possuem um tamanho bem variável, assumindo valores entre menor que 1mm ou maiores que 1cm. São associadas a áreas de necrose gordurosa e resíduos calcificados em ductos. A Figura 9 exemplifica esse tipo de calcificação.

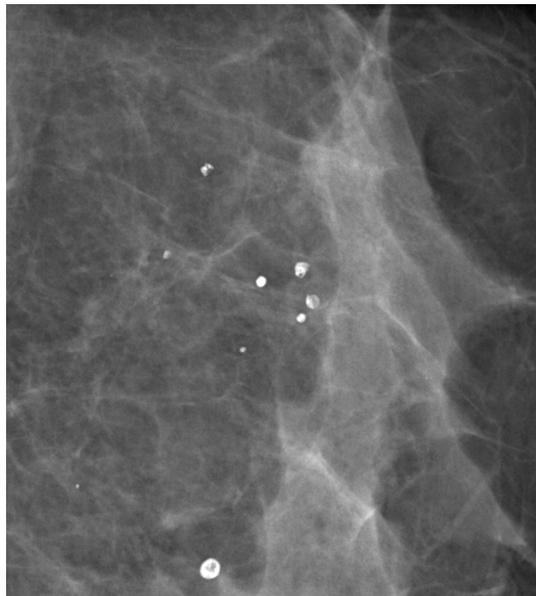


Figura 9: Calcificações centro radiotransparente (Imagem concedida pela CJLR).

- “*Em leite de cálcio*” – geralmente é amorfa ou esférica na incidência CC (Crânio Caudal) e curvilínea ou linear na MLO (Médio Lateral Oblíqua). A Figura 10 exemplifica esse tipo de calcificação.

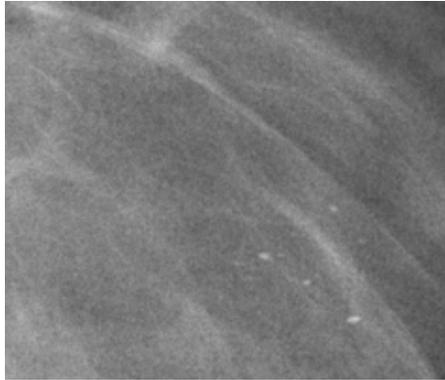


Figura 10: Calcificações em leite de cálcio (Imagem concedida pela CJLR).

- *Distróficas* – geralmente apresentam um tamanho superior a 0,5mm de diâmetro, são irregulares e podem apresentar centro radiotransparente. A Figura 11 exemplifica esse tipo de calcificação.

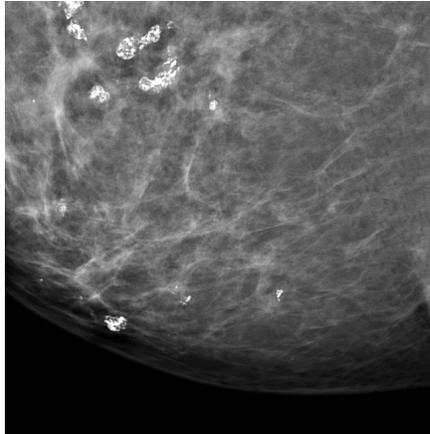


Figura 11: Calcificação distrófica (Imagem concedida pela CJLR).

- *Fios de sutura* – são filiformes, vermiculares, consistindo em depósitos de cálcio em fios de sutura que geralmente surgem após irradiação. A Figura 12 exemplifica esse tipo de calcificação.

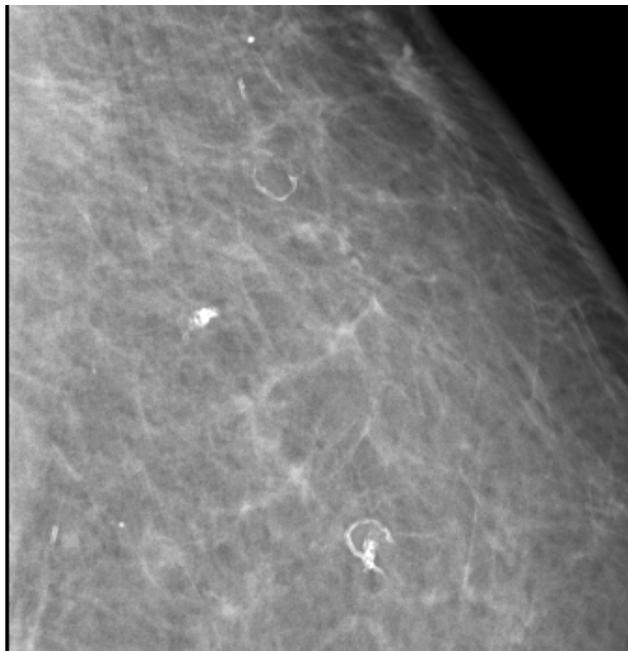


Figura 12: Calcificações em fio de sutura (Imagem concedida pela CJLR).

II. **Calcificações associadas ao câncer de mama ou suspeitas de malignidade** – são predominantemente pequenas (geralmente menores que 0,5mm), encontram-se agrupadas com mais de 5 elementos por cm³, sendo variáveis em morfologia (KOPANS, 2000; MARTINS, BARRA e LUCENA, 2010). Quanto a classificação são subclassificadas como:

- *Puntiformes* – são calcificações arredondadas muito pequenas geralmente menores que 0.5mm. A Figura 13 exemplifica esse tipo de calcificação.

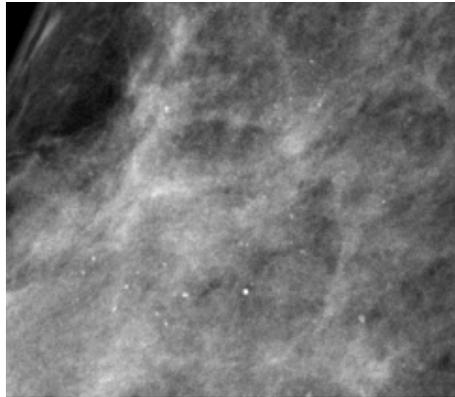


Figura 13: Calcificações puntiformes (Imagem concedida pela CJLR).

- *Heterogêneas* – apresentam tamanho e forma variável e podem caracterizar calcificações benignas e malignas. Exemplos desse tipo de calcificação podem ser vistos na Figura 14.

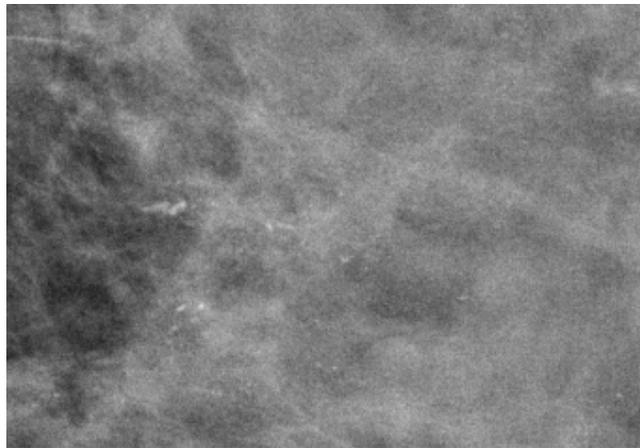


Figura 14: Calcificações heterogêneas (Imagem concedida pela CJLR).

- *Calcificações ramificadas* – apresentam um tamanho inferior a 0,1mm, são delgadas e irregulares. A Figura 15 apresenta um conjunto de calcificações com características típicas desse tipo de calcificação.

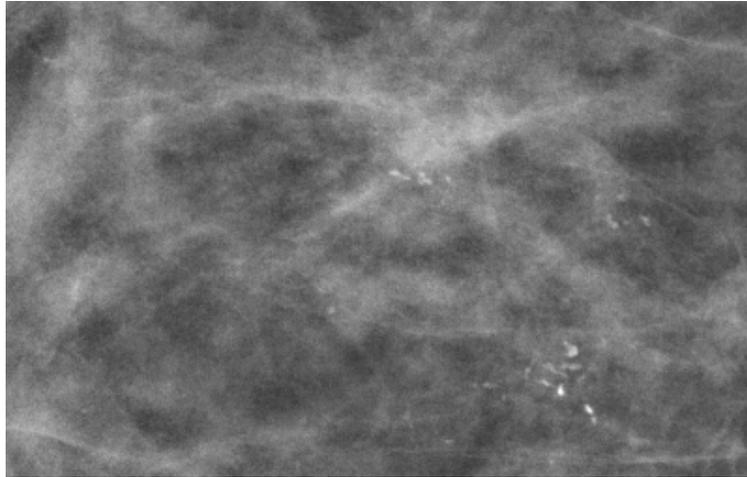


Figura 15: Conjunto de calcificações ramificadas (Imagem concedida pela CJLR).

Na literatura (KOPANS, 2000; DANTAS *et al.*, 2002; ACR, 2005), a análise e a classificação das calcificações geralmente é determinada pelas seguintes características:

- *A forma* – geralmente as patologias neoplásicas malignas apresentam forma irregular.
- *A quantidade* – o risco de malignidade está associada a quantidade de calcificações por cm^3 da mama. Muitas pesquisas (KOPANS, 2000; SICKLES, 1995; DANTAS *et al.*, 2002; ACR, 2005; MARTINS, BARRA e LUCENA, 2010) afirmam que há um aumento significativo do risco de malignidade se houverem cinco ou mais calcificações agrupadas por cm^3 .
- *O tamanho* – as calcificações benignas geralmente são grandes e grosseiras, enquanto que as malignas são frequentemente microscópicas.
- *A estabilidade* – esse fator está associado a probabilidade das calcificações não ampliarem a quantidade ou a distribuição e forma com o decorrer do tempo.

Quanto a distribuição, de acordo com ACR (2005) e Dantas *et al.* (2002), as calcificações podem ser classificadas em:

- *Agrupadas ou aglomeradas* – geralmente estão associadas a malignidade, embora possam ocorrer em processos benignos. Caracteriza um conjunto de calcificações que ocupam um volume de até 2 cm de diâmetro.

- *Lineares* – são calcificações que em virtude da orientação ductal apresentam-se linearmente distribuídas. Podem ocupar um pequeno ou grande espaço de extensão da mama e podem apresentar ramificações.
- *Segmentares* – geralmente ocupam uma área superior a 2 cm da mama. Podem ocorrer em processos benignos e malignos.
- *Difusas* – encontram-se distribuídas aleatoriamente por toda a mama.
- *Regionais* – geralmente ocupam um grande volume de tecido, sem configurar uma distribuição ductal e não são associadas a processos malignos.
- *Grupos Múltiplos* – relacionadas a mais de um grupo de calcificações semelhantes na morfologia e distribuição.

A Figura 16 ilustra um conjunto de calcificações de acordo com as suas características morfológicas e seus modificadores de distribuição.

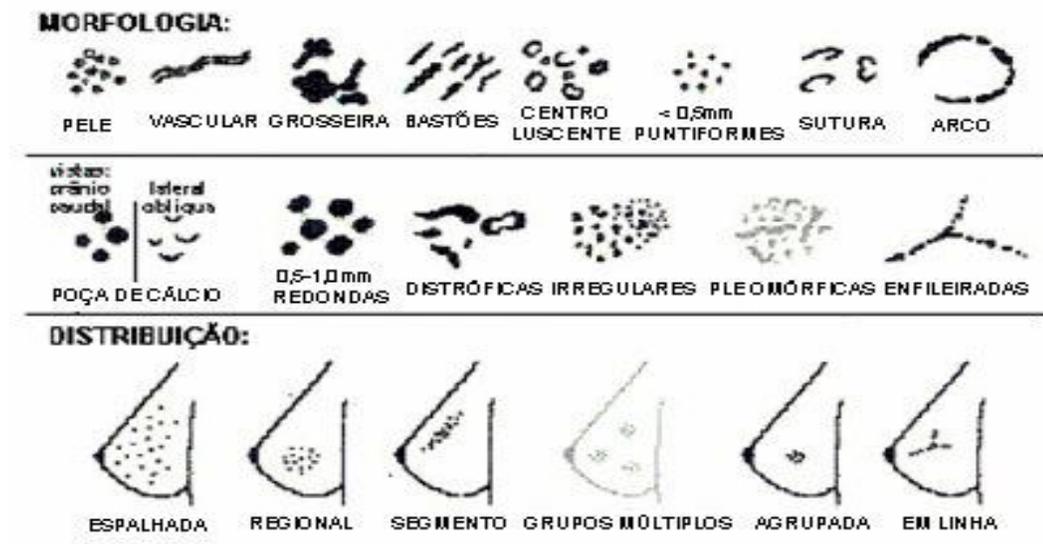


Figura 16: Características morfológicas e de distribuição das calcificações mamárias. (CENTRO DE CIÊNCIAS TECNOLÓGICAS).

2.3 FATORES PRÉ-DISPOSITIVOS DO CÂNCER DE MAMA

O câncer de mama é uma doença multifatorial advinda de um processo de múltiplas fases sob a influência de fatores adversos em cada fase. Trata-se de uma neoplasia heterogênea, com variações histopatológicas, rara em homens e em pessoas jovens que nas últimas décadas tem sido registrada como uma ocorrência ascendente em mulheres, conforme o avanço da idade feminina (KIRBY, BLAND e EDWARD, 1994). De acordo com Lamas (1998), muitos são os fatores relacionados ao desenvolvimento do câncer de mama, logo é praticamente impossível erradicar ou estimar com acurácia o surgimento da doença. O diagnóstico precoce é a melhor forma de mudar o curso natural da doença (LAMAS, 1998).

Nesse item foi realizada uma investigação dos fatores que são mais relacionados nas pesquisas sobre câncer de mama como fatores de predisposição da doença.

Histórico familiar

O histórico familiar representa um dos fatores mais relevantes para ocorrência do câncer de mama. O risco de ocorrência do câncer de mama cresce proporcionalmente ao número de parentes afetados e grau de parentesco com estes. Estudos (OTTMAN *et al.*, 1983; GUIMARÃES, 2004) revelam que parentes de pacientes com câncer mamário pré-menopáusico bilateral ocupam risco máximo de desenvolvimento da doença. Foi observado que esse risco é menor para as mulheres que são parentes de pacientes portadoras de câncer unilateral que desenvolveram a doença após a menopausa.

Etnia

A raça pode influenciar vários índices de risco tais como a menarca precoce, menopausa tardia, idade avançada da primeira gestação, dentre outros. Analisando a questão racial, vários autores (WATERHOUSE *et al.*, 1976; GRAY, PIKE e HENDERSON, 1980) sugerem que apesar das mulheres brancas representarem um maior número de casos global do câncer de mama em comparação as mulheres não-brancas, essa diferença só torna-se mais acentuada após os 40 anos (idade correspondente ao início da menopausa para a maioria das mulheres).

Doenças benignas da mama

As doenças benignas da mama, como a doença fibrocística, podem dobrar o risco de câncer de mama, segundo as pesquisas de Dupont e Page (1985) o risco apresenta uma relação maior com lesões com hiperplasia epitelial e evidência histológica de calcificação.

Evolução do ciclo reprodutivo feminino

Analisando-se o ciclo evolutivo feminino, observou que a menarca precoce (12 anos) aliada ao estabelecimento precoce de ciclos regulares (dentro de 1 ano após a menarca) pode triplicar a probabilidade de ocorrência de câncer de mama (HENDERSON, PIKE e CASAGRANDE, 1981). Por outro lado, as pesquisas revelam que o primeiro parto precoce e a alta paridade oferecem um efeito protetor contra o câncer de mama (LIPNICK *et al.*, 1984). Na menopausa foi observado que a tendência a câncer diminui, sendo duplamente reduzida se a menopausa ocorrer antes da idade de 45 anos. Foi observado ainda que indução artificial da menopausa pode oferecer um efeito protetor similar ao da menopausa natural.

Radiação

O efeito da radiação sobre a mama pôde ser vastamente investigada principalmente com os dados de acompanhamento das sobreviventes as explosões atômicas no Japão, bem como de mulheres que foram expostas a altas doses medicamentosas de radiação ionizante no tratamento de mastite e tuberculose. Os riscos de desenvolvimento de câncer de mama é inversamente proporcional à idade em que a mulher foi exposta a radiação e diretamente proporcional a dosagem de radiação. Tais constatações foram capazes de ocasionar dúvidas sobre a viabilidade do exame mamográfico. As pesquisas (FORREST *et al.*, 1987) mostram que os benefícios potenciais da técnica são superiores aos riscos ínfimos do exame, diante da estimativa de apenas 1 mulher com meia-idade em 200.000 desenvolver a doença após a realização de um único exame mamográfico.

Obesidade

Embora ainda não exista um consenso sobre os efeitos do peso e massa corpórea sobre o risco de desenvolvimento de câncer de mama, há estudos (WYNDER, ROSE e COHEN, 1986; ROSE, BOYAR e WYNDER, 1986) que apontam um dobro de chances de desenvolvimento da doença em mulheres obesas pós-menopáusicas. Essa relação foi

potencialmente observada, pela progressão dos registros da doença em paralelo ao aumento de peso em relação a altura das mulheres japonesas.

Terapia hormonal

A associação do uso de anticoncepcionais orais ao aumento de chances de câncer de mama ainda é uma questão controversa, há relatos na literatura que tais drogas podem proteger o organismo feminino ou que podem ampliar a predisposição ao câncer de mama. Dentre os fatores que contribuem para manter a indefinição destas afirmações constam: o fato que uso em larga escala de anticoncepcionais orais por grande parcela de mulheres jovens ainda pode ser considerado uma prática relativamente recente, a existência de uma grande lacuna de tempo latente entre a adoção da droga e a manifestação do câncer e as alterações do percentual de estrogênio e progesterona na composição dos anticoncepcionais orais que foram realizadas no decorrer dos últimos 20 anos. Diante de tais fatos o que geralmente é afirmado por muitos pesquisadores (PARKIN, 2001; KAHLENBORN, C. *et al.*, 2006; NAROD e FOULKES, 2004) é que as chances de desenvolvimento da doença podem ser reforçadas com o uso de anticoncepcionais orais combinados quando o uso deles é iniciado em uma idade muito precoce, ou antes, da primeira gestação, prolongando-se o tratamento durante uma longa fase da vida. A mesma pesquisa considerando mulheres que adotam a terapia de reposição hormonal sugere um leve aumento nas chances de desenvolvimento de câncer de mama para as mulheres peri ou pós-menopáusicas, principalmente nos casos em que elas são afetadas por doenças benignas da mama.

Influência do álcool e tabagismo

HAMAJIMA *et al.* (2002) realizou um estudo colaborativo de 53 casos epidemiológicos de mulheres tabagistas chegando a conclusão que o tabagismo exerce pouca ou nenhuma influência no risco de desenvolvimento de carcinoma mamário. Os estudos epidemiológicos de (INNES e BYERS, 2001) apontaram que houve acentuação do percentual de risco para mulheres pertencentes a subgrupos específicos, tais como as mulheres jovens fumantes durante o período gestacional.

Em suma, conforme definido por Campos (2006) a análise epidemiológica do câncer mamário é vasta, entretanto a associação dos fatores de risco da doença não explica a maioria das ocorrências dessa patologia. Sendo assim a melhor estratégia de prevenção da

saúde populacional é realizar o rastreamento conforme orientações por faixa etária (CAMPOS, 2006; LAMAS, 2000).

2.4 A MAMOGRAFIA E A CLASSIFICAÇÃO BI-RADS DOS ACHADOS

A mamografia, um dos testes mais sensíveis no rastreamento do câncer de mama em casos assintomáticos, viabiliza a detecção do câncer de mama em uma fase incipiente, não diagnosticada pelo exame clínico (QUEIROZ e HUMMEL, 1995; LAMAS, 1998; ANDOTRIA, 2004). Em comparação a outras técnicas, a mamografia é um dos exames mais empregados, fato que contribui no acúmulo e difusão de conhecimento entre especialistas da área (PIATO, 1995). Entretanto, a detecção de câncer de mama em mulheres assintomáticas depende de múltiplos fatores, entre eles a qualidade e forma de aquisição do mamógrafo, a forma como as imagens são processadas e a experiência do profissional responsável pela interpretação das imagens mamográficas (LAMAS, 2000).

O BI-RADS do Colégio Americano de Radiologia (*American College of Radiology* – ACR), resultou da colaboração de membros das comissões ACR e outras instituições tais como o *National Cancer Institute* (Instituto Nacional do Câncer), o *American Medical Association* (Associação Médica Americana), o *College of American Pathologists* (Colégio dos Patologistas Americanos), o *American College of Surgeons* (Colégio Americano de Cirurgiões), os Centros para Prevenção e Controle de Doenças e a *Food and Drug Administration* (FDA). O protocolo BI-RADS normatiza a descrição e manejo dos achados mamográficos simplificando a comunicação entre especialistas da área médica. Segundo ACR (2005), os achados, podem ser classificados, conforme o grau de risco para câncer de mama, em seis categorias BI-RADS:

- *Categoria 0* – categoria bastante empregada em exames de rastreamento, geralmente utilizada após um estudo mamográfico completo. Caracteriza achado para o qual é necessário uma avaliação adicional por imagem.
- *Categoria 1* – não é necessário nenhum comentário, entende-se que as mamas não apresentam alteração que exija investigação.

- *Categoria 2* – é uma avaliação normal, entretanto constam achados benignos que merecem serem descritos no relatório mamográfico. Nessa categoria estão inclusos fibroadenomas, calcificações involuídas, múltiplas calcificações secretórias, lesões com conteúdo de gordura (cistos oleosos, lipomas, galactoceles e hamartomas de densidade mistas), linfonodos intramamários, calcificações vasculares, implantes ou distorção arquitetural evidentemente associados a cirurgia.
- *Categoria 3* – caracteriza os achados não palpáveis de alta probabilidade de benignidade, onde 98% dos casos são provavelmente benignos, mas que sugerem um seguimento inicial em curto intervalo de tempo (6 meses). Nessa categoria entram o nódulo sólido circunscrito não calcificado, a assimetria focal e o agrupamento de calcificações redondas. Conforme relatos na literatura (SICKLES, 1995; LAMAS, 2004), 2% dessas lesões podem evoluir com sinais claros de malignidade.
- *Categoria 4* – são as anormalidades suspeitas, em que deve ser considerada uma biópsia para prosseguir a investigação. Essa categoria abrange os achados que, embora não sejam tipicamente malignos, possuem uma probabilidade maior de corresponder a lesões malignas. Na 4ª edição do BI-RADS essa classificação é subdividida em 4A, 4B e 4C, conforme o risco de câncer, o qual varia entre 2% até 95%.
- *Categoria 5* – são os achados altamente suspeitos de malignidade que exigem uma avaliação histológica. Nesse caso há uma probabilidade de câncer superior a 95% e que pode ser indicada uma intervenção cirúrgica sem uma biópsia preliminar.
- *Categoria 6* – há uma malignidade comprovada do achado, o qual já teve sua biópsia conhecida, sendo necessária uma conduta terapêutica adequada.

2.5 PROCESSAMENTO DIGITAL DE IMAGENS - PDI

As técnicas de PDI fundamentam-se em operações matemáticas visando dois objetivos principais: a melhoria da informação visual para a visão humana e o processamento de dados da imagem para análise automática de máquinas (GONZALEZ e WOODS, 2008). Abrangendo vários tópicos tais como o realce, a filtragem, a restauração, análise e a reconstrução de imagens. O PDI permite que o tratamento e extração de informações de imagens digitais sejam realizados de forma rápida e precisa, viabilizando a extração de dados que poderiam ser inviáveis de obter-se manualmente (GOMES, 2001).

Geralmente um sistema de processamento digital de imagens digitais apresenta as etapas ilustradas na Figura 17.

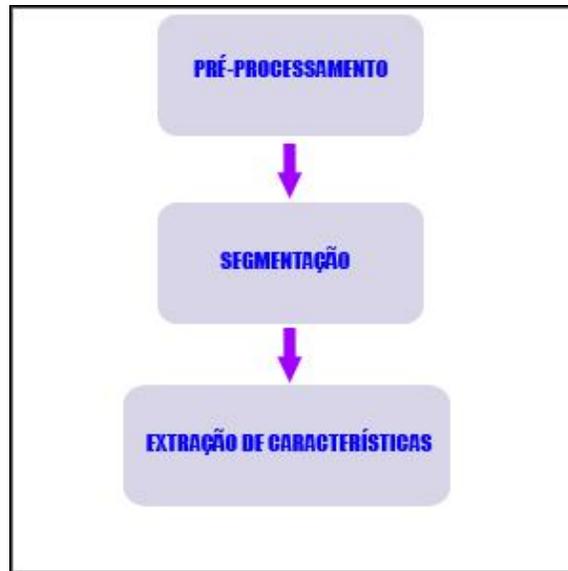


Figura 17: Etapas do PDI (Adaptado de GOMES, 2001).

As operações de pré-processamento são aplicadas na imagem para evidenciar características de interesse ou preparar a imagem para as demais etapas do processamento (GONZALEZ e WOODS, 2008). Nesse ínterim as técnicas mais conhecidas são a filtragem no domínio espacial e frequencial.

O processamento no domínio espacial atua diretamente sobre o arranjo espacial dos *pixels* da imagem, conforme a Equação 1 (GONZALEZ e WOODS, 2008):

$$g(x, y) = T[f(x, y)] \quad (1)$$

Equação 1: Operação de processamento espacial em imagem. No qual $f(x,y)$ denota a imagem de entrada e $g(x,y)$ a imagem de saída realçada pelo operador T que atua sobre a vizinhança do ponto (x,y) (GONZALEZ e WOODS, 2008).

A filtragem de imagens no domínio espacial é citada na literatura como a operação de convolução cujas operações básicas de filtragem são mencionadas a seguir, conforme JÄHNE *et al* (1999) e PRATT (2001):

- *Identidade* – corresponde a uma cópia idêntica, ponto a ponto da imagem.
- *Negativa* – geralmente utilizada para destacar sutis detalhes brancos ou cinza da imagem sobre um fundo preto, essa operação consiste em inverter a

intensidade luminosa de todos os pixels da imagem (similar ao negativo fotográfico).

- *Logarítmica* – aplica uma transformação logarítmica a cada ponto da imagem. Essa operação geralmente é utilizada para evidenciar partes da imagem que estavam muito escuros.
- *Potência* – atua de forma similar a operação logarítmica, sendo que nesse caso é aplicada uma transformação de potência a cada ponto da imagem.
- *Aritmética* – são as operações de soma, subtração, multiplicação e divisão entre *pixels* de duas imagens, resultando em uma terceira imagem. Tais operações são normalmente utilizadas para evidenciar níveis de intensidade luminosa da imagem, localizar distinções entre imagens similares, minimizar ruídos e gerar novas imagens a partir da média das imagens originais.
- *Lógicas* – empregam as operações lógicas E, OU, NAO e suas derivadas para destacar regiões específicas da imagem, além de serem bastante úteis nas operações morfológicas.

As operações realizadas no domínio da frequência (representado por $F(u,v)$) constituem outra forma de representar as informações da imagem. A Transformada Discreta de Fourier (TDF) realiza o mapeamento de dados no domínio espacial para F no domínio frequencial (CABRAL, DESCHAMPS e STEMMER, 2002). A TDF emprega um somatório de senos e cossenos com seus respectivos pesos para traduzir o sinal no domínio do tempo para o domínio da frequência (OPPENHEIM, WILLSKY e NAWAB, 1997; KREYSZIG, 1999). O processamento no domínio frequencial segue as etapas ilustradas na Figura 18.

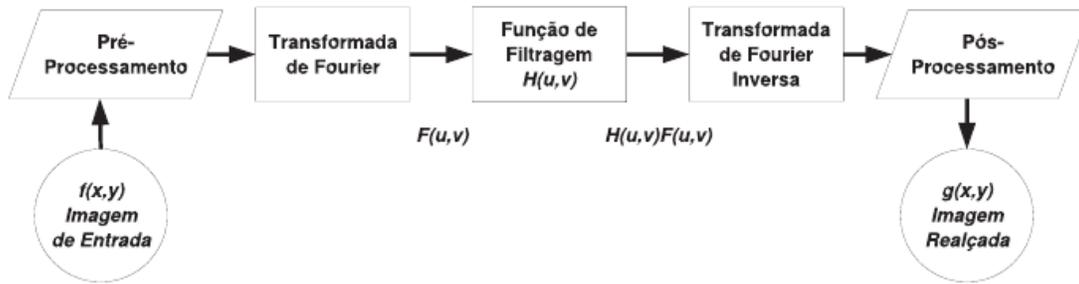


Figura 18: Etapas do processamento frequencial de imagens (GONZALEZ E WOODS, 2008).

2.5.1 Segmentação

A segmentação visa dividir a imagem em componentes significativos ou em ROIs para facilitar a análise. O nível de divisões da imagem é determinada conforme o objetivo almejado, geralmente até um nível de distinção satisfatório para distinguir os objetos de interesse na imagem analisada (GONZALEZ e WOODS, 2008).

Os algoritmos de segmentação baseiam-se principalmente em duas propriedades do nível de intensidade luminosa das imagens: a similaridade e a descontinuidade. Nos algoritmos de segmentação por similaridade as imagens são divididas conforme um padrão de semelhança entre regiões (textura, cor, nível de cinza). Enquanto que nos algoritmos de segmentação por descontinuidade a imagem é dividida utilizando como referência as mudanças abruptas do nível de intensidade luminosa entre as regiões da imagem (arestas, cantos, objetos da imagem) (PRATT, 2001).

2.5.2 Morfologia Matemática na extração de características

A extração de características tem como intuito principal o reconhecimento de padrões por meio de operações inteligentes na interpretação de sinais. As operações morfológicas, baseada na teoria de conjuntos e operações lógicas sobre imagens digitais, destacam-se como técnicas de extração de características de imagens digitais (GONZALEZ e WOODS, 2008). Estas operações são úteis na identificação e extração de características da imagem, além de auxiliar na identificação e minimização de ruídos que interferem na forma dos objetos contidos na imagem. Em geral, a aplicação das operações morfológicas simplifica os objetos de interesse e realça as características essenciais dos mesmos (SILVA, 1995). Os operadores morfológicos empregam um elemento estruturante para percorrer a imagem, de forma similar as máscaras de convolução utilizadas nas operações de filtragem. A seguir

consta um resumo das principais operações morfológicas, conforme M. Filho e V. Neto (1999):

- *Dilatação* – combina dois conjuntos através da adição vetorial de elementos dos conjuntos. O elemento estruturante é aplicado de forma concêntrica sobre um conjunto de pontos predefinido de uma imagem, de tal forma que o elemento estruturante adicione informação sobre a vizinhança desses pontos. Essa operação tem como principal objetivo o preenchimento de intervalos e lacunas indesejáveis na imagem.
- *Erosão* – realiza o inverso da dilatação, combina dois conjuntos e realiza a subtração vetorial entre os conjuntos por onde o elemento estruturante percorre. Esta operação tem como principal finalidade a extração de detalhes e ruídos indesejáveis na imagem.
- *Abertura* – consiste na aplicação de uma dilatação seguida de uma erosão. Com aplicação da erosão pretende-se eliminar ruídos e abrir lacunas em regiões de fraca conexão entre objetos. O resultado da operação anterior passa por uma dilatação que visa restaurar as dimensões reais dos objetos na imagem.
- *Fechamento* – é a operação inversa a abertura (erosão seguida de uma dilatação) e tem como intuito a restauração de conexões fracas entre objetos da imagem.

A Figura 19 ilustra os princípios de atuação dos operadores morfológicos mencionados.

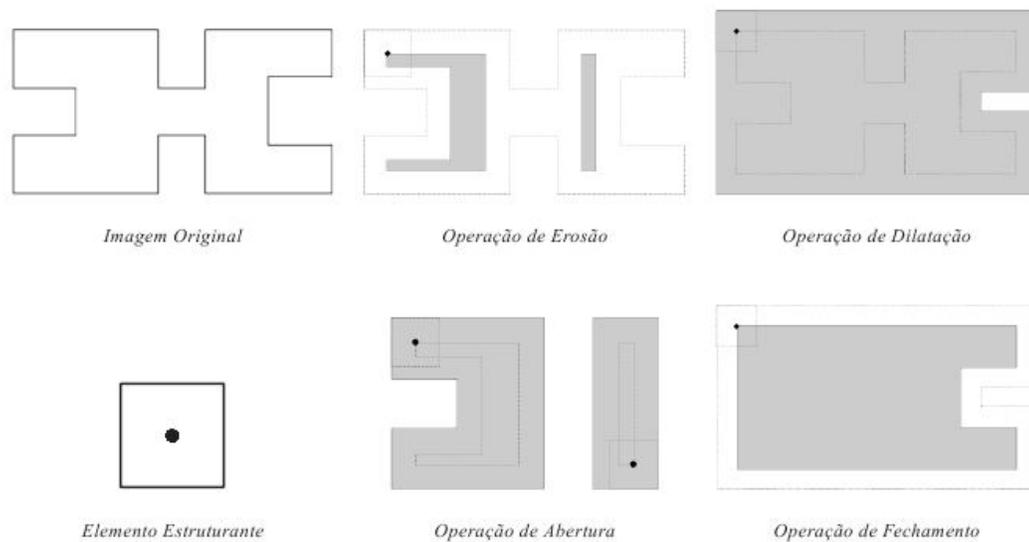


Figura 19: Fundamentos de atuação dos operadores morfológicos (GONZALEZ E WOODS, 2008).

2.5.3 Transformada de Wavelet

A Transformada de *Wavelet* (TW), criada em 1984, tem grandes aplicações na análise de sinais digitais e compressão de imagens. Muitos estudos na literatura (REZAI-RAD et al., 2005; JASMINE et al., 2009) têm relatado a eficiência do uso de *Wavelets* para detectar calcificações, visto que após a decomposição da imagem em faixas de frequência é possível observar que geralmente as calcificações concentram-se em uma determinada faixa. Esta transformada atua no domínio das frequências, similar a TDF, mas com uma atuação otimizada em comparação a *Transformada de Fourier de Curto Termo* (TFCT), visto que a TFCT analisa o sinal em uma escala fixa, o que pode ocasionar perda informações importantes se o sinal extrapolar essa janela. Já TW emprega janelas de tempo com tamanhos variados nos processos de “dilatação”. Na adição ela é calculada através do processo de “translação” o qual ocorre com o deslocamento do sinal ao longo do sinal (COX, 2004).

A TW adota combinações de variações de escala e deslocamento, sendo expressa matematicamente como o produto interno de uma função $x(t)$ com um conjunto de funções bases (MALLAT, 1999), as quais são definidas na Equação 2.

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-\tau}{a}\right) \quad (2)$$

Na Equação 2 a é denotado como parâmetro de dilatação e τ consiste no parâmetro de translação. O parâmetro de translação τ indica a posição wavelet $\psi_{\tau,a}(t)$, e a frequência é controlada pelo parâmetro de dilatação. O termo $\frac{1}{\sqrt{|a|}}$ visa conservar a energia com a dilatação das funções Wavelets filhas. A Equação 3 apresenta a função Wavelet de sinal não estacionário $x(t)$, nesse caso ψ^* é o complexo conjugado da Wavelet base.

$$W_x(\tau, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-\tau}{a}\right) dt \quad (3)$$

O algoritmo da Transformada de Wavelet Rápida (TWR) aborda a decomposição de um sinal discreto através de um conjunto de filtros que extraem as características médias (componentes de alta escala e baixas frequências do sinal - aproximações) do sinal. Em seguida são somados um conjunto de filtros baseados nas funções Wavelets, responsáveis pela extração de características de alta frequência (as componentes de baixa escala e alta frequência - detalhes) do sinal (MALLAT, 1999). Esse processo de decomposição corresponde na prática à aplicação de filtros passa-alta e passa-baixa através de múltiplas decomposições, conforme ilustrado na Figura 20.

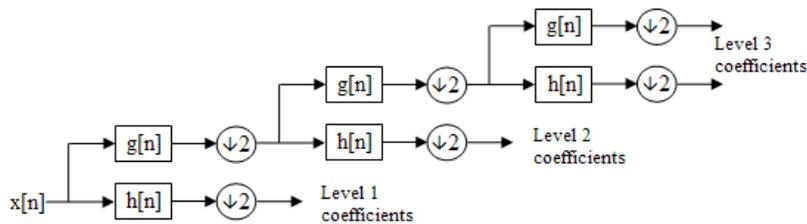


Figura 20: Árvore de decomposição Wavelet, contendo 3 níveis. Em cada nível são utilizados filtros passa-baixas obtendo as Aproximações (g_n) e filtros passa-altas obtendo os Detalhes (h_n) (PEREIRA, 2003).

2.6 O USO DE RBC COMO SISTEMA DE APOIO À DECISÃO

O uso de sistemas de apoio à decisão está vinculado a necessidade de auxiliar profissionais a difundir o conhecimento implícito e mapear computacionalmente os processos de raciocínio que levam a tomada de decisão. Nesse contexto, os sistemas de RBC consistem em técnicas computacionais derivadas da IA que utilizam o conhecimento tácito para modelá-lo em um sistema inteligente hábil na utilização de soluções anteriores para resolução de novos casos (WANGENHEIM, 2003). Diferente dos Sistemas Especialistas convencionais que utilizam o Raciocínio Baseado em Regras (WATTERMAN, 1986) com sequencias lógicas compostas por premissas e conclusões, o RBC modela o conhecimento a partir de uma base de casos que constituem o conhecimento previamente experimentado (WANGENHEIM, 2003; SILVA, 2005). Geralmente cada caso possui uma descrição do problema e a solução, podendo conter o conjunto de sucessos e fracassos dos casos relacionados resolvidos (AAMODT e PLAZA, 1994). Na base de casos o RBC usa estratégias de inferências para identificar a situação atual e localizar na memória a experiência semelhante para solução do novo problema (KRAUS, 2009).

2.6.1 Etapas de desenvolvimento do sistema RBC

Segundo Silva (2005) apesar dos sistemas RBC apresentarem forte dependência do conteúdo e da estrutura de sua base de casos, o conhecimento dele não está restrito a sua memória de casos, dependendo também das suas etapas de desenvolvimento. O modelo de processo RBC mais aceito é o proposto por Aamondt e Plaza (1994), ilustrado na Figura 21, contém as seguintes etapas:

Recuperação – busca na sua base o(s) caso(s) mais similar ao novo problema. Nessa etapa é realiza a busca de índices e o cálculo de similaridade entre o caso recuperado e o novo problema.

Reutilização – adota a solução vinculada ao(s) caso(s) recuperado(s) no contexto do problema atual.

Revisão – se necessário, identifica as diferenças para adaptar a solução do(s) caso(s) recuperado(s) na etapa de reutilização. Esse processo caracteriza também um mecanismo de aprendizado por falhas.

Retenção – mecanismo de aprendizagem que incorpora o que for útil na solução do novo problema na biblioteca de casos.

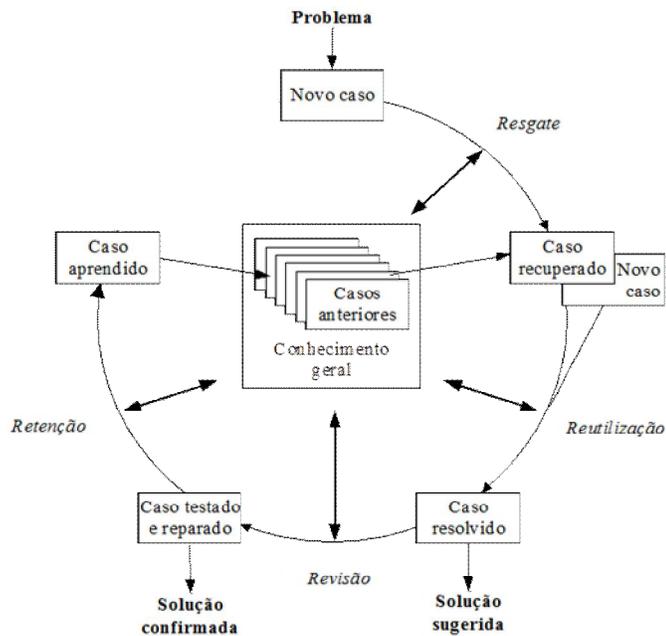


Figura 21 Ciclo RBC (AAMODT e PLAZA, 1994).

Geralmente a concepção de um sistema RBC contempla as seguintes etapas:

Seleção da base de informações – consiste na definição do banco de dados, ou conjunto de registros que representem o conhecimento. Na ausência da base de dados formal, torna-se necessário obter os casos diretamente do(s) especialista(s). Fernandes (2003) caracteriza essa etapa como crítica, visto que o sistema RBC depende fundamentalmente de sua base de conhecimento, portanto a decisão do quê e como armazenar torna-se uma tarefa complexa. Os casos armazenados na base podem estar presentes no domínio de forma natural, quando eles são dispersos e não abrangem satisfatoriamente o domínio ou são suficientes para cobrir o domínio. Há também os domínios artificiais nos quais os problemas tendem a repetir-se de forma sistemática. Em domínios artificiais geralmente são empregadas técnicas de adaptação mais simples para solução de casos recuperados, sendo possível até a inexistência de adaptação, visto que as soluções dos problemas são simples e não necessitam de adaptação.

Definição de atributos relevantes a solução de problemas - através da consulta ao(s) especialista(s) em conjunto com a análise do próprio domínio do conhecimento.

Definição de índices – decisão da estrutura apropriada para a caracterização, recuperação e reutilização dos casos.

Definição de métodos de recuperação de casos – utilização de técnicas de verificação de similaridade entre os casos contidos na base e os novos casos que o sistema irá consultar. A similaridade pode ser definida matematicamente, porém o método deve receber a validação do(s) especialista(s). A métrica da similaridade é uma função que mede a similaridade entre dois casos, ela é utilizada para determinar os casos mais similares ordenando-os de acordo com o nível de maior similaridade (Weber, 1996). Segundo Wangenheim (2003) a determinação de similaridade em RBC pode ser formalizada sob os seguintes conceitos:

- I. *Predicado* – concebe a similaridade como uma relação de pertinência ou não pertinência entre objetos ou fatos;
- II. *Relação de preferência* – considerada a proporção maior ou menor similaridade entre os dados analisados;
- III. *Medida* – trata a similaridade através da quantificação ou extensão da semelhança. Essa é a forma mais utilizada.

As similaridades geralmente são normalizadas em uma faixa de 0 (dissimilaridade total) a 1 (coincidência absoluta). De acordo com Wangenheim (2003) a similaridade é avaliada em um contexto global e local. A Similaridade Global analisa a utilidade de um caso em relação a uma determinada pergunta, onde a similaridade entre a pergunta e o caso deve ser determinada. Conforme Wangenheim (2003) e Kraus (2009) as medidas de similaridade Global mais difundidas em RBC são as seguintes:

- a) *Vizinho mais próximo* – técnica que busca determinar o vizinho geometricamente mais próximo, após a definição de uma medida de distância, definido pela Equação 4:

$$SIM(T, S) = \sum_{i=1}^n Sim_i(T_i, S_i) \quad (4)$$

Onde:

T é o caso de entrada

S é o caso da base

n é o número de atributos de cada caso

i é um atributo individual

Sim é a função de similaridade para o atributo i nos casos T e S

- b) *Vizinho mais próximo ponderado* – é o vizinho mais próximo alterado pelo peso dos atributos do caso, conforme expresso na Equação 5.

$$SIM(T, S) = \sum_{i=1}^n Sim_i(T_i, S_i) \times w_i \quad (5)$$

Onde w é o peso dado ao atributo i .

- c) *Vizinho mais próximo normalizado* – é o vizinho mais próximo ponderado dividido pelo total de casos da base conforme denotado pela Equação 6.

$$SIM(T, S) = \frac{\sum_{i=1}^n Sim_i(T_i, S_i) \times w_i}{\sum_{i=1}^n w_i} \quad (6)$$

- d) *Distância Euclidiana* – corresponde a distância real entre dois pontos em um espaço, conforme denotado pela Equação 7.

$$SIM(T, S) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (7)$$

Onde a e b correspondem respectivamente à pares de atributo-valor dos índices do caso pesquisado e caso da base. Da mesma forma que o vizinho mais próximo a Distância Euclidiana também pode ser utilizada de forma ponderada (considera o valor dos pesos dos atributos) e normalizada (considera o peso dos atributos e normaliza o resultado empregando a quantidade de casos analisados no cálculo).

- e) *Medidas de contraste* – consiste em uma alternativa aos modelos geométricos de similaridade. O cálculo da similaridade é realizado através da combinação linear das diferenças ponderadas existentes entre atributos correspondentes e divergentes. Uma forma de obter medida de contraste é através da criação dos seguintes conjuntos de atributos:
- Conjunto que contém os atributos disponíveis tanto na situação analisada quanto no caso avaliado da base;
 - Conjunto de atributos que constam somente no caso avaliado da base;
 - Conjunto de atributos disponíveis somente na situação analisada.

A Equação 8 define como esses conceitos são empregados para realizar o cálculo da similaridade por medida de contraste.

$$SIM(T, S) = (a \sum_{i \in S \cap C} wi) - (b \sum_{i \in C - S} wi) - (c \sum_{i \in S - C} wi) \quad (8)$$

A Similaridade Local avalia a similaridade no nível dos atributos ou características relevantes de um caso. As estratégias mais empregadas para mensurar a similaridade local é a combinação de medidas de similaridade local ou a consideração de importância ou pesos distintos para os diferentes atributos. A seguir constam algumas das principais abordagens de mensuração da similaridade local, segundo Wangenheim (2003) e Kraus (2009):

- a) *Função escada* – calcula a distância entre dois valores, conforme especificado pela Equação 9. O resultado da similaridade local é 1 quando o resultado da diferença é menor que o limiar estipulado do contrário a similaridade local será 0.

$$Sim(v_i, v_j) = \begin{cases} 1, se |v_i - v_j| \leq S \\ 0, se |v_i - v_j| > S \end{cases} \quad (9)$$

- b) *Função linear* – nessa função a similaridade cresce com o decréscimo da distância entre valores, sendo ponderada pelo tamanho do intervalo assumido pelo domínio

dos valores do atributo, onde ls é o limite superior da faixa de valores do atributo e li o limite inferior, conforme expresso na Equação 10.

$$Sim(v_i, v_j) = \begin{cases} 1, & se\ v_i = v_j \\ 1 - \frac{|v_i - v_j|}{ls - li}, & se\ v_i \neq v_j \end{cases} \quad (10)$$

c) *Função Igual* - consiste em uma função binária que verifica se os atributos possuem valores iguais, conforme Equação 11.

$$Sim(v_i, v_j) = \begin{cases} 1, & se\ v_i = v_j \\ 0, & se\ v_i \neq v_j \end{cases} \quad (11)$$

d) *Função Elemento Máximo* – expressa o valor máximo de similaridade entre todos os elementos de um conjunto de atributos, conforme pode ser visto na Equação 12.

$$Sim(v_i, v_j) = MAX(v_i, v_j), \forall v_i \in v_j \quad (12)$$

e) *Função Interseção* – consiste numa função multivalorada, cujo resultado considera o numero total de atributos para calcular a quantidade de valores correspondentes entre v_i e v_j sem utilizar o peso das características, conforme Equação 13:

$$Sim(v_i, v_j) = \frac{|v_i \cap v_j|}{|v_i \cup v_j|} \quad (13)$$

f) *A Função do Contraste* – utiliza um contraste linear das diferenças ponderadas entre os atributos comuns e discordantes, conforme especificado pela Equação 14.

$$Sim(v_i, v_j) = a(v_i \cap v_j) - b(v_i - v_j) - c(v_j - v_i) \quad (14)$$

Onde as variáveis a , b e c são os pesos de importância dos casos analisados:

Definição da forma de adaptação dos casos – em domínios pouco complexos é possível construir bases que abordem todas as possibilidades de combinação, tornando a adaptação inexistente. Entretanto quando essa etapa é contemplada deve ser baseada no conhecimento e em heurísticas definidas pelo(s) especialista(s). Essa etapa também é conhecida como a responsável pela aquisição do conhecimento, a qual segundo Wangenheim (2003) é de suma importância para que o sistema mantenha-se atualizado e evolua continuamente.

2.6.2 Vantagens e aplicações dos sistemas RBC

Segundo Silva (2005) e Kraus (2009), dentre as várias vantagens do uso de sistemas RBC destacam-se:

- A não necessidade de modelo explícito do domínio como nos sistemas baseados em modelo;
- Reutilização da informação por meio da utilização de bases de dados pré-existentes que possuem o conhecimento tácito;
- Possibilidade de aprendizado automático com novos casos;
- A incerteza e raciocínio implícito presentes no domínio não requerem um tratamento específico dos casos, pois se o problema abordado for mapeado adequadamente, o sistema não precisará entender todas as peculiaridades para propor uma solução (THÉ, 2001).

Tais vantagens contribuíram para que os sistemas RBC sejam adotados em várias áreas tais como os exemplos a seguir:

Sistemas de apoio à decisão e ao diagnóstico. Exemplos: CASEY (avalia causa e solução de problemas cardíacos), CELIA (avaliação de automóveis) (LORENZI, 1998; WANGENHEIM, 2003).

Help Desk. Exemplos: CAMPAQ (suporte ao cliente, diagnóstico e consertos) (LORENZI, 1998; WANGENHEIM, 2003).

Projeto. Exemplos: CHEF (elaboração de receitas), CLAVIER (projeto e modelos de autoclave), ARCHIE (projetos arquitetônicos), CADET (projeto de componentes mecânicos) (LORENZI, 1998; WANGENHEIM, 2003).

Planejamento e mediações. Exemplos: NETTRAC (planejamento e monitoramento do tráfego de redes de telefonia público), MEDIATOR (atua no planejamento, avaliação e consertos em disputas), PLEXUS (planejamento de atividades cotidianas), etc. (LORENZI, 1998; WANGENHEIM, 2003).

De acordo com Kraus (2009) e Wangenheim (2003) o uso de RBC é indicado em situações onde:

- O domínio é complexo, entretanto as soluções podem ser justificadas por meio de casos;
- As experiências passadas podem ser empregadas para resolver problemas futuros;
- Em domínios onde a teoria é insuficiente, imprecisa ou o encadeamento de regras é muito longo inviabilizando o uso de regras para mapear o problema, entretanto constam casos ou o especialista do domínio.
- Quando já existe uma base de casos, ou quando as regras formuladas requerem mais entradas de dados do que normalmente está disponível ao usuário.

2.6.3 Validação do sistema RBC na área médica

Dentre as metodologias difundidas para confirmar ou descartar um diagnóstico na área médica estão a sensibilidade e especificidade. De acordo com ACR (2005) a sensibilidade caracteriza a probabilidade de detectar um câncer quando existe um câncer ou número de cânceres diagnosticados após serem identificados em mamografias de uma determinada população no intervalo de um ano do exame de diagnóstico por imagem, dividido por todos os cânceres presentes naquela população no mesmo período de tempo. Enquanto que a especificidade é o oposto da sensibilidade, ou seja, é a probabilidade de se interpretar um exame como negativo, quando não existe um câncer. Os conceitos de sensibilidade e especificidade podem ser aplicados em uma escala maior conforme definido por Friedland (2001) nas seguintes possibilidades distintas:

- *Verdadeiro-positivo* – resultado positivo em paciente com doença;
- *Falso-positivo* – resultado negativo em paciente não portador da doença;
- *Falso-negativo* – resultado positivo em paciente portador da doença;
- *Verdadeiro-negativo* – resultado negativo em paciente não portador da doença.

O método sensibilidade e especificidade, conforme denotado pelas Equações 15 e 16 respectivamente, consiste em uma das estratégias de validação de sistemas RBC na área médica (SILVA, 2005).

$$Sensibilidade = \frac{VP}{(VP + FN)} \quad (15)$$

$$Especificidade = \frac{VN}{(FP + VN)} \quad (16)$$

2.7 RECONHECIMENTO DE PADRÕES EM IMAGENS COM RNA PMC

Dentre as diversas faculdades mentais que visam ser automatizadas e modeladas por um computador, a aprendizagem e o reconhecimento de padrões são aspectos fortemente abordados em uma subárea de estudo da IA – as RNAs. Soluções diversas neste campo das RNAs vêm sendo apresentadas desde os primeiros modelos matemáticos baseados no comportamento do neurônio biológico, introduzidos por McCulloch e Pitts (1965). De modo generalizado, as diversas RNAs existentes buscam um modelo alternativo ao modelo de computação tradicional baseado nas máquinas de Von Neumann. As RNAs inovaram ao trazer novos conceitos na resolução de problemas computacionais, tais como a aprendizagem por exemplos por padrões de dados de entrada, e a conseqüente inferência de regras deste conjunto; a não necessidade de programação, uma vez que se atém a executarem operações não lógicas; a capacidade de testar todas as possibilidades de reações a estímulos em paralelo e a generalização diante da incompletude dos dados apresentados para o problema.

A RNA *Perceptron* Múltiplas Camadas (PMC) é caracterizada por um conjunto de neurônios interligados por múltiplas camadas. As PMCs podem usar um grande número de técnicas de aprendizado, sendo que a mais popular é a propagação reversa ou *backpropagation*. Neste caso, os valores da saída são comparados com a resposta correta para computar o valor de alguma “função-erro” predefinida, em seguida o erro é alimentado de volta na RNA. Usando essa informação, o algoritmo ajusta os pesos de cada conexão para reduzir o valor da “função-erro”. Após repetir este processo por um número suficiente, os pesos sinápticos são ajustados para fazer com que a resposta real se aproxime da resposta desejada (HAYKIN, 1999). A Figura 22 apresenta um modelo de RNA PMC:

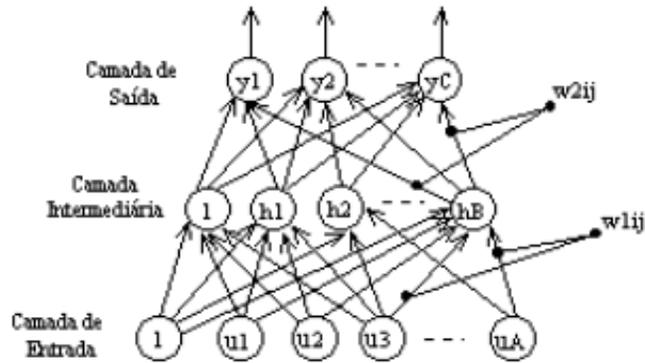


Figura 22: Exemplo de uma RNA PMC típica com uma camada oculta (AZEVEDO et al., 2000).

Como pode ser visto na Figura 22, cada camada da RNA PMC é conectada à outra por meio de uma matriz de pesos. Esses pesos são inicialmente definidos de forma aleatória e vão sendo ajustados na fase de retro-propagação. Na RNA PMC, cada camada, exceto a última, possui um *bias*.

As entradas da rede buscam espelhar as características do padrão que será aprendido. As saídas da rede denotam as classes de pertinência que serão empregadas para classificar o padrão. Os valores de entrada da rede são multiplicados pela matriz de pesos, resultando numa saída, que é a soma ponderada da entrada. Em tarefas de reconhecimento de padrões em imagens, o conjunto de treinamento é formado por imagens. Estas são submetidas a etapas de pré-processamento visando à redução de ruído, melhoria da qualidade e realce de regiões de interesse objetivando melhorar a qualidade do conjunto de treinamento, uma vez que ele está diretamente relacionado com a capacidade de aprendizado e generalização da rede neural (HAYKIN, 1999; ARTERO, 2008).

A soma ponderada da entrada da RNA é passada para uma função de ativação, responsável por executar o processo sináptico em cada neurônio, podendo causar um estímulo nos neurônios da camada seguinte. Os valores de entrada da RNA são multiplicados pela matriz de pesos, resultando numa saída, que é a soma ponderada da entrada (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000). Essa soma é passada por alguma função de ativação onde as mais empregadas na RNA PMC são a função

Sigmóide e a Tangente Hiperbólica, cujas fórmulas são apresentadas nas Equações 17 e 18, respectivamente:

$$y(x) = I / \{ 1 + \exp(-kx) \} \quad (17)$$

$$y(x) = \tanh(kx) = \frac{\exp(kx) - \exp(-kx)}{\exp(kx) + \exp(-kx)} \quad (18)$$

A Equação 17, explora a forma de não-linearidade sigmóide na sua forma geral, onde saída $y_j(x)$ é o campo local induzido do neurônio j . De acordo com esta não-linearidade, a amplitude da saída se encontra dentro do intervalo $0 \leq y_j \leq 1$. A medida que k cresce, a função torna-se mais abrupta nas proximidades de $x=0$. Quando k tende a infinito, a função é definida por $+1$ se $x \geq 0$ e 0 se $x < 0$ (HAYKIN, 1999; ARTERO, 2008).

A função Tangente Hiperbólica, apresentada na Equação 18, é a função logarítmica reescalada e modificada por um *bias* (onde k é um escalar positivo). A medida que k cresce, a função torna-se mais abrupta nas proximidades de $x=0$. Quando k tende a infinito, a função atua de forma similar a Sigmóide ($+1$ se $x \geq 0$ e 0 se $x < 0$) (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000).

Após o cálculo dos erros de cada unidade, o ajuste dos pesos de cada conexão é calculado, conforme a Equação 19.

$$\Delta W_{kj}(n+1) = \alpha W_{kj}(n) + \eta \delta_j y_j \quad (19)$$

Na Equação 19 o gradiente local $\delta_j(n)$ definido depende se o neurônio j é um nó de saída ou se é um nó oculto. O y_j corresponde ao sinal de entrada do neurônio j .

Após o cálculo dos pesos de cada conexão, os pesos são atualizados, segundo a regra delta generalizada (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000), conforme a Equação 20:

$$W_{kj}(n+1) = w(n) + \Delta W_{kj}(n) \quad (20)$$

A correção $\Delta w_{kj}(n)$, definida na Equação 20, é aplicada ao peso sináptico conectando o neurônio k ao neurônio j . A saída de uma camada da RNA é utilizada como entrada para a

camada seguinte. Os valores de entradas para cada camada vão sendo ajustados pelos pesos dessas, até obter uma saída condizente com a classificação almejada (HAYKIN, 1999).

2.7.1 A aquisição do conhecimento das RNAs

A resolução de problemas de RNA é processada por meio de duas etapas, a aprendizagem e testes. Na etapa de aprendizagem a RNA cria um mapeamento interno do problema analisado, cujas principais abordagens são o aprendizado supervisionado ou não-supervisionado (HAYKIN, 1999; BRASIL et. al. 2012). No aprendizado supervisionado um supervisor externo, detentor do conhecimento sobre o ambiente, produz o custo para cada padrão em um conjunto de treinamento visando reduzir a soma dos custos para esses padrões (FERNANDES, 2007), ou seja, busca-se ajustar os pesos sinápticos da RNA de modo a obter uma associação entre os valores de entrada e saída dos pares de treinamento. Enquanto que no paradigma de aprendizado não-supervisionado não existe um instrutor para direcionar a aquisição do conhecimento da RNA, em geral o sistema forma “agrupamentos naturais” dos padrões de entrada para assimilar as informações (DUDA, 2001).

Durante o treinamento da RNA PMC pode ocorrer um fenômeno denominado *overfitting* que caracteriza o ajuste demasiado dos dados fazendo que a rede piore sua taxa de acerto para entradas diferentes das empregadas na aprendizagem. Ou ainda pode ocorrer que durante o ajuste dos pesos da rede na retropropagação eles entrem em uma situação conhecida como *mínimo local*, que consiste em uma solução que não atende o problema para qual a RNA PMC está sendo treinada. Dentre os vários métodos empregados para evitar essas situações estão a adoção do termo *momentum* e *taxa de aprendizado* com valores condizentes com uma curva de aprendizado suave (HAYKIN, 1999; BRASIL et. al. 2012). Outra estratégia é a validação cruzada, a qual consiste na divisão do conjunto de dados aleatoriamente em um conjunto de treinamento e testes. O conjunto de treinamento é constituído pela divisão dos dados em dois subconjuntos distintos, nos quais de acordo com Haykin (1999) 80% dos dados devem ser empregados no treinamento da RNA PMC e o restante dos dados devem ser usados para validar o modelo. A validação da generalização da RNA é analisada pelo conjunto de testes o qual possui dados diferentes do conjunto de validação do modelo. De acordo com Fernandes (2007) o uso dos três conjuntos visa validar o modelo com um conjunto de dados diferentes dos empregados na estimação dos parâmetros da RNA.

A etapa de teste é utilizada para validar a capacidade de generalização da RNA, onde é analisado se o conhecimento adquirido pela RNA é suficiente para gerar saídas adequadas para vetores de entrada não apresentados durante a etapa de treinamento. Para avaliar o desempenho da RNA geralmente acompanha-se um Erro Médio Quadrático (EMQ) cujo critério adotado para definir o valor aceitável varia conforme a aplicação.

2.7.2 Algoritmo de aprendizagem *Backpropagation*

Basicamente, a aprendizagem por retropropagação de erro consiste de dois passos através das diferentes camadas da RNA: um passo para frente, a propagação, e um passo para trás, a retropropagação. No passo para frente, um padrão de atividade (vetor de entrada) é aplicado aos nós sensoriais da RNA e seu efeito se propaga através da mesma, camada por camada. Finalmente, um conjunto de saídas é produzido como a resposta real da RNA. Durante o passo de propagação, os pesos sinápticos da RNA são todos fixos. Durante o passo para trás, por outro lado, os pesos sinápticos são todos ajustados de acordo com uma regra de correção de erro. Especificamente, a resposta real da RNA é subtraída de uma resposta desejada (alvo) para produzir um sinal de erro. Este sinal de erro é então propagado para trás através da RNA, contra a direção das conexões sinápticas. Os pesos sinápticos são ajustados para fazer com que a resposta da RNA se mova para mais perto da resposta desejada, em um sentido estático (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000).

As RNAs que utilizam *backpropagation* trabalham com uma variação da regra delta, apropriada para RNA multi-camadas: a regra delta generalizada. A regra delta padrão essencialmente implementa um gradiente descendente no quadrado da soma do erro para funções de ativação lineares. RNAs sem camadas intermediárias podem resolver problemas onde a superfície de erro tem a forma de um parabolóide com apenas um mínimo (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000). Entretanto, a superfície do erro pode não ser tão simples e suas derivadas mais difíceis de serem calculadas. Nestes casos devem ser utilizadas redes com camadas intermediárias (MINSKY e PAPERT, 1988). Segundo HECHT-NIELSEN (1987), o matemático Kolmogorov publicou em 1957 um teorema que comprovou que tudo o que uma RNA pode aprender com n camadas intermediárias pode ser assimilado por uma RNA que possua uma única camada intermediária. De acordo com F. Azevedo, L. Brasil e R. Oliveira (2000), um parâmetro importante que deriva dessa questão é a quantidade

necessária de neurônios nas camadas intermediárias, visto que a quantidade de neurônios nas camadas de entrada e saída depende do problema em questão. Uma quantidade excessiva de neurônios nas camadas intermediárias pode levar a RNA a especializar-se e perder a capacidade de generalização, em contrapartida um número de neurônios menor que o necessário pode interferir na capacidade de aprendizado da RNA. Determinar a melhor quantidade de neurônios para as camadas intermediárias é um problema em aberto que até os tempos atuais possui apenas heurísticas a respeito. Uma possível solução para esse problema é a realizada por BRASIL (1999), que explora o uso de Algoritmos Genéticos para determinar o número de neurônios nas camadas intermediárias a partir de uma população inicial de RNAs com números distintos de neurônios nas camadas intermediárias.

A regra delta generalizada funciona quando são utilizadas na rede unidades com uma função de ativação semi-linear, que é uma função diferenciável e não decrescente (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000).

A taxa de aprendizado (η) é uma constante de proporcionalidade no intervalo $[0,1]$, pois este procedimento de aprendizado requer apenas que a mudança no peso seja proporcional (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000). Ou seja, esse parâmetro tende a manter a estabilidade no processo de minimização de erros. De acordo com (HAYKIN, 1999) a atribuição de um valor baixo para η induz uma queda suave do erro em direção a superfície de solução. Entretanto, o gradiente descendente real requer que sejam tomados passos infinitesimais. Assim quanto maior for essa constante, maior será a mudança nos pesos, aumentando a velocidade do aprendizado, o que pode levar à uma oscilação do modelo na superfície de erro. O ideal seria utilizar a maior η possível que não levasse à uma oscilação, resultando em um aprendizado mais rápido (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000).

O treinamento das RNAs PMC com *backpropagation* pode demandar muitos passos no conjunto de treinamento, resultando um tempo de treinamento consideravelmente longo. Se for encontrado um mínimo local, o erro para o conjunto de treinamento para de diminuir e estaciona em um valor maior que o aceitável. Uma maneira de aumentar a taxa de aprendizado sem levar à oscilação é modificar a regra delta generalizada para incluir o termo *momentum*, uma constante que determina o efeito das mudanças passadas dos pesos

na direção atual do movimento no espaço de pesos (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000).

Desta forma, o termo *momentum* leva em consideração o efeito de mudanças anteriores de pesos na direção do movimento atual no espaço de pesos. O termo momentum torna-se útil em espaços de erro que contenham longas gargantas, com curvas acentuadas ou vales com descidas suaves (HAYKIN, 1999; AZEVEDO, BRASIL e OLIVEIRA, 2000).

3 METODOLOGIA

Este capítulo descreve os procedimentos realizados pela metodologia aplicada na realização desse trabalho. A Figura 23 apresenta uma visão geral de cada uma das etapas que serão explanadas no decorrer desse capítulo.

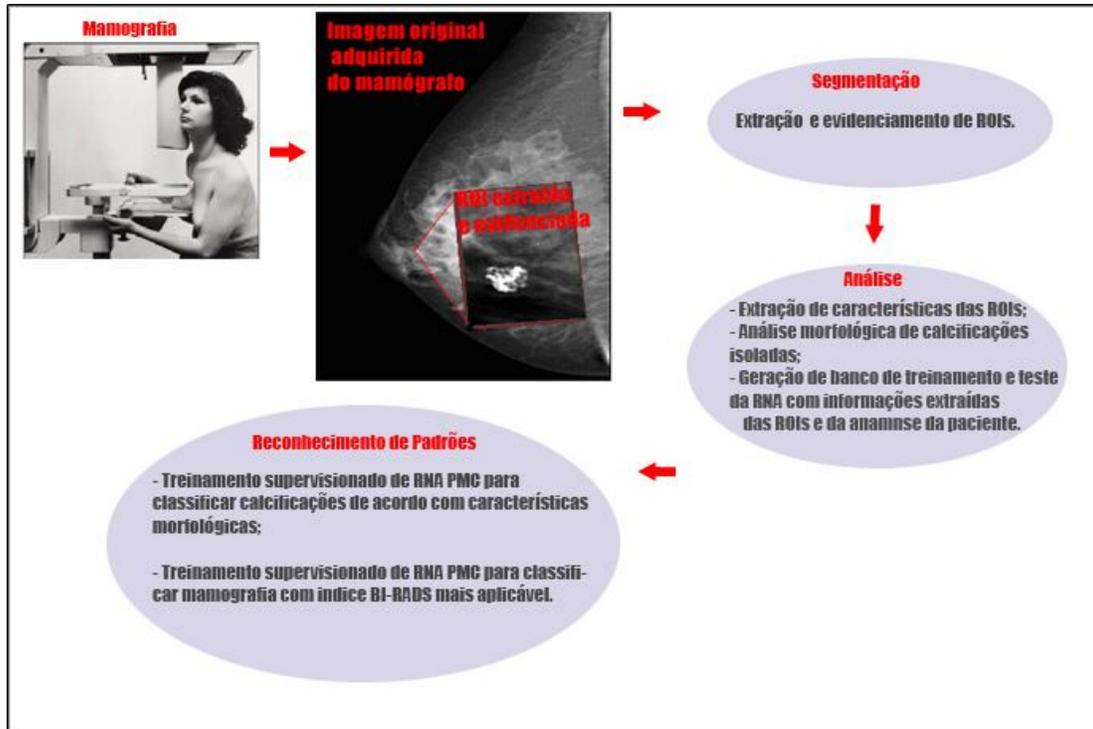


Figura 23: Etapas aplicadas na metodologia.

3.1 DEFINIÇÃO DA PESQUISA

A natureza dessa pesquisa é aplicada, pois está dirigida a solução de um problema específico: o desenvolvimento de sistema híbrido de RNA PMC e RBC para auxiliar na classificação BI-RADS de mamografias baseado na análise de calcificações. É uma pesquisa observacional descritiva, visto que analisa as características de uma determinada população de mamografias estabelecendo correlações para auxiliar uma conclusão do laudo mamográfico. É pesquisa bibliográfica, pois abrange o estudo sistematizado do conhecimento difundido por artigos, periódicos, livros, revistas, acervos de dados

eletrônicos, etc. Além disso, é também uma pesquisa de campo, ao passo que são abordadas técnicas de RNA e RBC, constituídas por treinamento, testes e acompanhamento dos resultados traçando-se um paralelo com casos reais.

3.2 DELIMITAÇÃO DO ESTUDO

A presente proposta destina-se a disponibilizar um sistema de auxílio na avaliação de calcificações mamárias a qual compreende uma RNA para classificar as calcificações mamárias em concordância com a nomenclatura e os critérios de diferenciação adotados pelo *BI-RAD* e um sistema RBC para aplicar categoria BI-RADS mais aplicável a mamografia de acordo com a análise das calcificações realizada anteriormente. Não será contemplada, neste trabalho, a análise e classificação de quaisquer outras estruturas ou a abordagem de necessidades que fujam ao escopo dessa pesquisa.

3.3 COLETA DE DADOS PARA RNA

3.3.1 FORMAÇÃO E AQUISIÇÃO DE ROIS

Os dados adotados nesse trabalho foram extraídos de imagens mamográficas, cujo formato inicial é o *Digital Imaging and Communications in Medicine* (DICOM) (NEMA, 2000) As mamografias empregadas foram cedidas pela clínica de estudo, CJLR. A amostra empregada nessa pesquisa é constituída por 50 exames mamográficos distintos contendo casos variados de calcificações. Cada exame possui cerca de 5 imagens diferentes incluindo a projeção Crânio Caudal (CC) caracterizada por ser ortogonal ao eixo vertical da paciente (EKLUD e CARDENOSA, 1992), a projeção Médio Lateral Oblíqua (MLO) a qual consiste em um ângulo oblíquo em relação ao eixo vertical da paciente (MUIR, KIRKPATRICK e ROBERTS, 1984) e *Spot* com compressão seletiva e ampliação (SICKLES, 1989), os quais buscam evidenciar melhor a estrutura da mama.

As ROIs utilizadas na pesquisa foram extraídas manualmente das imagens mamográficas e passaram por um processo de limiarização para transformá-las em binárias (onde os *pixels* assumem valores de 0 ou 1). Em seguida, realiza-se outra subdivisão para gerar ROIs referentes a calcificações individuais, extraídas de agrupamentos.

Ao término do procedimento de extração de ROIs referentes a calcificações individuais, foram reservadas 200 amostras as quais foram distribuídas dentre as 4 possibilidades de classificação analisadas no escopo desse trabalho, conforme a nomenclatura e os critérios de diferenciação do BI-RADS (ACR, 2005). Com a subdivisão das 200 ROIs obteve-se a seguinte amostra por tipo morfológico: 50 redondas, 50 em forma de pipoca, 50 puntiformes e 50 bastonetes.

Além da amostra de mamografias originais, na etapa de treinamento da RNA foram adotadas ROIs sintéticas simulando calcificações reais geradas por um algoritmo próprio (ELPÍDIO et. al., 2012) desenvolvido para constituir um banco controlado e ampliar o número de amostras por tipo morfológico das calcificações. Foram geradas 1000 ROIs de calcificações individuais para cada um dos 4 tipos morfológicos analisados nessa pesquisa. Em suma foram adotadas 4200 amostras ROIs para coleta dados da RNA. A Figura 24 consiste em um diagrama de atividades (GUEDES, 2009) que sumariza o processo de extração de características morfológicas das ROIs para geração das bases de treinamento e teste da RNA.

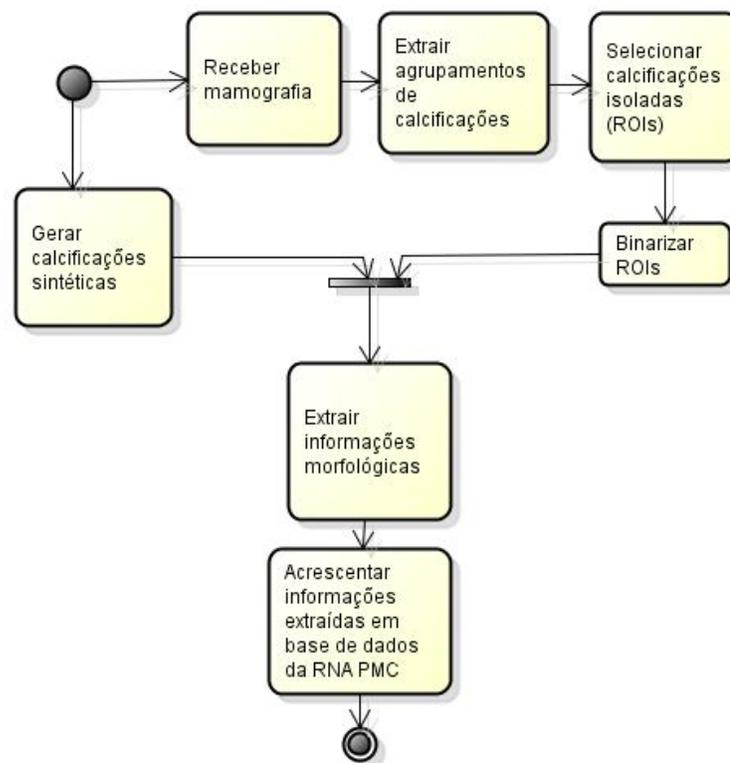


Figura 24: Diagrama de atividades para extração de informações morfológicas de ROIs.

3.3.2 EXTRAÇÃO DE ATRIBUTOS DE ROIS (PRÉ-PROCESSAMENTO DA RNA PMC)

Após o procedimento de extração de ROI, cada uma é processada por um algoritmo de extração de características que extrai as seguintes informações para mensurar as características morfológicas relevantes ao diagnóstico (MANGASARIAN, 1993; RANGAYYAN, MUNDIGONDA e DESAUTELS, 2000) de calcificações mamárias:

- *Área (A)* - obtida através da função *bwarea* do Matlab (2008). Ela é útil na análise do tamanho das calcificações. Segundo os estudos (ACR, 2005; MARTINS, BARRA e LUCENA, 2010), calcificações com tamanho reduzido (menor que 0,5mm) geralmente são associadas a malignidade.
- *Perímetro (p) das bordas* - mensurado através de uma função criada para calcular a soma da soma do resultado da função *bwperim* do Matlab (2008). Este fator é empregado em conjunto com outras características para analisar o aspecto que caracteriza a calcificação (Exemplos: circular, ovular, irregular).
- *Variância* - calcula o nível de dispersão (proximidade ou distância) entre as calcificações detectadas em aglomerados. Esse parâmetro é calculado através da função *find* do Matlab que retorna a disposição dos *pixels* que assumem o valor 1 e 0.
- *Excentricidade (E)* - define a distribuição espacial entre os eixos do objeto. É calculada pela razão entre o menor e maior eixo horizontal ou vertical (MANGASARIAN, 1993; RANGAYYAN *et al.*, 2000), conforme a Equação 21:

$$E = \frac{(\mu_{02} - \mu_{20}) + 4\mu_{11}}{A}, \quad (21)$$

onde μ_{pq} é o momento central dado pela Equação 22:

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q, \quad (22)$$

com (\bar{x}, \bar{y}) o centro de massa do objeto em estudo e sendo M e N sua largura e comprimento, respectivamente.

- *Nível de compactidade (c)* - combina o perímetro (p) e área (A) para estabelecer uma medida de compactação (MANGASARIAN, 1993; RANGAYYAN *et al.*, 2000), conforme a Equação 23:

$$c = \frac{p^2}{4\pi A} \quad (23)$$

- *Circularidade (C)* – utiliza o perímetro (p) e a área (A) para definir o quão próxima está a forma da ROI com um círculo (MANGASARIAN, 1993; RANGAYYAN *et al.*, 2000). É obtida por meio da Equação 24:

$$C = \frac{4\pi A}{(p)^2} \quad (24)$$

3.3.3 GERAÇÃO DE BASE DE TREINAMENTO E TESTE DA RNA

O conjunto de informações morfológicas extraídas das ROIs, conforme descrito no item 3.3.2, foi empregado para gerar as bases de treino e teste da RNA PMC.

Devido o pequeno tamanho da amostra dos dados reais, na fase de pré-processamento da RNA PMC foram adotadas ROIs provenientes da base de imagens sintéticas. O conjunto de treinamento foi dividido em dois subconjuntos sendo que um era composto por 70% e o outro 30% dos casos, de acordo com o método de validação cruzada descrito no item 2.7.1. E a validação da RNA foi realizada usando a base constituída com os parâmetros extraídos das ROIs reais extraídas das mamografias.

3.3.4 MODELO DE RNA PROPOSTO

A RNA adotada para indicar a classificação morfológica (ACR, 2005) das ROIs consiste na função do modelo PMC, *newff*, disponível na plataforma MATLAB (2008). Inicialmente foi empregada uma RNA PMC com a seguinte arquitetura: algoritmo de

aprendizado *Backpropagation*, 5 neurônios na entrada que correspondem a quantidade de informações morfológicas (previamente descritas no item 3.2.2) extraídas das ROIs, com exceção do parâmetro de variância, o qual é aplicado apenas na análise de aglomerados de calcificações. São adotados 7 neurônios na camada oculta (definidos empiricamente e seguindo a regra de EBERHART; DOBBINS, 1991) e 4 neurônios na camada de saída em consonância com as possibilidades de classificação morfológicas de calcificações analisadas nessa pesquisa. As funções de transferência *Tangente Hiperbólica* e *Sigmóide* foram empregadas em testes distintos para avaliar o desempenho da RNA. A Figura 25 demonstra a arquitetura da RNA PMC adotada.

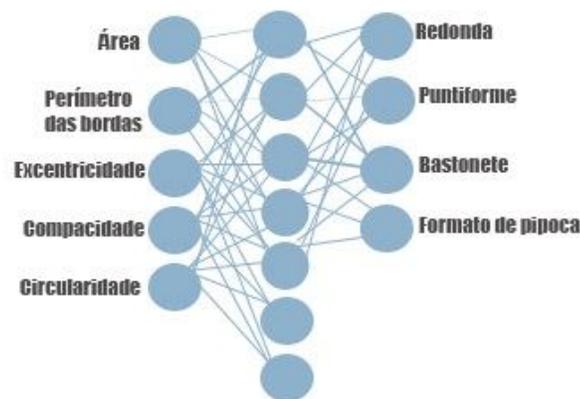


Figura 25: Arquitetura da RNA PMC adotada.

3.4 AQUISIÇÃO DE CONHECIMENTO E MODELAGEM DO SISTEMA RBC

Partindo da premissa que para qualquer aplicação há diferentes tipos de usuários com interesses geralmente distintos, por isso mesmo um sistema simples pode ter diferentes pontos de vista envolvidos, apesar de isso não caracterizar perspectivas inteiramente independentes. Nas abordagens orientadas a ponto de vista da engenharia de requisitos (PRESSMAN, 1995; SOMERVILLE, 2003) esse pontos de vista diferentes são utilizados para estruturar e organizar o processo de levantamento e os próprios requisitos. O processo de a aquisição e modelagem do conhecimento do RBC foi contemplado pelo método de definição de requisitos orientado a ponto de vista (VORD, do inglês *Viewpoint Oriented Requirements Definition*) conforme ilustrado na Figura 26:

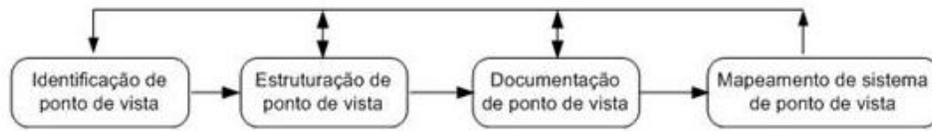


Figura 26: Método VORD utilizado na etapa de levantamento de requisitos do RBC (SOMMERVILLE, 2003).

3.4.1 LEVANTAMENTO DE REQUISITOS

Inicialmente foi realizado um estudo sistematizado do domínio, além de uma análise do funcionamento da técnica de RBC através de pesquisa bibliográfica dos conceitos a fim de assimilar as definições necessárias à representação do conhecimento e concepção do sistema. Nessa etapa também foi essencial obter informações específicas do domínio com a especialista da área, na Clínica Janice Lamas Radiologia. A especialista contribuiu com a definição das características e índices mais relevantes para a avaliação da similaridade global e local, além da definição dos pesos para o cálculo da similaridade global.

O levantamento de dados foi realizado através de técnicas de elicitación do conhecimento (PRESSMAN, 1995; BRASIL, 1994; SOMERVILLE, 2003; BRASIL, 2008) com a especialista e estudo de casos relacionados ao tema, onde as principais estratégias adotadas foram as seguintes:

- *Brainstorming* – consiste em uma técnica de geração de idéias realizada através de reuniões para que os participantes sugiram e explorem idéias (PRESSMAN, 1995). A técnica foi empregada para que o conhecimento do domínio fosse discutido amplamente fazendo que o especialista orientasse sobre os principais pontos que devem ser considerados na delimitação do problema.
- *Questionários* – essa abordagem é bastante útil, pois permite obter o conhecimento de forma direcionada, prática e concisa (PRESSMAN, 1995; BRASIL, 2008). Os Anexos 8 e 9 exemplificam os tipos de formulários adotados, os quais abordam questões de múltipla escolha, lista de verificação e questões descritivas no intuito de agrupar a informação por tópicos específicos. Os questionários foram

desenvolvidos baseados em listas de importância para que o especialista indique os pesos dos índices.

- *Entrevistas* – foram traçados roteiros conforme itens de conhecimento que estavam sendo analisados para obter a informação de forma eficiente.
- *Prototipagem* – essa técnica permitiu acompanhar aspectos críticos dos requisitos, agilizando o desenvolvimento e minimizando os riscos da construção do sistema.

Essa etapa gerou como artefatos a documentação formal dos requisitos do sistema e definição do escopo da pesquisa quanto ao uso da técnica RBC.

3.4.2 DEFINIÇÃO DO PROBLEMA E REPRESENTAÇÃO DOS CASOS COM RBC

As técnicas de RBC são adotadas no escopo desse trabalho para indicar a categoria BI-RADS (ACR, 2005), conforme análise de casos de calcificação mamária. Dentre as diversas formas de representação dos casos (WANGENHEIM, 2003; SILVA, 2005), optou-se nessa pesquisa pela representação atributo-valor (WANGENHEIM, 2003) através de entidades de modelo relacionamento na forma de uma tabela de banco de dados relacional (CODD, 1970; CHEN, 1970). Portanto cada caso é representado por um registro na tabela e cada campo da tabela representa uma característica do caso. A escolha da representação dos casos como atributo-valor está relacionada a facilidade e simplificação da implementação, além de ser uma estratégia aprimorada pela adoção do banco de dados relacional viabilizando uma solução prática e eficiente para armazenamento e recuperação.

Segundo Lorenzi (1998), a representação dos casos pode ser feita utilizando dados diretamente da estrutura onde se encontram ou gerando uma segunda estrutura contendo apenas os dados relevantes a composição dos casos. Na presente proposta os dados são extraídos da base de dados através da interpretação da informação contida no domínio, conforme métodos estabelecidos pelo especialista, em concordância com o padrão *BI-RADS* (ACR, 2005). Esses métodos de interpretação foram traduzidos em um algoritmo que gera a base de casos correspondente ao domínio de forma automática. No algoritmo de geração de base de casos cada instância do caso e os valores dos atributos são computados através de uma estratégia que realiza uma análise combinatória (ALMEIDA, MARTINHO,

2003) entre os atributos apontados pelo BI-RADS (ACR, 2005) para atribuir categoria final de risco. O conjunto de atributos empregados foi: a morfologia das calcificações, modificadores de distribuição das calcificações, padrão de densidade mamária, estabilidade das calcificações e achados associados, previamente descritos no Capítulo 2.

Nessa pesquisa foram adotados os 15 tipos morfológicos descritos BI-RADS, 8 tipos de distribuição, 11 tipos de formato das calcificações, 6 tipos de tamanho, 4 tipos de padrão de densidade mamária e 8 tipos de achados associados, resultando numa amostra de 253.440 casos. Entretanto foi necessário filtrar dados inconsistentes para garantir a qualidade de base. Com o auxílio da especialista foram aplicados filtros que excluíram casos com informação duvidosa, resultando numa amostra final de 78.336 casos. Além dessa amostra, foram utilizados 50 casos advindos diretamente da estrutura, os quais também foram tratados para serem moldados apenas com as características relevantes no caso. Os casos advindos diretamente da estrutura foram empregados na validação do sistema. A Figura 27 apresenta uma visão geral de como os casos são representados nesse trabalho.

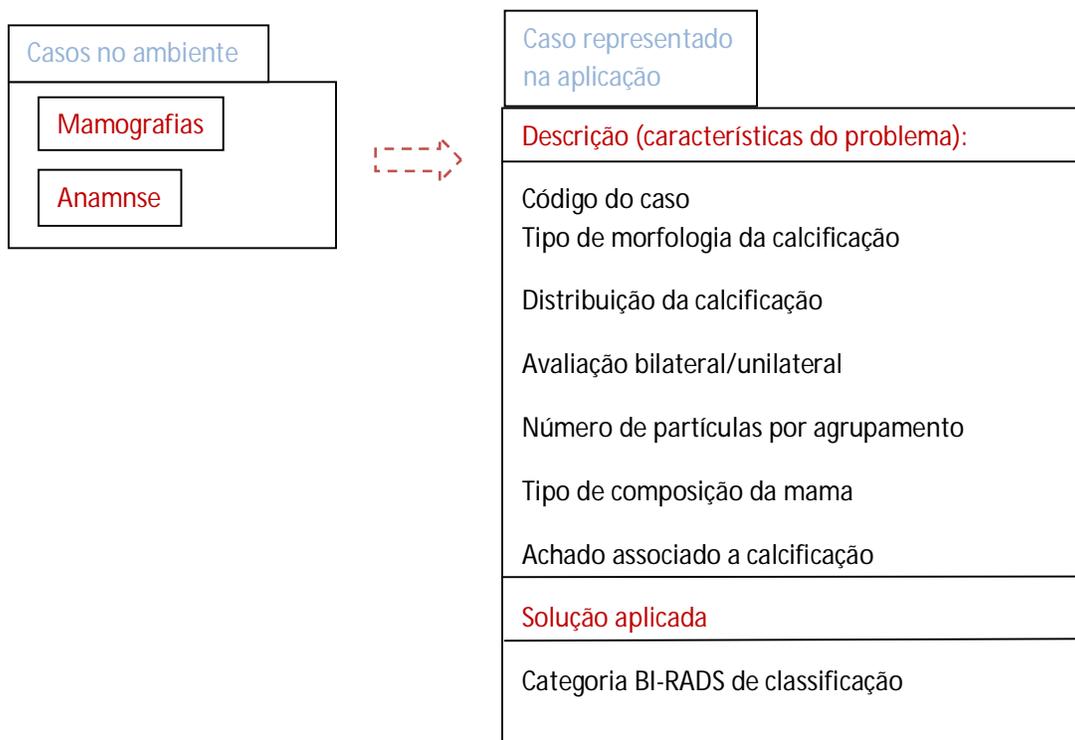


Figura 27: Estrutura de representação dos casos.

3.4.3 ATRIBUIÇÃO DOS PESOS ÀS CARACTERÍSTICAS DOS CASOS

Durante a indexação torna-se necessário atribuir pesos as características dos casos para viabilizar a recuperação dos casos através do algoritmo de similaridade. Na presente proposta cada característica dos casos recebeu um peso conforme o grau de importância para classificação BI-RADS. Conforme o protocolo BI-RADS, estratégias de levantamento de requisitos com a especialista e revisão da literatura alguns dos itens de maior relevância na avaliação das calcificações mamárias são:

- *A morfologia* - conforme descrito no Capítulo 2, as calcificações mamárias podem estar inseridas em um dos 19 tipos de morfologia descritos no protocolo BI-RADS. De acordo com o tipo de morfologia geralmente as calcificações serão subdivididas em 3 níveis: tipicamente benignas, intermediárias ou altamente suspeitas.
- *A distribuição* - conforme descrito no capítulo 2, os modificadores de distribuição indicam a disposição das calcificações na mama. O BI-RADS indica 6 tipos de distribuição das calcificações na mama que são avaliados quanto ao grau de suspeição das calcificações.
- *Avaliação bilateral ou unilateral* - em geral é analisado se o agrupamento de calcificações apresenta a mesma distribuição e a mesma morfologia em ambas as mamas. Os casos onde as características são unilaterais geralmente são mais associados a suspeita de malignidade.
- *O padrão de composição mamária* - conforme descrito no capítulo 2, o padrão de composição mamária merece ser considerado porque de acordo com o tipo de mama a mamografia pode apresentar uma sensibilidade menor de detecção.
- *A estabilidade das calcificações* – uma das formas de avaliar a estabilidade das calcificações mamárias é através de exames anteriores, onde é analisado se houve alteração no número de partículas, distribuição, morfologia ou achados associados.
- *O número de partículas em agrupamentos de calcificações* – conforme descrito no capítulo 2, acima de 5 partículas há um grau de suspeição maior.

- *Achados associados* – são outras estruturas que quando presentes junto com as calcificações poderão aumentar o grau de suspeição. Em determinadas situações os achados associados recebem um grau de importância prioritário em relação às calcificações para a avaliação final. São exemplos de achados associados os nódulos, a assimetria focal, a distorção arquitetural, etc.

Desta forma, os pesos das características dos casos foram definidos atribuindo 70% do peso entre a morfologia e a distribuição das calcificações, além de considerar se a avaliação é bilateral e unilateral. Os 30% restantes foram divididos entre o padrão de composição mamário, avaliação da quantidade de partículas presentes no agrupamento de calcificação e achados associados. Como o foco desse trabalho é avaliar as calcificações buscou-se utilizar associações com achados onde as calcificações tinham maior importância na avaliação final da mamografia. A estabilidade das calcificações mamárias não foi considerada, pois as mamografias foram avaliadas no escopo dessa pesquisa considerando a hipótese de inexistência de exames prévios.

3.4.4 MODELAGEM DA SOLUÇÃO PROPOSTA

A modelagem de diagramas descrita em (GUEDES, 2009) do *Unified Modeling Language* (UML) foi adotada para documentar os requisitos do *software* e detalhar os componentes do sistema. A solução final (desenvolvimento da ferramenta) foi gerada empregando a linguagem de *software* PHP (ZANDSTRA, 2006).

O banco de dados relacional escolhido para trabalhar com a base de casos foi o *PostgreSQL* versão 9.0. A base de casos foi gerenciada usando o sistema gerenciador de banco de dados relacional *pgAdmin* e a linguagem de acesso a manipulação dos dados *Structured Query Language* (SQL). Uma das vantagens do uso da SQL é que ela acessa os dados de forma transparente para o usuário (OSBORNE e MORTON, 2010).

Representação dos casos – conforme explanado no item 3.4.2, a representação dos casos neste trabalho é representada como um conjunto de *pares atributo-valor*.

Indexação (seleção dos índices) - a determinação dos atributos que devem ser comparados para avaliar a similaridade dos casos de entrada e da base de casos foi realizada utilizando a técnica baseada na explicação (LORENZI, 1998; WANGENHEIM,

2003) atendendo critérios estabelecidos pela especialista. Pois apesar da existência de vários métodos automáticos de indexação dos casos, Wangenheim (2003) defende que a melhor estratégia ainda é a escolha dos índices manual, conforme adotado nesse trabalho.

Recuperação dos casos - é feita de forma sequencial, onde a medida de similaridade é calculada sequencialmente para todos os casos na base para determinar os n casos mais similares.

Em seguida verifica-se a relação de preferência específica para a situação, através das medidas de similaridade do sistema, sucessivamente, a cada um dos casos da base. Todos os casos pertencentes à base são então ordenados de acordo com o resultado da função de similaridade. O conjunto de similaridades locais é calculado para cada atributo dos casos através dos critérios do padrão BI-RADS aplicados em uma *função escada*. A forma de busca na base para encontrar casos passados com similaridade global com o novo caso é realizada usando o algoritmo *vizinho mais próximo mais Próximo Ponderado Normalizado* em conjunto com métricas de *medida de contraste*.

Adaptação – neste trabalho optou-se pela adaptação nula, a qual é geralmente empregada em situações de problemas complexos, mas com soluções simples. Conforme discutido previamente, o raciocínio tácito envolvido na distinção de calcificações benignas das suspeitas de malignidade é um processo complexo, entretanto a resolução aplicada no escopo da presente proposta é consideravelmente simples: indicar a categoria BI-RADS mais aplicável.

Validação do sistema RBC – a avaliação da solução foi realizada com testes com a especialista, além disso, foi adotada a técnica de validação preditiva, a qual segundo Silva (2005) necessita de casos históricos para testes com resultados conhecidos ou obtidos diretamente do especialista.

Retenção – o armazenamento de novos casos é realizado manualmente, ou seja, haverá a necessidade de intervenção humana para cadastrar novos casos no sistema.

A Figura 28 apresenta uma síntese da estrutura de raciocínio adotada no escopo dos componentes funcionais do RBC que são utilizados nesse trabalho.



Figura 28: Estrutura de raciocínio empregada na pesquisa (Adaptado de SILVA, 2005).

4 RESULTADOS

4.1 PRÉ-PROCESSAMENTO DE ROIS

O processo de identificação e extração de ROIs foi realizado manualmente com o auxílio de um algoritmo desenvolvido para permitir que a especialista indique a ROI antes da segmentação. A Figura 29 apresenta uma visão geral do processo de extração das ROIs da mamografia.

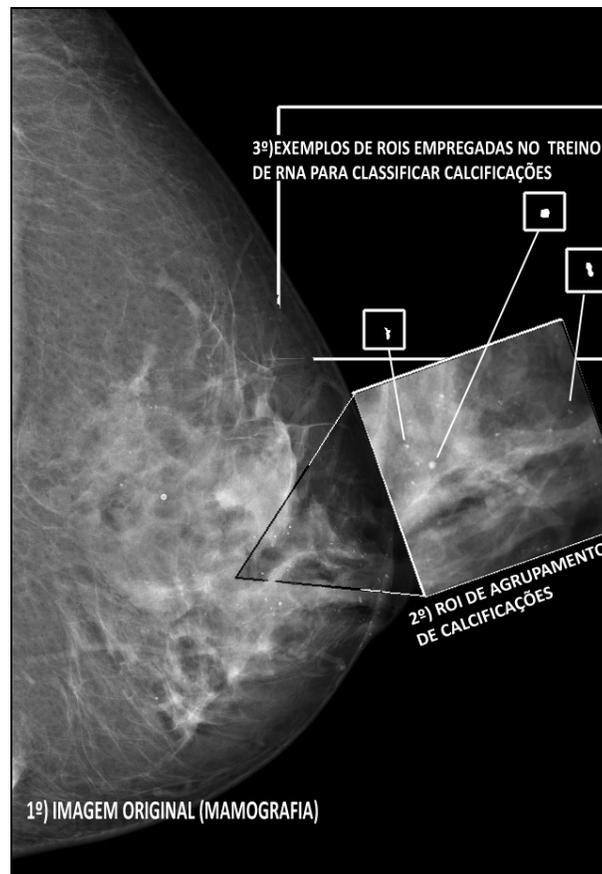


Figura 29: Procedimento para extração de ROIs (Adaptado de imagem original concedida pela CJLR).

Conforme explanado nos itens 3.3.2 e 3.3.3 foram utilizadas ROIs sintéticas para gerar bases de teste controlado. Com as ROIs sintéticas foi possível ampliar o número de amostras de calcificações para avaliar o desempenho da RNA PMC como classificador de

padrões na metodologia proposta. A Figura 30 exemplifica amostras de ROIs sintéticas comparadas com as calcificações reais que elas buscam representar.

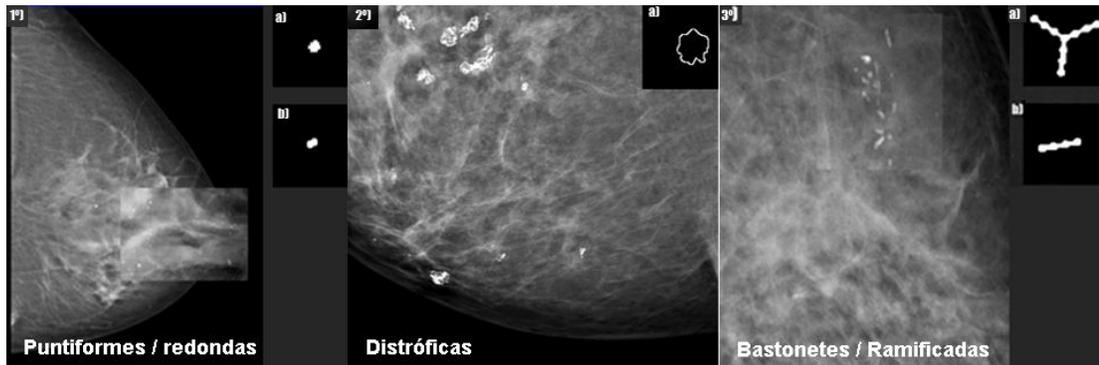


Figura 30: Exemplo de ROIs de calcificações reais adotadas no estudo em comparação com calcificações fictícias geradas para simular as reais (Adaptado de imagens originais concedidas pela CJLR).

De acordo com os exemplos apresentados é possível observar que há uma grande similaridade entre as calcificações reais e as sintéticas. As calcificações sintéticas foram geradas buscando representar inclusive as irregularidades nas bordas, formatos e tipo de disposição mais freqüente das calcificações reais. Em suma, o algoritmo de geração das ROIs sintéticas gera como forma inicial um círculo variando o tamanho da forma, quantidade e amplitude de oscilações das bordas de acordo com o tipo morfológico que seria representado.

Conforme discutido no Capítulo 2, as calcificações reais *Redondas* e *Puntiformes* têm como maior distinção o tamanho (as *Puntiformes* são muito menores que as *Redondas*), mas basicamente o formato delas é igual (redondo ou oval), portanto o critério para geração das ROIs sintéticas que emulam esse tipo de calcificação gerou os 2 tipos morfológicos sintéticos variando apenas a escala de tamanho. No caso das calcificações em *formato de Pipoca* foi introduzido um maior nível de oscilações para não gerar uma circunferência ou parábola perfeita, dando a idéia de deformidades similar ao que pode ser visto nas calcificações reais. Já as calcificações sintéticas que representam o *formato de Bastonetes*, são geradas com uma seqüência das calcificações sintéticas *Puntiformes* aglutinadas em linha. A Figura 31 apresenta um exemplo comparando as ROIs sintéticas com as ROIs reais extraídas de agrupamentos de calcificações detectados nas mamografias.

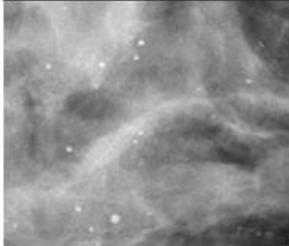
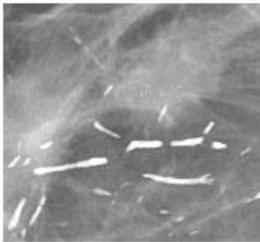
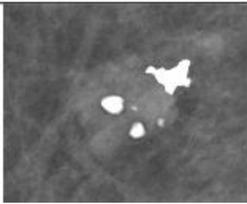
ROI de calcificações agrupadas por morfologia	ROI individual real extraída do agrupamento	ROI individual real extraída do agrupamento binarizada	ROI individual sintética gerada
	Morfologia Puntiforme		
			
	Morfologia Redonda		
			
	Morfologia em Forma de bastonetes		
			
	Morfologia em Formato de pipoca		
			

Figura 31: Exemplo de ROIs de calcificações extraídas de mamografias em paralelo com as calcificações fictícias geradas.

Antes da etapa de extração de características morfológicas, descrita no Capítulo 3.3.2, o tamanho de cada ROI foi ajustado para 300X300 *pixels*. A uniformização de tamanho das ROIs foi realizada para padronizar o tamanho evitando conflitos na análise e classificação dos padrões.

4.2 ANÁLISE DE ATRIBUTOS MORFOLÓGICOS EXTRAÍDOS DAS ROIS

O conjunto de características morfológicas, mencionadas no item 3.2.2, é utilizado para distinguir e caracterizar calcificações de acordo com o tipo morfológico analisado. As características extraídas são suficientemente versáteis para serem empregadas em testes distintos. As características de perímetro e área, por exemplo, são úteis para discriminar calcificações onde o principal diferencial é o tamanho. A variância é útil na distinção de aglomerados de calcificações benignos e malignos. Enquanto que as características de excentricidade, compacidade e circularidade são usadas em conjunto com a área e perímetro para classificar tipos de calcificações segundo a morfologia predominante.

As Figuras 32 e 33 ilustram, respectivamente, a distribuição dos parâmetros extraídos (área e perímetro) das ROIs sintéticas e reais em um plano cartesiano, onde é possível observar que com apenas esses 2 parâmetros é possível fazer uma distinção entre as duas classes de classificação (ROIs de calcificações Redondas e Puntiformes), pois os parâmetros são linearmente separáveis. Por essa razão a entrada da RNA para distinguir calcificações redondas de puntiformes necessitou conter apenas 2 neurônios. O fato dos atributos extraídos das ROIs sintéticas, utilizados no treinamento da RNA, estarem distribuídos de forma distinta maximiza a probabilidade da RNA PMC apresentar um bom desempenho na classificação dos exemplos não vistos durante o teste com as ROIs de calcificações reais, apresentados na Figura 33.

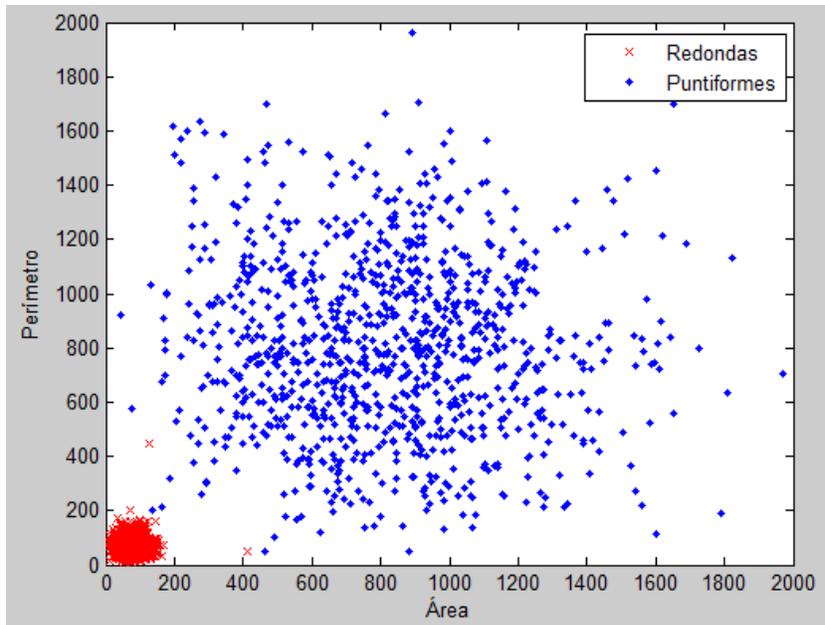


Figura 32: Distribuição de classes por parâmetros extraídos das ROIs de calcificações sintéticas.

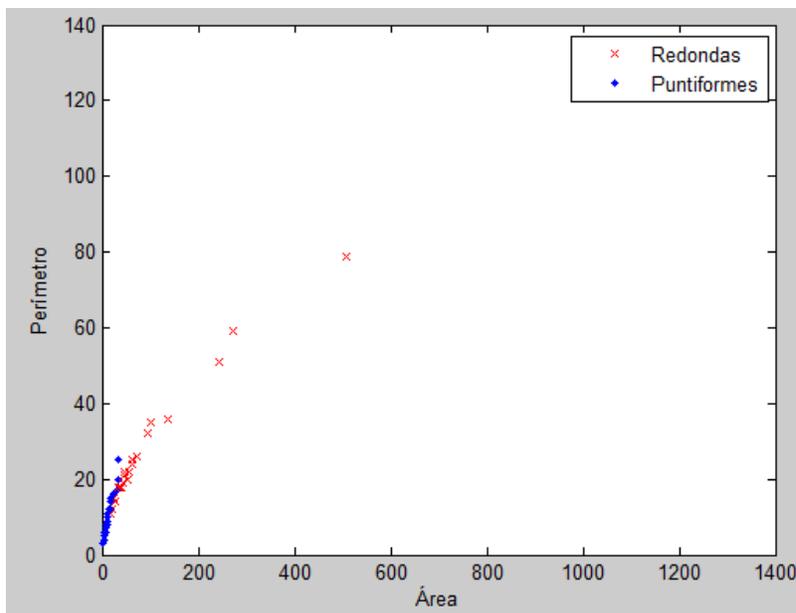


Figura 33: Distribuição de classes por parâmetros extraídos das ROIs de calcificações reais.

A Figura 34 apresenta distribuição das 4 classes de calcificações analisadas nessa pesquisa em função dos parâmetros área e perímetro. Nesse caso ocorre uma sobreposição entre as classes, logo a utilização dos demais parâmetros extraídos das ROIS (previamente descritos no item 3.2.2) é uma solução para ampliar a capacidade de aprendizado e generalização da RNA.

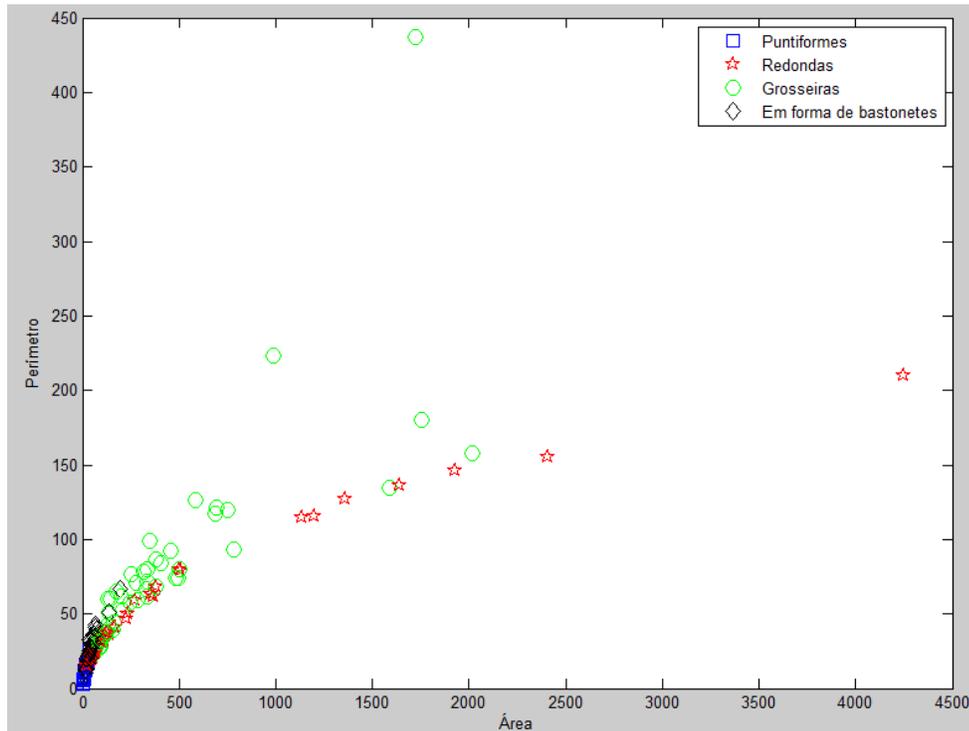


Figura 34: Distribuição de classes calcificação em função dos parâmetros área e perímetro.

4.3 TREINAMENTO DE RNA PMC PARA CLASSIFICAR CALCIFICAÇÕES MAMÁRIAS

Para demonstrar o comportamento da rede com uma entrada e saída bi-dimensional será demonstrado a seguir o funcionamento da RNA PMC, ilustrada na Figura 35, na distinção das calcificações do tipo *Puntiforme* de calcificações do tipo *Redonda*.

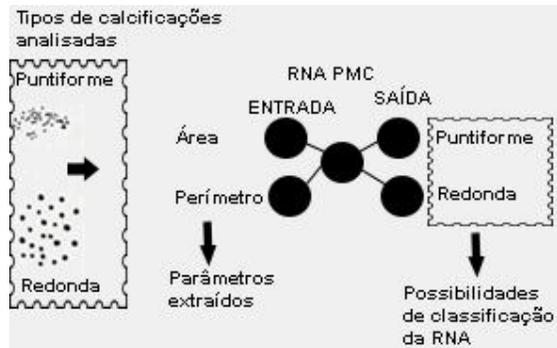


Figura 35: Arquitetura da RNA PMC adotada.

A RNA PMC recebeu como entrada a matriz de parâmetros extraídos das ROIs sintéticas (área e perímetro) e o vetor de saída com o valor estimado para cada entrada (Exemplo: 1 – Puntiforme ou 0 - Redonda). A função *newff* do Matlab, utilizada na implementação da RNA PMC, realiza uma normalização das entradas necessitando que sejam informados os limites máximos e mínimos dos padrões de entrada. A Figura 36 apresenta a distribuição dos parâmetros extraídos de uma amostra contendo 400 ROIs sintéticas que simulam calcificações *Puntiformes* e *Redondas*. Conforme análise da distribuição dos parâmetros é possível mensurar os limites para aplicar na função de treino da RNA PMC adotada.

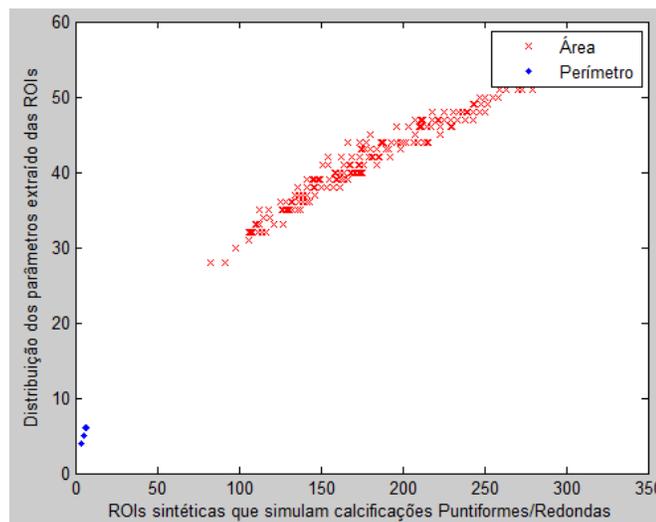


Figura 36: Distribuição dos parâmetros extraídos de ROIs sintéticas utilizadas no treinamento da RNA.

A RNA PMC é então treinada com o algoritmo *Backpropagation* e configurada com 2 neurônios na entrada correspondentes a quantidade de parâmetros morfológicos extraídos das ROIs, 1 neurônio na camada oculta e 2 neurônios na saída para discriminar as 2 classes de classificação (*Redonda* ou *Puntiforme*). Das 400 amostras metade é utilizada para treino, metade para teste e a validação da RNA é feita com outra amostra cujos parâmetros morfológicos foram extraídos de ROIs geradas a partir de calcificações reais. No exemplo a seguir será apresentado o desempenho da RNA PMC adotada usando a função de transferência Tangente Sigmóide. A Figura 37 demonstra o decaimento do EMQ da RNA PMC no decorrer do treinamento.

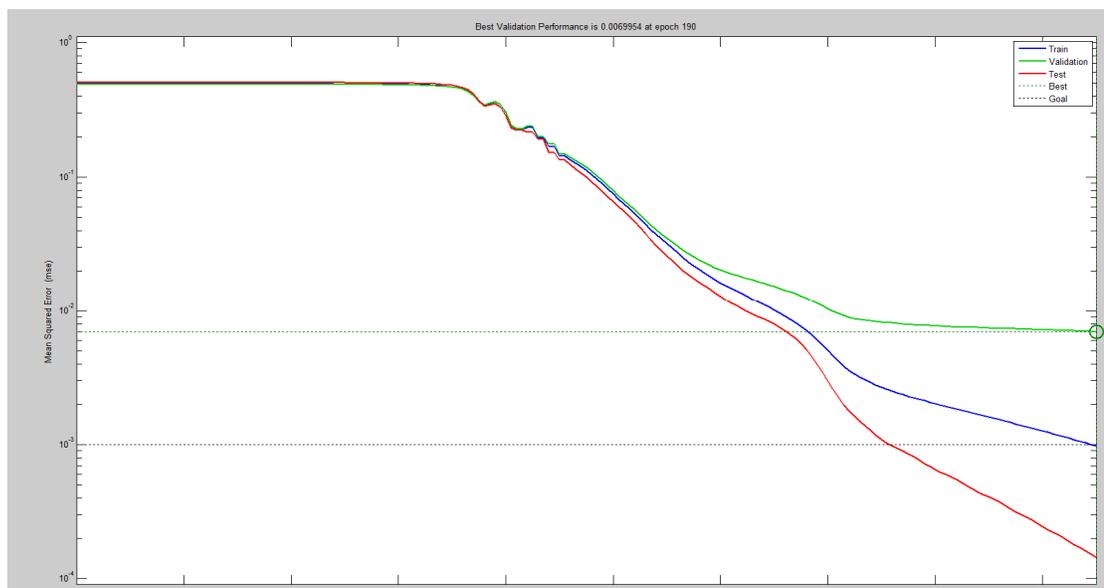


Figura 37: Variação do EMQ da RNA PMC empregada para distinguir calcificações redondas de puntiformes.

A Figura 38 apresenta a distribuição de classe do arquivo utilizado para teste da RNA treinada com as ROIs sintéticas.

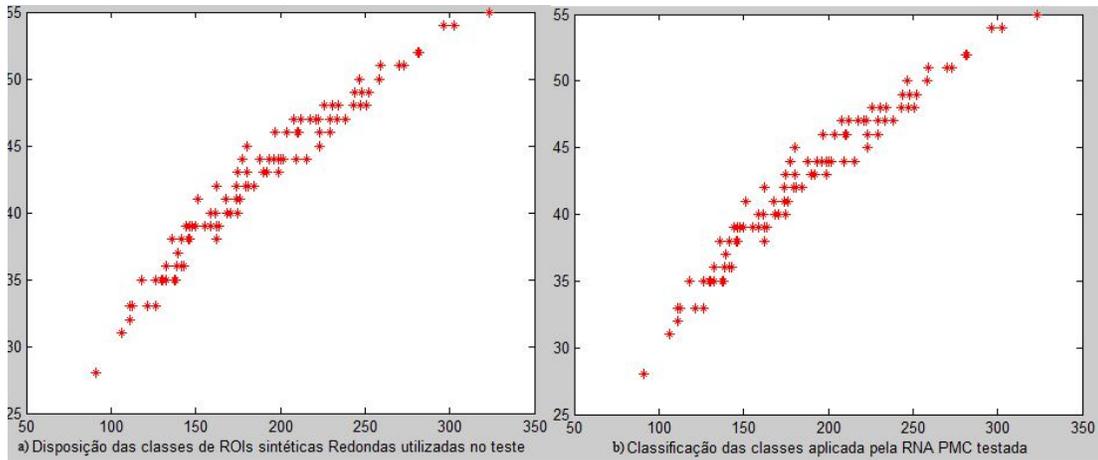


Figura 38: Disposição das classes no arquivo de treinamento utilizado como teste da RNA PMC treinada.

A Figura 39 apresenta o desempenho da RNA PMC treinada na classificação das ROIs reais. Neste exemplo foram adotadas 9 amostras de ROIs de calcificações Redondas reais. Ainda, nesse exemplo, conforme pode ser observado em ambos os casos (teste com ROIs sintéticas e validação com ROIs reais) a rede conseguiu classificar o padrão obtendo 100% de acerto.

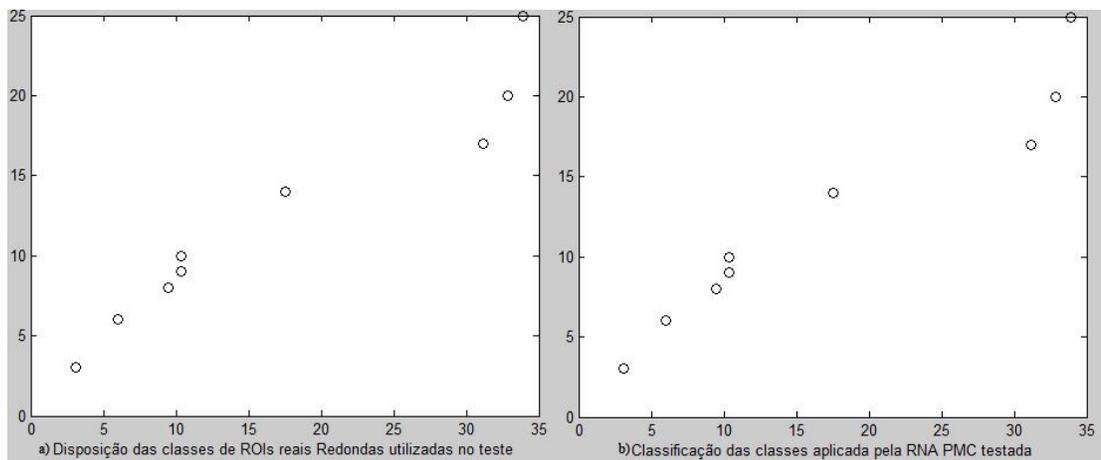


Figura 39: Desempenho da RNA PMC testada durante a classificação de um arquivo teste com as ROIs reais.

4.4 FUNCIONAMENTO DO SISTEMA RBC

Conforme explanado no item 3.3.2 a representação dos casos neste trabalho é realizada com um conjunto de pares atributo-valor. A Figura 40 apresenta o Diagrama de Entidade-Relacionamento (RAMEZ, SHAM, GUIMARÃES; 2005) do sistema RBC desenvolvido.

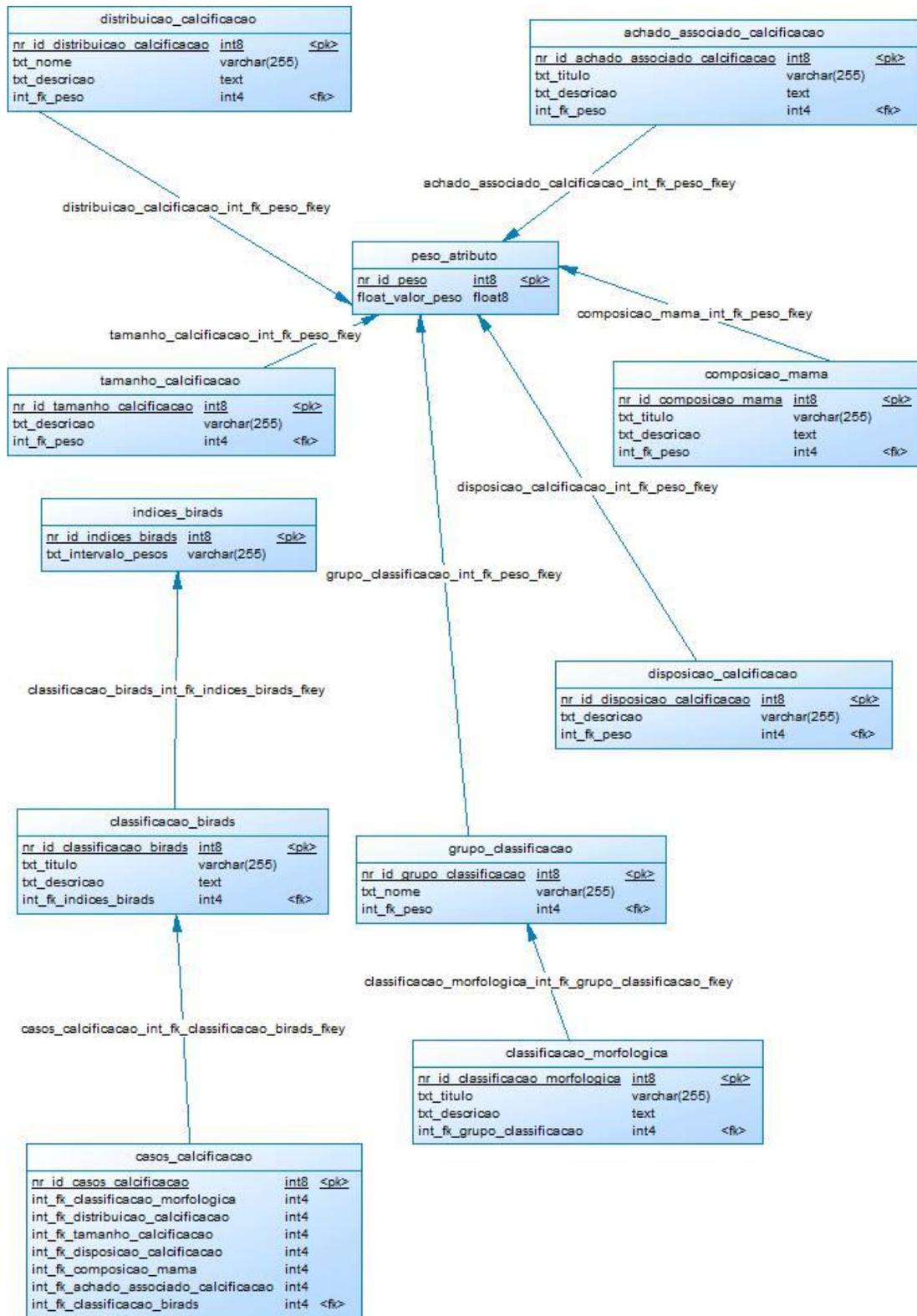


Figura 40: Diagrama Entidade Relacionamento da aplicação.

A comparação de um caso consultado com os casos registrados na base do Sistema de Avaliação de Calcificações Mamárias (SACM) é realizada através da comparação das características do novo caso com as características dos casos registrados no banco de dados do sistema usando o *Algoritmo do Vizinho mais próximo*. O algoritmo é executado seguindo os seguintes passos:

- Considerando que a base de casos dos SACM está montada contendo X casos.
- É realizada uma consulta de caso C no SACM. O sistema realiza uma busca por *medida de contraste* e retorna um conjunto de casos onde é aplicado o *Algoritmo do Vizinho mais próximo* para obter o(s) caso(s) com maior similaridade com C . A Tabela 2 apresenta um conjunto de casos de exemplos que serão utilizados para explicar a lógica de funcionamento do sistema durante a busca.

Tabela 2 – Comparação de caso de consulta com casos da base do SACM.

Característica	Peso	Caso 1	Caso 2	Consulta de caso
Morfologia	0.35	0	0	0
Distribuição	0.35	0.8	0.8	0.8
Avaliação bilateral/unilateral	0.02	0.5	0.8	1
Número de partículas do agrupamento	0.03	0.2	0.3	0
Composição da mama	0.05	0.1	0.1	0.1
Achado associado	0.20	0	0	0

Utilizando como exemplo os casos da Tabela 2 e o caso de consulta C , os cálculos realizados para recuperar o(s) caso(s) com maior similaridade com o problema são os seguintes.

X =Caso 1;

C =Consulta de caso;

$$\text{Distância}(x,c)=(0.35 * |0-0|^2)+(0.35 * |0.8-0.8|^2)+(0.02 * |1-0.5|^2)+(0.03 * |0-0.2|^2)+(0.05 * |0.01-0.01|^2)+(0.20 * |0-0|^2) 1/2;$$

$$\text{Distância}(x,c) =(0.35 * 0)+(0.35 * 0.25)+(0.02 * 1)+(0.03 * 0.04)+(0.05 * 0)+(0.20 * 0) 1/2;$$

$$\text{Distância}(x,c)\approx 0.33;$$

X=Caso 2;

C=Consulta de caso;

$$\text{Distância}(x,c)=(0.35 * |0-0|^2)+(0.35 * |0.8-0.8|^2)+(0.02 * |1-0.8|^2)+(0.03 * |0-0.3|^2)+(0.05 * |0.01-0.01|^2)+(0.20 * |0-0|^2)1/2;$$

$$\text{Distância}(x,c)= (0.35 * 0)+(0.35 * 0)+(0.02 * 0.04)+(0.03 * 0.09)+(0.05 * 0)+(0.20 * 0) 1/2;$$

$$\text{Distância}(x,c)\approx 0.06;$$

De acordo com os cálculos realizados e em conformidade com Abel (1996), o caso mais parecido com o novo problema é aquele que apresenta a menor distância. Portanto, no exemplo apresentado o caso que possui a maior similaridade com o caso de consulta é o Caso 2. A seguir contam as *interfaces* do sistema durante as etapas para a recuperação de casos usando o SACM.

- I. Inicialmente é acessada a opção de “Consulta de caso” do SACM, conforme apresentado na Figura 41.

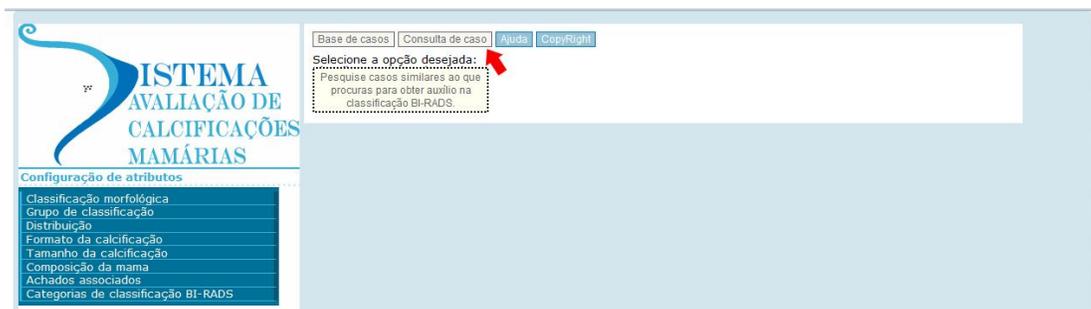


Figura 41: Tela de acesso a consulta de casos no SACM.

II. Em seguida, é necessário preencher as características do caso, conforme ilustrado pela Figura 42.

Figura 42: Tela de consulta de caso na base do sistema SACM.

III. Com as informações do caso de consulta preenchidas clica-se na opção “Consulta” e o sistema realiza uma busca do(s) caso(s) com maior similaridade para apresentar na tela qual a classificação BI-RADS mais indicada de acordo com a solução de casos similares presentes na base de dados do sistema SACM anteriores. A Figura 43 apresenta o resultado de uma consulta realizada.

Verifique a classificação BI-RADS indicada (caso não concorde altere).

Base de casos | Consulta de caso | Ajuda | Copyright

Consulta de caso (indicação da classificação BI-RADS):
 Seleccione o tipo de morfologia predominante das calcificações:

<input type="radio"/> Cutânea	<input type="radio"/> Vasculares	<input type="radio"/> Grosseiras ou...	<input type="radio"/> Semelhantes a bastonetes
<input type="radio"/> Redondas	<input type="radio"/> Puntiformes	<input type="radio"/> Centro radiotransparente	<input type="radio"/> Distróficas
<input type="radio"/> Amorfas ou indistintas	<input type="radio"/> Heterogêneas grosseiras	<input type="radio"/> Pleomórficas finas	<input type="radio"/> Finas lineares ou...
<input type="radio"/> Casca de ovo ou em anel	<input type="radio"/> Fios de sutura	<input type="radio"/> Leite de cálcio	

Seleccione a disposição da(s) calcificação(es):
 Isolada Bilateral Unilateral

Seleccione a opção correspondente ao número de partículas do agrupamento de calcificação(es):
 Maior ou igual a 5 Menor que 5

Seleccione o tipo de distribuição da calcificação:
 Seleccione o tipo de distribuição da calcificação:

Seleccione a composição da mama:

Mama lipossustituída (tecido glandular < 25%)
 Tecido glandular entre 25% e 50%
 Tecido glandular aproximadamente entre 51% e 75%
 Tecido glandular > 75%

Seleccione os achados associados as calcificações:
 Nódulo sólido circunscrito não calcificado
 Assimetria focal
 Linfo nodos intramamários
 Nenhum

Classificação BI-RADS indicada pelo método vizinho mais próximo:
 Categoria 4 Categoria 4

Consultar Apagar dados

Figura 43: Tela apresentando sugestão de classificação BI-RADS conforme consulta por similaridade entre as características do caso de consulta e os casos presentes na base do SACM.

5 DISCUSSÃO E CONCLUSÃO

Em trabalhos anteriores foi constatada a relevância do estudo das calcificações mamárias na detecção precoce do câncer de mama. Adicionalmente a adoção do sistema BI-RADS na classificação das mamografias já está consagrado como um dos métodos mais aplicados na padronização dos relatórios mamográficos. Não consta na literatura uma abordagem que utilize o conjunto de técnicas abordadas por essa pesquisa (um sistema híbrido de RNA PMC e RBC) para auxiliar na classificação BI-RADS final de mamografias pela análise morfológica de calcificações mamárias.

Tanto as RNAs quanto o RBC são técnicas de IA consagradas na aquisição e utilização da informação para gerar soluções de forma similar ao raciocínio humano em campos especializados. A vantagem de empregar um sistema híbrido, embora em tarefas distintas, reside em ampliar e complementar o potencial de cada abordagem. Se por um lado as RNAs são ferramentas poderosas na aquisição do conhecimento, elas apresentam deficiências na explicação da solução proposta (o conhecimento das RNAs pode ser visto como uma caixa preta mapeada nos pesos entre as conexões dos neurônios). Por outro lado, o RBC além de mapear o conhecimento por meio de sua base de casos, é capaz de apresentar o motivo que o levou a tomar uma decisão.

Para o desenvolvimento do sistema RBC, embora existam soluções consistentes no mercado, as opções existentes geralmente possuem restrições quanto a uma *interface* amigável ou que se encaixe com o interesse dessa abordagem. Portanto, optou-se pela elaboração de um sistema RBC particular moldado a essa pesquisa.

Quanto a decisão de quais características deveriam ser adotadas para caracterizar computacionalmente as calcificações previamente detectadas nas mamografias, foi realizado um levantamento de requisitos junto com a especialista da área, estudo de casos relacionados e de pesquisas na área. Como a presente proposta realiza uma análise por imagens das características selecionadas para extração de parâmetros das ROIs empregados na RNA PMC e representação de casos no RBC, foram desconsiderados muitos parâmetros referentes a fatores de risco presentes na anamnese, que não são conclusivos para caracterizar os achados. Foram adotados, portanto, os parâmetros referenciados no padrão BI-RADS que possuem maior significância na análise de calcificações por mamografia.

Uma das problemáticas enfrentadas no desenvolvimento do sistema de avaliação de calcificações mamárias é que, em geral, o protocolo BI-RADS descreve que o médico radiologista deve descrever no seu laudo de avaliação da mamografia os achados de maior suspeição ou os que possam gerar dúvidas na interpretação do laudo por outros profissionais. Entretanto, a presente proposta realiza uma avaliação de calcificações mamárias independente do nível de suspeição. Todavia, houve dificuldade na coleta de informações das situações em que a aplicação estava analisando calcificações tipicamente benignas, as quais geralmente não são referenciadas nos laudos em virtude do seu baixo grau de suspeição. Para abranger essas situações e criar uma base suficientemente ampla, abordando todos os casos possíveis, foi utilizada a técnica de análise combinatória baseada nos índices analisados para gerar a base de casos do RBC. Em seguida foi atribuída a classificação BI-RADS segundo os critérios estabelecidos pela especialista e pelo padrão BI-RADS.

A presente proposta disponibiliza um sistema para auxiliar na avaliação de calcificações mamárias. A metodologia empregada fundamenta-se em técnicas de processamento digital de imagens para realizar a extração de informações morfológicas das ROIs e uma RNA PMC para classificar as calcificações analisadas de acordo com a classificação morfológica definida pelo BI-RADS. O desempenho da RNA PMC foi avaliado em um teste controlado com ROIs sintéticas geradas para simular calcificações reais e validado com calcificações reais extraídas de mamografias cedidas pela clínica de estudo. Adicionalmente foi desenvolvido um sistema RBC que atua em conjunto com a RNA PMC para indicar a classificação BI-RADS final da mamografia conforme análise das calcificações mamárias. Com a metodologia aplicada atende os objetivos propostos por esse trabalho. Os resultados obtidos validam que a presente proposta disponibiliza um processo consistente de análise das calcificações mamárias através do uso das técnicas de IA e PDI, além de realizar um referencial teórico substancial que pode servir como fonte de consulta para os interessados no estudo do tema em questão.

6 TRABALHOS FUTUROS

Como trabalhos futuros sugerem-se:

- *Automatização da etapa de segmentação de ROIs* – nesse intuito foram testadas e implementas possíveis soluções para reduzir ruídos das imagens de mamografia digital, realçar/evidenciar ROIs, viabilizar a segmentação da imagem da mamografia digital para finalmente efetuar a classificação referente a sugestão de laudo para a ROI. Uma das alternativas exploradas para detectar automaticamente e destacar as regiões de calcificações das mamografias analisadas foi a aplicação de testes utilizando a TW, visto que após a decomposição da imagem em faixas de frequência é possível observar que geralmente as calcificações concentram-se em uma determinada faixa (LASHKARI, 2010). O algoritmo de TW aborda a decomposição de um sinal discreto através de um conjunto de filtros que extraem as características médias (aproximações) do sinal. Em seguida, é somado um conjunto de filtros baseados nas funções Wavelets, responsáveis pela extração de características de alta frequência (detalhes) do sinal (MALLAT, 1999). Na prática, esse processo de decomposição consiste na aplicação de filtros passa-altas e passa-baixas através de múltiplas decomposições. A Figura 44 ilustra como esses conceitos foram adotados nesse experimento.

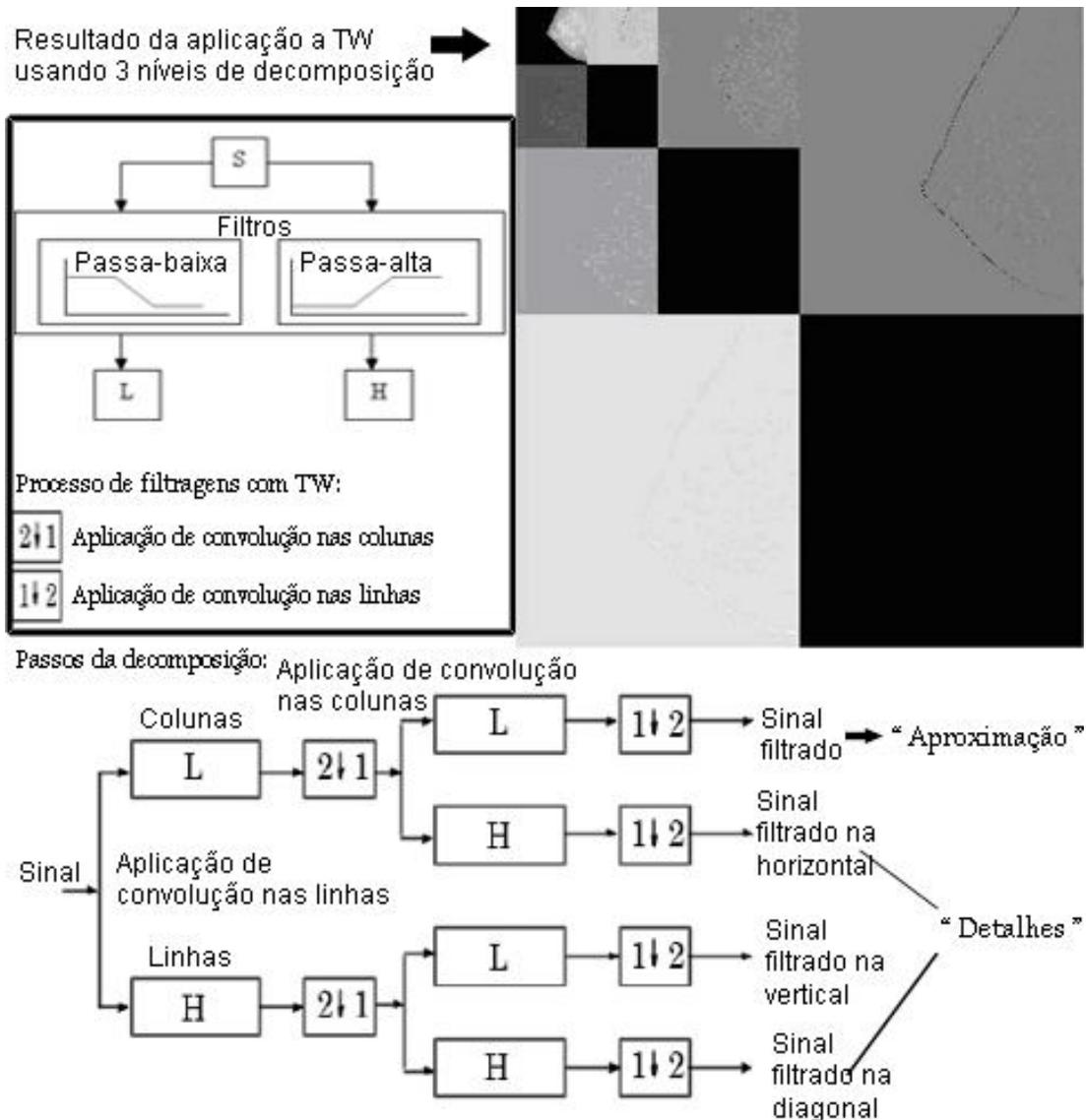


Figura 44: Teste para detectar automaticamente as regiões de calcificação em imagem de mamografia.

Outra possibilidade explorada para evidenciar as ROIs das imagens mamográficas digitais estudadas consiste em um algoritmo que adota o histograma da imagem para dividi-la em faixas de frequências. O resultado desse trabalho pode ser avaliado na Figura 45.



Figura 45: Processo de segmentação de imagem por histograma.

- *Inclusão do parâmetro estabilidade das calcificações* – em futuras evoluções do sistema RBC, pode-se ampliar as técnicas adotadas para incluir esse item, o qual também tem grande relevância na análise de calcificações mamárias.
- *Adaptação de RNA PMC para classificar demais tipos morfológicos de calcificações mamárias* – na presente abordagem inicialmente a RNA PMC desenvolvida é capaz de discriminar 4 tipos de calcificação mamária (Redondas, Puntiformes, Bastonetes e “forma de pipoca”). A metodologia aplicada mostra que é possível ampliar a classificação para os demais tipos.
- *Disponibilização WEB do sistema desenvolvido para consulta acadêmica;*
- *Implantação do sistema na clínica de estudo para avaliação estatística da metodologia proposta in loco.*

Por fim, pode-se adotar a solução aplicada na análise de calcificações mamárias nessa pesquisa no desenvolvimento de um CAD que indique a classificação BI-RADS de outros achados relevantes na avaliação de mamografias.

REFERÊNCIAS BIBLIOGRÁFICAS

ABEL, M. Raciocínio Baseado em Casos. Porto Alegre: Instituto de Informática da UFRGS, 1996.

ABRAMSON, J.H. Making sense of data. A self-Instruction manual on the interpretation of epidemiological data. Oxford University Press, 1994.

ALMEIDA, C.; MARTINHO, H. Processos de resolução de problemas de análise combinatória numa turma do 7º ano. In: COSME A. et. al. Atlas do XIV Seminário de Investigação em Educação Matemática, pp. 505-524. Lisboa: Associação de Professores de Matemática, 2003.

ACR. BI-RADS: Sistema de laudos e registro de dados de imagem da mama. 1. ed. São Paulo: American College of Radiology, 2005.

ANGELO, M. F. Sistema de Processamento de Imagens Mamográficas e Auxílio ao Diagnóstico via-Internet. Tese (Doutorado), Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2007.

AAMODT, A., PLAZA, E. Case Based Reasoning: Foundational Issues, Methodological variations and System Approaches. In: AICOM- Artificial Intelligence Communications, Vol. 7, Ed. IOS Press, EUA, 1994.

ARTERO, A. O. Inteligência Artificial: Teórica e prática. Ed. Livraria da Física, p. 123-132, São Paulo, Brasil, 2008.

AZEVEDO, F. M.; BRASIL, L. M.; OLIVEIRA, R. C. L. Redes Neurais com Aplicações em Controle e em Sistemas Especialistas. 2000.

BIFFI, R. G. A dinâmica familiar de um grupo de mulheres com câncer de mama. Tese de Doutorado, Escola de Enfermagem de Ribeirão Preto. USP, 2003.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDERMIR, T. B. Redes Neurais Artificiais: Teoria e aplicações. ed. LTC - Livros Técnicos e Científicos Editora S.A., Rio de Janeiro, 2000.

BRASIL - Agência Nacional de Saúde Suplementar (ANS). Rol de procedimentos e eventos em Saúde. Resolução Normativa RN nº 167, janeiro de 2008.

BRASIL L. M. Proposta de Arquitetura para Sistema Especialista Híbrido e a Correspondente Metodologia de Aquisição do Conhecimento, Ph.D. Thesis, UFSC, Florianópolis, Brasil, 1999.

BRASIL, L. M. Aquisição de conhecimento aplicada ao diagnóstico de Epilepsia. Florianópolis: [s.n.], 1994.

BRASIL, L. M. Informática em Saúde. Brasília/Londrina: Universa e Edue, 2008.

BRITO, H. R. et. al. Evolução da posição do complexo aréolo-mamilar e da extensão das cicatrizes após mamoplastia redutora com retalho de pedículo inferior areolado Rev. Bras. Cir. Plást., vol. 23(3): p.167-72, 2008.

BOYD N. F. et. al. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev, 7 :1133-44, 1998.

CABRAL, E. P.; DESCHAMPS, F.; STEMMER, M. R. Algumas Aplicações da Transformada Discreta de Fourier no Processamento Digital de Imagens. In Anais do II Simpósio Catarinense de Processamento Digital de Imagens - SCPDI, volume 1, Florianópolis - SC - Brasil, 2002.

CAMPOS, M. C. V. Análise crítica e estimativa dos valores preditivos positivos das calcificações de aspecto mamográfico não benigno. Tese de Doutorado em Ciências apresentada à Faculdade de Medicina da Universidade de São Paulo, São Paulo, 2006.

CENTRO DE CIÊNCIAS TECNOLÓGICAS, Universidade do Estado de Santa Catarina. Disponível em: <<http://www.joinville.udesc.br/processamentodeimagens/deteccao/deteccao.html>>. Acesso em: 18 maio 2011.

CHAKRABARTI, J. et al. Accuracy of mammography in predicting histological extent of carcinoma ductal in situ (DCIS). Eur J Surg Oncol. 32(10): p. 1089-92, 2006.

CHEN, Peter Pin-Shan. The Entity-Relationship Model – Toward a Unified View of Data. Massachusetts Institute of Technology, 1970.

CHENG, H. D. et al. Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition, v.36, p.2967-2991, 2003.

CHENG, H. D. et al. Automated breast cancer detection and classification using ultrasound images: A survey. ACM, Available online 30 May 2009.

CODD, Edgar Frank. A Relational Model of Data for Large Shared Data Banks. 1970.

CHUA e YANG. Cellular neural networks: theory. IEEE Transactions on Systems and Circuits, vol. 35, n. 10, p. 1257-1272, 1988.

COX, P. H. Análise e síntese de um processador digital wavelet. Tese de doutorado, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, São Paulo, Brasil, 2004.

Davis, R., Stacey, A.J. The detection and significance of calcifications in the breast: a radiological and pathological study. The British Journal of Radiology, 49:12-26, 1976.

DINKEL, H.P., GASSEL, A.M., TSCHAMMLER, A. Is the appearance of the microcalcifications useful in predicting histological grade of malignancy in ductal cancer in situ? The British Journal of Radiology; 73:938-944, 2000.

- DIGABEL, H.; LANTUÉJOUL, C. Iterative algorithms. In J.-L. Chermant, ed. Proc. Second European Symp. Quantitative Analysis of Microstructures in Material Science, Biology and Medicine, p. 85-99, Stuttgart, Germany, Riederer Verlag, 1978.
- DUARTE, D. L. A mama em imagens. p. 5-6, Rio de Janeiro, ed. Guanabara Koogan, 2006.
- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. 2. ed. New York: Wiley-Interscience, 2001.
- DUPONT, W. D.; PAGE, D. L. Risk factors for breast câncer in women with proliferative breast disease. New Engl. J. Med 312: 146-151, 1985.
- EKLUND, G. W.; CARDENOSA, G. The art of mammographic positioning, vol.30 p. 21-53, Radiol Clin of North Am, 1992.
- ELPÍDIO, F.G.G. et al; Morphological analysis for feature extraction and classification of breast calcifications. In: Pan American Health Care Exchanges (PAHCE), Miami, Flórida , 2012.
- FEIG, S. A. Mammographic evaluation of calcifications. In: Kopans, D.B. Mendelson, E.B. eds. Syllabus: a categorical course in breast imaging. Oak Broak, III: Radiological Society of North America, p. 93-105, 1995.
- FERNANDES, F. C. Sistema Neuro-Fuzzy para classificação de calcificações em mamogramas. Tese de mestrado da Universidade Católica de Brasília, Distrito Federal, 2007.
- FERNANDES, A. M. R. Inteligência Artificial: noções gerais. Ed. Visual Books, Florianópolis, 2003.
- FREITAS, F. et al; Rotinas em ginecologia. Artmed, v. 6 São Paulo, 2006.
- FRIEDLAND, D.J. et al. Medicina Baseada em Evidências: Uma Estrutura para a Prática Clínica. Guanabara Koogan, 2001.
- SHEN, L.; BAI, L. A review of Gabor wavelets for face recognition, Patt. Anal. Appl. 9: p. 273-292, 2006.
- GUEDES, G. T. A. UML 2: uma abordagem prática. São Paulo: Novatec Editora, 2009.
- GOMES, O. F. M. Processamento e análise de imagens aplicados à caracterização automática de materiais. Dissertação (Mestrado), Departamento de Ciência de Materiais e Metalurgia, Pontifícia Universidade Católica do Rio de Janeiro, Distrito Federal, 2001.
- GONZALEZ, R. C. ; WOODS, R. Digital Image Processing. Upper Saddle River, New Jersey: Prentice-Hall, Edinburgh, 2008.
- GRAY, G. E.; HENDERSON, B. E.; PIKE, M. C. Changing ratio of breast cancer incidence rates with age of blanck females compared with females in the United States. J Natl Cancer Inst 64(3): 461-463, 1980.

GUIMARÃES, J. R. Q. Manual de Oncologia. São Paulo: BBS, 2004.

GUIMARÃES, L. T. G. Subtração digital como ferramenta para detecção de tumores em imagens mamográficas de mamas densas: uma abordagem utilizando simulação computacional. Tese (mestrado) Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.

HALKIOTIS, S.; TAXIARCHIS B.; RANGOUSI M. Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks. Journal Signal Processing Vol. 87 Issue 7, July, Elsevier North-Holland, Inc. Amsterdam, The Netherlands, 2007.

HAMAJIMA et al. Alcohol, tobacco and breast cancer: collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer. 2002;87:1234-45.

HAYKIN, S., Neural Networks: A Comprehensive Foundation, 2nd. Ed., USA, Prentice Hall, 1999.

HECHT-NIELSON, R. Kolmogorov's mapping neural network existence theorem, First IEEE International Conference on Neural Networks, vol. III, p. 11-14, San Diego, CA, 1987.

HENDERSON, B. E.; PIKE, M. C.;CASAGRANDE J. T. Breast cancer and the oestrogen window hypothesis. Lancet ii:363-364, 1981.

HIGHNAM, R.; BRADY, M. Mammographic Image Analysis. Computational Imaging and Vision. 1. ed. Dordrecht, Boston, Londres: Kluwer Academic Publishers, p. 392, 1999.

HOLLAND, R., HENDRIKS, J.H.C.L. Microcalcifications associated with ductal carcinoma in situ: mammographic-pathologic correlation. Seminars in Diagnostic Pathology, 1994.

Hu, M. K., Visual Pattern Recognition by Moments Invariants - IRE Trans. Inf Theory, vol. 8, p. 179-187, 1962.

INCA. Câncer de mama. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/acoos_programas/site/home/nobrasil/programa_controle_cancer_mama/>. Acesso em: 08 de outubro de 2011.

INNES, K. E.; BYERS, T. E. Smoking during pregnancy and breast cancer risk in very young women (United States). Cancer Causes Control;12:179-85, 2001.

JÄHNE, B., HAUSSECKER, H.; GEISSLER, P. Handbook of Computer Vision and Applications - Signal Processing and Pattern Recognition, vol. 2., Academic Press, San Diego - CA, 1 edition, 1999.

JASMINE, J. S. L. et al. Microcalcification detection in digital mammograms based on wavelet analysis and neural networks. Control, Automation, Communication and Energy Conservation, INCACEC 2009, Perundurai, Tamilnadu, 2009.

- KAHLENBORN, C. et al. Oral contraceptive use as a risk factor for premenopausal breast cancer: a meta-analysis. *Mayo Clin Proc.* 81:1290-302, 2006.
- KETTRITZ, U.; MORACK, G.; DECKER, T. Stereotatic vaccum-assisted breast biopsies in 500 women with microcalcifications: radiological and pathological correlations. *Eur J Radiol*; 55(2):270-6, 2005.
- KIRBY, I.; BLAND, M. D.; COPELAND, E. M. *A mama: tratamento compreensivo das doenças benignas e malignas.* Tradução: Nelson Gomes de Oliveira, São Paulo, ed. Manole, 1994.
- KOPANS, D.B. *Breast Imaging*, ed.2, Lippincott-Raven, 2000.
- KOUSKOS, E. et al. Wire localisation biopsy of non-palpable breast lesions: reasons of unsuccessful excision. *Eur J Gynaecol Oncol.*;27(3):262-6, 2006.
- KREYSZIG, E. *Advanced Engineering Mathematics.* John Wiley and Sons, Estados Unidos, ed. 8, 1999.
- KRAUS, H. M. *Ferramenta para desenvolvimento de Sistemas Baseado em Raciocínio Baseado em Casos.* Dissertação de Mestrado, Computação Aplicada, Universidade do Vale do Itajaí, Santa Catarina, São José, 2009.
- LAMBE, M. et al. Parity, age at first and last birth, and risk of breast cancer: a population-based study in Sweden. *Breast Cancer Res Treat.* vol.38: p. 305-11, 1996.
- LAMAS, J. M. *Prevalencia e fatores de risco de câncer de mama e de lesões pré-malignas em mulheres assintomáticas no Distrito Federal.* p.27-29, Dissertação (Mestrado), Curso de Pós Graduação em Medicina, Universidade de Brasília, Distrito Federal, 1998.
- LAMAS, J. M. *Avaliação dos resultados de exames mamográficos para detecção precoce do câncer de mama, no Distrito Federal, segundo indicadores de qualidade.* Tese de Doutorado, Faculdade de Medicina do Rio de Janeiro, Rio de Janeiro, 2000.
- LAMAS, J. M. et. al. *Lesões mamográficas provavelmente benignas: estudo brasileiro apresentado em Viena.* 23rd Internacional Congress of Radiology of the International Society of Radiology, Montreal, June 25-29, 2004.
- LAMI. *On The Breast Comprehensive management of benign and malignat disease*, WB. Saunders Company, 1991.
- LANYI M. *Diagnosis and differential diagnosis of breast calcifications.* 1st ed. Berlin: Springer-Verlag, 1986.
- LASHKARI, A. *Full automatic micro calcification detection in mammogram images using artificial neural network and Gabor wavelets.* *Machine Vision and Image Processing (MVIP)*, 2010 6th Iranian, 2010.

Le Gal M, Chavanne G, Pellier D., Valeur diagnostique des microcalcifications groupées découvertes par mammographies. Bull Cancer. Paris, v.71, n(1), p.57-64, 1984.

LEMACKS, M. R. *et al.* A dual energy subtraction technique for microcalcifications. Imaging in digital mammography : A signal to noise analysis. Medical physics and imaging, Wisconsin, v. 29, p. 1739-1751, Aug. 2002.

LEUNG, J. W.; SICKLES, E. A. The probably benign assessment. Radiol Clin North Am. 45(5):773-89, 2007.

LIBERTMAN, L. *et al.* The Breast imaging report and data system: positive predict value of mammographic features and final assement categories. AJR, vol. 171 p. 35-40, 1998.

LIPNICK, S. F. E. *et. al.* Case control study of risk indicators among women with premenopausal and early post menopausal breast cancer. Cancer 53:1020-1024, 1984.

LORENZI, F. Uso da metodologia de Raciocínio Baseado em Casos na investigação de irregularidades nas internações hospitalares. Dissertação de mestrado em Ciência da Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1998.

MACCHETI, A. H.; MARANA, H. R. C. Densidade mamográfica como fator de risco para o câncer de mama. Rev. Bras. Ginecol. Obst. 29(10):493-6, 2007.

MAJID, A. *et al.*, Missed breast carcinoma: pitfalls and pearls. Radiographics, 23:881-95, 2003.

MANGASARIAN, O. L. "Multi-surface method of pattern separation. Street". W.N., Wolberg W.H. and Mangasarian. O.L., "Nuclear Feature Extraction For Breast Tumor Diagnosis", International Symposium on Electronic Imaging: Science and Technology, vol. 1950, pp. 861-870, San Jose, CA, 1993.

MARQUES FILHO, O.; VIEIRA NETO, H. Processamento Digital de Imagens, Rio de Janeiro: Brasport, 1999.

MARTINS, L. A. L. M.; BARRA, A. A.; LUCENA, C. E. M. Microcalcificações Mamárias Suspeitas de Malignidade.vol.56(2): p. 251-258, Revista Brasileira de Cancerologia, 2010.

MASTERS, T. Signal and Image Processing with Neural Networks - John Wiley & Sons, Inc. - 1994.

McCulloch W. S. Embodiments of Mind. The MIT press, Cambridge, Massachusetts, 1965.

McCulloch, W. S.; PITTS, W. H.A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, v. 5, n. 1 p. 115-133, 1943.

MALLAT, S. A Wavelet tour of signal processing. 2ed. Elsevier, 1999.

MINSKY M. L.; PAPERTE S. A. Perceptrons: an introduction to computational geometry. The MIT Press, 3rd impressão modicada) do original de 1969, 1988.

- MORRISON, A. S. Screening in chronic disease. 2ed. New York, Oxford University Press, 1992.
- MORTON, Karen; OSBORNE, Kerry; Pro Oracle SQL: Exploit the full power of SQL and supporting features in Oracle Database. 2010.
- MOTA, B. S.; RICCI, M. D. Número de casos de câncer de mama registrados por faixa etária em São Paulo. Disponível em: <<http://www.desvendandocancerdemama.com/2011/01/mito-na-imprensa-cancer-de-mama-aumenta.html>>. Acesso em: 13 de Outubro de 2011.
- MUIR, B.; KIRKPATRICK, A.; ROBERTS, M. Oblique-view mammography: adequacy for screening. vol.151 p.39-41, Radiology, 1984.
- NAROD, S. A.; FOULKES, W. D. BRCA1 and BRCA2. Nature Reviews. (4):665-76, 2004.
- NEMA, Digital Imaging and Communications in Medicine (DICOM), Part 5: Data Structures and Encoding. National Electrical Manufacturers Association. EUA, 2000.
- NETTER, Frank H. Atlas de anatomia humana. 4. ed. Rio de Janeiro, RJ: Elsevier, p.548, 2008.
- NUNES, F. L. S.; SCHIABEL, H.; GOES, C. E. Contrast Enhancement in Dense Breast Images to Aid Clustered Microcalcifications Detection. PUBMED, Published online July 6, 2006.
- OPPENHEIM, A. V.; WILLISKY, A. S.; NAWAB, S. H. Signals and Systems. Prentice Hall, Estados Unidos, ed. 2, 1997.
- OTTOMAN, R. et al. Practical guide for estimating risk for familial breast cancer. Lancet ii:556-558, 1983.
- PAQUERAULT *et al.* Investigation of reading mode and relative sensitivity as factors that influence reader performance when using computer-aided detection software. Acad. Radiol., 16:1095-107, 2009.
- PARKIN, D. M.; BRAY, F. I.; DEVESA, S. S. Cancer burden in the year 2000. The global picture. Eur J. Cancer, (37):4-66, 2001.
- PATROCINIO, A. C.; SCHIABEL, H. Classifying clusters of microcalcification in digitized mammograms by artificial neural network. XIV Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'01), pp.266, Florianopolis, Brazil, 2001.
- PEREIRA, M. C. V. Avaliação de Técnicas de Pré-Processamento de Sinais do EEG para Detecção de Eventos Epileptogênicos utilizando Redes Neurais Artificiais, 2003.
- PIATO, S. Mastologia. São Paulo: Roca, 1995.
- PODHOLA, M., URMINSKA, H. Combined radiologic and microscopic examination of nonpalpable breast lesions with microcalcification. Cesk Pathol, 40(1):12-7, 2004.

- PRATT, W. K. Digital Image Processing: PIKS Inside. John Wiley and Sons, Estados Unidos, 3 edition, 2001.
- RAMEZ, E.; SHAM, N.; GUIMARÃES, M. P. Sistemas de banco de dados, Ed. Library Holding, 4. Ed. 2005.
- PRESSMAN, R. S. Engenharia de Software. São Paulo, Ed. Markon Books, 1995.
- QUEIROZ, R. Y.; HUMMEL, A. C.; O valor das microcalcificações agrupadas no diagnóstico precoce do câncer de mama. Revista brasileira de cancerologia. vol. 41(1): 29-31, 1995.
- RANGAYYAN R. M.; BANIK, S.; DESAUTELS, J. E. L. Computer-Aided Detection of Architectural Distortion in Prior Mammograms of Interval Cancer. Society for Imaging Informatics in Medicine 2009, J Digit Imaging; vol. 23(5): p.611–631, October, 2010.
- RANGAYYAN, R. M. ; MUNDIGONDA, N. R.; DESAUTELS, J. E. L., “Boundary modellign and shape analysis methods for classification of mammographic masses”. Medical and Biological Engineering and Computing, vol. 38, pp. 487-495, May 2000.
- REZAI-RAD, G. et al. Detecting microcalcification clusters in digital mammograms using combination of wavelet and neural network. Computer Graphics, Imaging and Vision: New Trends, 2005.
- ROSE, D. P.; BOYAR, A. P.; WYNDER, E. L. International comparisons of mortality rates for cancer of the breast, ovary, prostate, colon and per capita food consumption. Cancer (58):2363-2371, 1986.
- SAMPAIO, W. B. Detecção de massas em imagens mamográficas usando redes neurais celulares, funções geoestatísticas e máquinas de vetores de suporte. Dissertação de mestrado, Curso de Pós Graduação em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luís, 2009.
- SANTOS, V. T.; Segmentação de imagens Mamográficas para detecção de nódulos em mamas densas. Dissertação de mestrado, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2002.
- SICKLES E. A. Management of probably benign lesion. In: Kopans D.B, Mendelson E.B., eds. Syllabus: a categorical course in breast imaging. Oak Brook, III: Radiological Society of North America, p. 133-138, 1995.
- SICKLES E. A. Combination spot-compression and other special views to maximize mammographic information. p. 173-571, Radiology, 1989.
- SILVA JÚNIOR, E. C. Modelo de Processamento de Imagens Mamográficas para detecção de Agrupamentos de Microcalcificações. Tese (Doutorado) Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.

SILVA, A. M. M. et al. Aplicação de técnicas de morfologia matemática para identificação de fases em imagens de raios-X característicos. XVIII Encontro Nacional de Física da Matéria Condensada, Sociedade Brasileira de Física Caxambu, MG, 06 a 10 de Junho, 1995.

SILVA, R. P. Modelo de apoio ao diagnóstico no domínio médico aplicando Raciócio Baseado em Casos. Dissertação (mestrado) apresentada ao ao Programa de Pós-Graduação Strictu Sensu em Gestão do Conhecimento e da Tecnologia da Informação, da Universidade Católica de Brasília, Brasília, 2005.

SOMERVILLE, I. Engenharia de Software. 6º Ed. Tradução Maurício de Andrade. São Paulo: Ed Addison-Wesley, 2003.

STOMPER, P.C., GERADTS, J., EDGE, S.B., LEVINE, E.G. Mammographic predictors of the presence and size of invasive carcinomas associated with malignant microcalcification lesions without a mass. AJR Am J Roentgenol, 18(6)1679-84, 2003.

THÉ, M.A L.. Raciócio Baseado em Casos. Uma Abordagem Fuzzy para Diagnóstico Nutricional. Dissertação (Mestrado em Engenharia de Produção), UFSC, 2001.

TURGAY, A. et al. Computer-aided diagnostic models in breast cancer screening. Imaging Med. vol.2(3): p.313–323, Junho, 2010.

WANGENHEIM, C.H.; WANGENHEIM, A.V. Raciócio Baseado em Casos, SP: Manole, 2003.

WATERHOUSE, J. et al. Cancer Incidence in Five Continents. Scientific Publication n. 15, Lyon, IARC, 1976.

WATTERMAN, D. A. A Guide to Expert Systems. Addison Wesley, 1986.

WOLF J. N. Breast patterns as an index of risk for developing breast cancer. Am J Roentgenol, 126 :1130-7 1976.

WYNDER, E. C.; ROSE, D. P.; COHEN, L. A. Diet and breast cancer in causation and therapy. Cancer (58):804-1813, 1986.

WHO. Breast cancer. Disponível em: <http://www.who.int>. Acesso em: 08 outubro 2011.

ZANDSTRA, M. Entendendo e Dominando o PHP. São Paulo: Digerati Books, 2006.

ZHENG, B. et al. Method to test the reproducibility and to improve performance of computer-aided detection schemes for digitized mammograms. Medical Physics, vol. (31):11, 2964-2972, 2004.

ANEXOS – MODELOS DE DOCUMENTOS

ANEXO 1: ARTIGO PUBLICADO NO ANAIS DO CBEB XXII

AUTOMATIC IDENTIFICATION OF AEADES AEGYPTI EGGS DEPOSITED IN OVITRAPS' SLIDES USING IMAGE PROCESSING TECHNIQUES

Elpídio G. G. Fátima*, Costa F. R. Luan**, Pucci L. Guilherme**, Andrade M. Marcelino***, Costa Jr. A. Edson**, Soares A. Fabiano* and Brasil M. Lourdes***

*Electrical Engineering Department/University of Brasília, Brasília, Brazil

**Engineering Course, University of Brasília, Gama, Brazil

***Post-Graduation in Biomedical Engineering, University of Brasília, Gama, Brazil

e-mail: lmbrazil@gmail.com

Abstract: Dengue, a disease transmitted by the *Aedes Aegypti* mosquito, is indicated as a major public health problem in tropical and semi-tropical countries around the world. According to the World Health Organization, around 100 countries in all continents (except Europe) are affected by the disease. The use of ovitraps has shown significant results in population monitoring and control of dengue vector. In general, the process of counting the eggs of *Aedes Aegypti* mosquitoes present in the slides of ovitraps (ovipaddle images), is performed manually. This paper describes a proposal for automatic counting of eggs deposited in layers of ovitraps, the results show a high correlation ($\rho > 0.88$) compared to the manual counting process.

Palavras-chave: Dengue, Ovitrap, Digital Image Processing.

Introduction

The dengue, a disease caused by a virus, has already become a major problem of public health worldwide. According to World Health Organization (WHO) this disease is present in over 100 countries on every continent except Europe, as showed in Figure 1. Accordingly to WHO, 50 to 100 million people are infected each year by the dengue virus [1]. in accordance with the latest survey conducted by the Health Ministry in Brazil, until the month of February/2010, the country already had an increase of 109.43% in the number of cases of infection by the dengue virus over the past year. Consistent with the Brazilian Ministry of Health, 108,640 people have contracted the disease this year, while in 2009 the number of cases was 51.873 [2]. Brazil relies on socio-environmental conditions that contribute to the resistance of the vector that transmits dengue, mosquito *Aedes aegypti*. The course and progression of the disease could be eradicated with measures focused in education and conscientization of the population as well

as investments in basic sanitation. But in practice, the measure taken is the application of poisons in the air and water to eradicate the dengue vector. As stated by studies this practice may compromise the health of persons because of the effect of neurotoxins and allergenic substances used on the poison [3]. Given this situation it is necessary to consider less aggressive solutions on the fight against this disease.

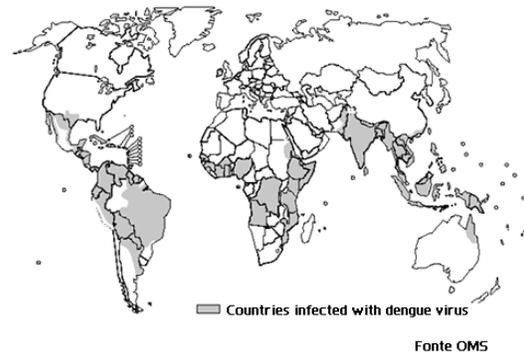


Figure 1: Countries affected by the dengue virus [4].

Studies have reported the efficient use of ovitraps (sort of trap used to capture the eggs of *Aedes Aegypti* mosquitoes) in combating the proliferation of dengue vector [5]. Additionally, the use of ovitraps allows the estimation of the progress of dengue disease in the environment by counting eggs [6, 7]. However, since the counting of the mosquitoes eggs in the ovitraps slides is done manually, it is a highly slow process. One approach to automate the process of counting the eggs contained in the Slides of Ovitrap (SO) is the use of digital image processing techniques (DIP) on the SO. The use of computers in this activity can considerably reduce the counting time and the tedious aspects for the human experts in this task. However, the image processing techniques used in biological tasks tends to deal with form irregular shapes. This factor denotes a

greater challenge in the use of image processing for automation of practical applications in the field of biological sciences [8]. Therefore, the interpretation of images by means of quantitative analysis is an important support for many scientific areas, with regard to the applications of techniques oriented for: segmentation, statistical analysis, pattern recognition and feature extraction [9].

The work in question is the scope of the project entitled "System Acquisition and Image Processing of Ovitrap" (SAPIO [10], from the Portuguese), coordinated by the Federal University of Pernambuco in partnership with the Aggeu Magalhães and the University of Brasilia. The aim of this paper is the use of traditional methods of DIP for, among other things, create a tool that automates the process of eggs counting present in SOs to support the fight against the Dengue Disease.

Materials and Methods

The algorithm developed for automatic counting of eggs present in SO was implemented with Matlab software version 7.6. To form a data bank for testing and validation of the algorithm, initially a matrix with size 2300x6000 pixels was obtained from a scanned image of a SO previously collected in the field. Then 100 sample images with size 512x512 pixels were generated from the original image randomly selecting 100 sub images of 512x512 pixels, divided into two groups A and B of 50 images each. The typical ovitrap used by Central Public Health Laboratory in Pernambuco (LACEN, from Portuguese) is shown in Figure 2.

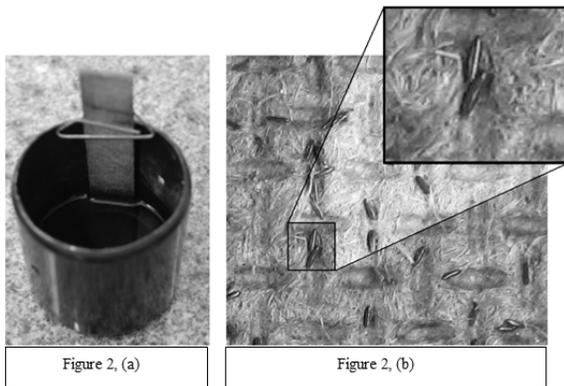


Figure 2: Typical ovitraps (a) and RGB color sub-image of ovipaddle (b)

Random sampling was used to minimize possible bias in the formation of the sample groups, while the size (512x512) and quantity (100) images were established to minimize possible human error in Manual Counting (MC) used in the validation of the Automatic Counting Algorithm (ACA).

The ACA can be divided into three basic stages. In the first step it is performed a binarization of the color images of the scanned SO. The second step the identification and classification of the regions of eggs

occurrence is performed and the final step is the estimation of the total number of eggs in each image. Such information is illustrated in Figure 3.

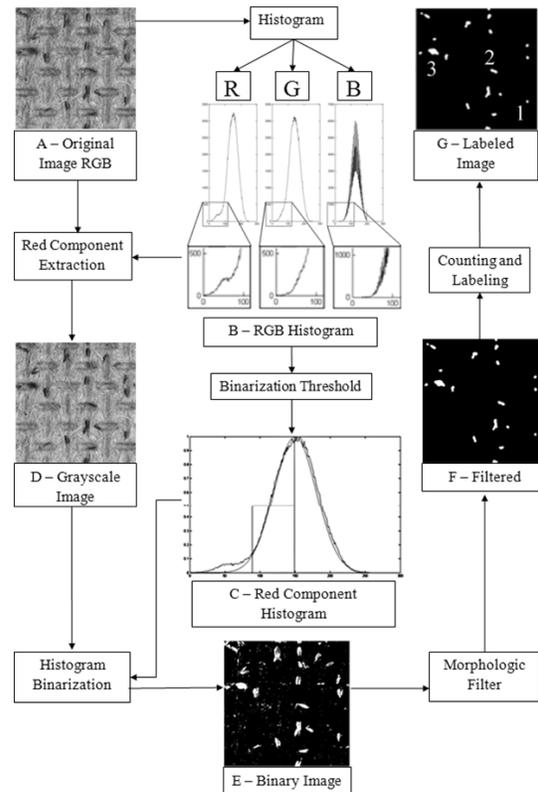


Figure 3: Diagram describing the flow of ACA eggs present in SO.

Starting from the original RGB image (Figure 3A) it was found by analysis made on the histogram of the three components (Figure 3B) that the red component presented the best contrast between the eggs and the region of the blade. This feature can be seen in Figure 3-C and in more detail in Figure 4, in which the red component histogram has a bimodal representation.

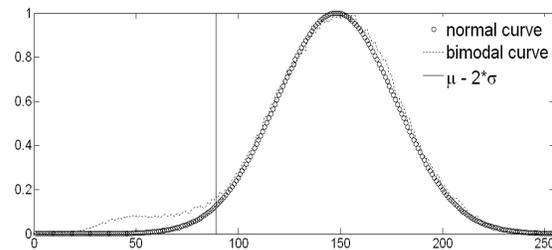


Figure 4: Red Component Histogram

The Gaussian approximation applied to the histogram of the red component allows the visualization of a more clearly bimodal curve. With this information it was possible to discern that the highest peak of the bimodal curve has relationship with the region of the blade and

the lower peak has relationship with the region of the eggs.

Figure 3-D, which represents the red component in grayscale, was subjected to a process of binarization. The threshold for binarization (L) was determined by the equation 1

$$L = \mu - 2\sigma \quad (1)$$

where μ is the mean and σ is the standard deviation of the normal curve that approximates the distribution of pixels corresponding to the SO (Figure 3-D, normal curve). This choice was made with the intention of segmenting the image highlighting the eggs from the SO (the levels $r > L$ account for 97.5% of the histogram curve that represents the SO). The result of binarization by histogram is shown in Figure 3-E.

Figure 2-F was obtained by the application of morphological operations as follows: 1. An morphological opening was applied in the binarized image (Figure 2E). The structuring element was a disk of radius 5 pixels; 2. A morphological close was applied to the result of the opening. The structuring element used in that closure was a disk of radius 7 pixels. The choice of structural elements of morphological filtering was determined respecting the dimensions of the eggs and noise. Thus, morphological opening minimizes the noise in the image and morphological closing eliminates the failures present in some eggs images. It was also found that this sequence of morphological operators soften the threshold of the eggs on the image.

The counting and classification process began with the labeling of eggs from the N regions identified in Figure

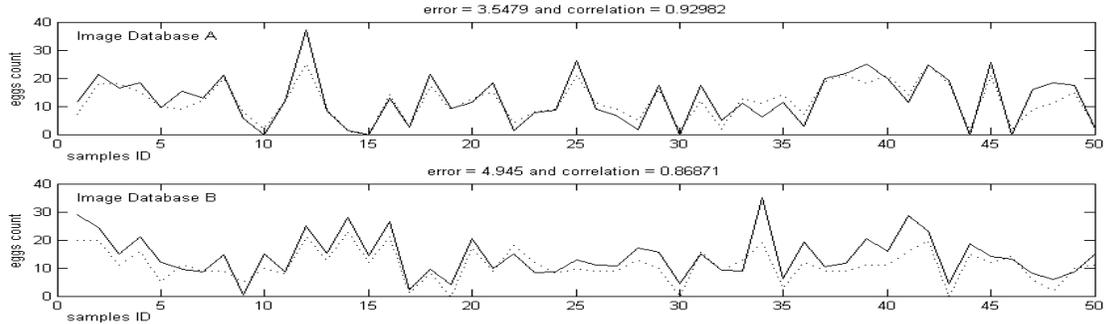


Figure 5: Egg count for the banks image. The continuous line represents the manual counting. Y-axis represents the number of eggs found and x-axis represents the image evaluated.

3-F. Then the labeled regions R_i were classified from the area A_i , in pixels, of each region as: areas with union of three eggs (R_t), with union of two eggs (R_d), a single egg (R_e) or noise (R_n) as stated by the following criteria: where $|R_x|$ is the number of regions of type x in the image (Figure 3-G). In which the regions are classified as regions without overlapping (R_e), regions with double overlap (R_d) and regions with triple overlap (R_t). These regions are labeled in accordance with the size of the regions. It is noteworthy that from the 100 samples tested, most of the eggs are presented in isolation, what justifies the measure R_e and consequently the other measures $|R_x|$.

Manual counting of all images was performed by three volunteers and the average of the numbers of eggs indicated by them was taken as a benchmark for purposes of validation of the automatic counting algorithm.

Results

Figure 5 shows the comparison between the numbers of eggs counted using each method. For each of the two image banks it was computed the mean square error of the automatic counting in relation to manual counting, as well as the correlation between both counts.

To analyze the performance of the automatic counting algorithm, the number of eggs given by the algorithm was compared with the average value of the counts of the two volunteers. Statistical measures of distance (mean squared error and standard deviation) and similarity (mean and correlation) were used for the comparative analysis.

$$C(R_i) = \begin{cases} R_t, & \text{if } A_i \geq \mu_A + 3\sigma_A; \\ R_d, & \text{if } \mu_A + 2\sigma_A \leq A_i < \mu_A + 3\sigma_A; \\ R_e, & \text{if } \mu_A - 2\sigma_A \leq A_i < \mu_A + 2\sigma_A; \\ R_n, & \text{if } A_i < \mu_A - 2\sigma_A; \end{cases} \quad (2)$$

where μ_A is the mean and σ_A is the standard deviation of the areas A_i . The estimative of number of eggs E was computed using equation 3.

$$(3) \quad + 2|R_d| + 3|R_t|,$$

Table 1 shows the results of both methods of counting in terms of mean and standard deviation of eggs laid in ovitraps on site.

Image Database	Method	Mean	Standard Deviation
A	Automatic	11.540	6.753
	Manual	12.427	8.647
B	Automatic	11.120	5.749
	Manual	14.200	7.576

Table 1: Comparison of Methods to Automatic Counting Developed on Algorithm.

Discussion

The low complexity of the proposed algorithm (Figure 2) allows both its implementation in low cost embedded systems and the development of an integration module to the portable Geographic Information System (GIS) in project SAPIO [10]. This module would allow the counting of eggs laid in ovitraps on site.

Evaluating the intervals of confidence of the eggs counting procedures manual (IM) and automatic (IA) through the mean values and standard deviations presented in Table 1, it is evident that for the two banks evaluated, this result indicates a possible equivalence between the two counting strategies ($IA \subset IM$).

The mean squared error was used to measure the statistical distance in a way to evaluate the performance of the algorithm. However errors are also expected in manual counting on a large scale process. Also, the strategy of computational counting prevents the intra-group error.

In the evaluation of similarity through the correlation method it was sought whether the trends identified in the manual and automatic counts retained some loyalty. According to the graph of Figure 3, the curves of both methods of counting had correlation coefficients ρ above 0.88. Applying the statistical test of hypothesis [11] it was obtained a significance level exceeding 99%, confirming, in spite of the mean square error, the strong correlation between the counting methods.

One of the biggest challenges of working with images of ovitraps is to distinguish the total number of eggs in areas with overlap, which could be minimized with the design of ovitraps slides that avoid this phenomenon. In the present work the algorithm considered quantitative and statistic metrics only. In future work the authors are planning to incorporate geometric and morphological information to improve the results.

It is evident that the strength of this method depends on the quality and uniformity of the collected images. Therefore, it is essential to standardize the process of capturing images for the proper performance of the proposed computational strategy.

Conclusion

In the present work, an algorithm based on image processing techniques was developed to automatically count the number of eggs in ovitraps. The quantitative approach used in the study shows high similarity between the manual and automatic eggs counting process, indicating the potential of the developed technique in the fight against dengue.

Acknowledgment

This research is partially sponsored by FINEP-Brazil and CNPq-Brazil. The authors are grateful to Aggeu Magalhães Research Center for the concession of the ovitraps used in this study.

References

- [1] Brasil. Ministério da Saúde. *Programa Nacional de Controle da Dengue*. Available at: http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=23614.
- [2] Brasil. Organização Mundial da Saúde. *Secretaria de Vigilância em Saúde Informe Epidemiológico da Dengue Análise de situação e tendências – 2010*. Available at: HTTP://new.paho.org/bra/index.php?option=com_content&task=view&id=1007&Itemid=256.
- [3] CÂMARA N. F. H., AUGUSTO L. G. S (2000), *Condições Sanitárias do Ambiente Urbano e o uso de Pesticidas Domésticos: implicações para a saúde*. In: XXVII Congresso Internacional de Engenharia Sanitária e Ambiental. Porto Alegre. Livro de Resumos. Porto Alegre: ABES-AIDIS. v.1.
- [4] Brasil. Ministério da Saúde. *Agência Nacional de Agência Sanitária*. Available at: <http://www.anvisa.gov.br/paf/viajantes/dengue.htm>.
- [5] BRAGA I. A., VALLE D.(2007), *Aedes aegypti: vigilância, monitoramento da resistência e alternativas de controle no Brasil*. *Epidemiol. Serv. Saúde*, Brasília, v.16, n.4, p.295-302.
- [6] BONAT W. H. et al. (2009) *Investigando Fatores Associados a Contagens de ovos de Aedes Aegypti Coletados em ovitrampas em Recife/P*. *Rev. Bras. Biom.*, São Paulo, v.27, n.4, p.519-537.
- [7] Health's Ministry (2009) *National Guidelines for Prevention and Control of Epidemic Dengue*, Health's Ministry, Secretariat of Health Surveillance, Department of Epidemiological Surveillance – Serie A. Standards and Technical Manuals, Brasília, p. 160.
- [8] Glasbey C.A., Horgan G.W. (1994), *Image Analysis for the Biological Sciences*, Ed.: New York: John Wiley and Sons, Scotland, Edinburgh, p. 5.
- [9] Gonzalez R.C., Woods R. (2008), *Digital Image Processing*, Upper Saddle River, New Jersey: Prentice-Hall, 3rd edition.
- [10] Mello C.A.B., Wellington P.S., Rodrigues M.A.B., Candeias A.L.B., Gusmão C.M.G. (2008), *Image Segmentation of Ovitrap for Automatic Counting of Aedes Aegypti Eggs*. In: 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24.
- [11] Levin J. (1987), *Estatística Aplicada a Ciências Humanas*, Brasil: HARBRA, p.143-169.

ANEXO 2: ARTIGO PUBLICADO NO ANAIS DO ECT2010

Identificação de Ovos de *Aedes Aegypti* Presentes em Lâminas de Ovitrapas Através de Redes Neurais Artificiais

Fátima Gisele G. Elpídio¹, Lourdes M. Brasil², Marcelino M. Andrade³

^{1,2,3} Departamento de Engenharia Elétrica – Universidade de Brasília (UNB) Área Especial 2 Lote 14
Setor Cep: 72405-610 – Central Gama – DF – Brasil.

fgizele@hotmail.com, {lmbrasil, andrade}@unb.br

Abstract. *Dengue is identified as one of the main problems global health. Brazil has socio-environmental factors that contribute to the resistance of the vector transmitter, Aedes Aegypti, reflecting the disease progress. The use of ovitraps has itself shown a simple and inexpensive alternative for monitoring and controlling Dengue vector. Generally the process of counting Dengue mosquito eggs present in the slides of ovitraps, is performed manually. This article describes a proposal to automate the identification process of eggs of the mosquito Aedes Aegypti deposited in regions of the ovitraps through the use of Artificial Neural Networks.*

Resumo. *A Dengue é apontada como um dos principais problemas de saúde mundial. O Brasil conta com condições sócio-ambientais que contribuem para a resistência do vetor transmissor, Aedes Aegypti, refletindo no progresso da doença. O uso de Ovitrapas apresenta-se como uma alternativa simples e barata para monitorar e controlar o vetor da Dengue. Geralmente o processo de contagem dos ovos do mosquito da Dengue, presentes em lâminas de ovitrapa, é realizado manualmente. Esse artigo descreve uma proposta de automatização do processo de identificação de ovos do mosquito Aedes Aegypti depositados em regiões da ovitrapa, através do uso de Redes Neurais Artificiais.*

1. Introdução

A Dengue, doença causada por um vírus e transmitida por um mosquito, é um problema típico de áreas tropicais e semi-tropicais do mundo. Segundo dados da Organização Mundial da Saúde diversos países são afetados com inúmeras ocorrências dessa doença, resultando em necessidades de hospitalização ou até óbito dos doentes [1]. Pesquisas revelam que o uso de ovitrapas (espécie de armadilha usada para capturar os ovos do mosquito) tem apresentado resultados significativos no combate a Dengue [2], além de contribuir na investigação da distribuição sazonal de ovos do mosquito e viabilizar o controle do impacto das medidas adotadas [3].

A presente proposta aborda o uso de Redes Neurais Artificiais (RNA) *Perceptron* de Múltiplas Camadas [4] em conjunto com técnicas de Processamento Digital de Imagens [5]. O objetivo é automatizar o processo de identificação de ovos do *Aedes Aegypti* em regiões da Lâmina de Ovitrapa (LO), com o intuito de minimizar a análise manual de cada ovitrapa. Esse trabalho é parte de um projeto de extensão maior (Sistema de Aquisição e Processamento de Imagens de Ovitrapas - SAPIO)

desenvolvido pela Universidade de Brasília (UnB) - Faculdade Gama (FGA) e Universidade Federal de Pernambuco (UFPE) [6].

2. Processamento Digital de Imagens e Redes Neurais Artificiais

Segundo [5], o processamento de imagens aborda, dentre outros contextos, a execução de transformações sobre a imagem com o intuito de melhorá-la e facilitar a extração de informação da mesma.

As RNAs têm sido freqüentemente utilizadas em problemas de reconhecimento e classificação de padrões devido a sua capacidade de aprender por meio de exemplos. Essa capacidade advém do processo de aprendizado, onde um conjunto de treinamento é interativamente apresentado à RNA, a qual a partir de um algoritmo de aprendizado constrói uma representação interna do problema, realizando um mapeamento, geralmente não-linear, das entradas e de suas respectivas saídas [4].

A RNA *Perceptron* Múltiplas Camadas (PMC) é caracterizada por um conjunto de neurônios interligados por múltiplas camadas. As RNAs PMCs podem usar um grande número de técnicas de aprendizado, sendo que a mais popular é a propagação reversa. Neste caso, os valores da saída são comparados com a resposta correta para computar o valor de alguma “função-erro” predefinida, em seguida o erro é alimentado de volta na rede. O algoritmo ajusta os pesos de cada conexão para reduzir o valor da “função-erro”. Após repetir este processo por um número suficiente, os pesos sinápticos são ajustados para fazer com que a resposta real se aproxime da resposta desejada [4, 7].

3. Materiais e Métodos

O algoritmo elaborado para detecção de ovos presentes na LO foi desenvolvido no software Matlab. Na criação do conjunto de treinamento e teste da RNA foram realizadas as seguintes etapas no processo de segmentação:

Foram empregadas 5 imagens de 2300x6000 *pixels*, as quais foram escaneadas de LOs previamente coletadas em campo.

Com o intuito de melhorar o desempenho computacional e aperfeiçoar o treinamento da RNA, as imagens foram subdivididas em sub-imagens com tamanhos de 512x512 *pixels* por meio de um algoritmo desenvolvido para segmentação seqüencial. Nessa etapa foram originadas 102 sub-imagens.

Por fim, foram extraídas as Regiões de Interesse (em inglês *Region of Interest* – ROI) [7] de cada sub-imagem. Para cada ROI foi originada uma imagem binarizada.

Antes do processo de classificação dos ovos com a RNA PMC, cada sub-imagem passou por um processo de pré-processamento, onde elas eram binarizadas (pela tonalidade da ROI). Foi aplicado um filtro mediana e as operações morfológicas de abertura e fechamento para minimização de efeitos de ruídos. Por fim realizou-se um processo de rotulação dos objetos presentes na imagem binarizada, cujos resultados foram empregados como o valor desejado para classificação da RNA. O valor desejado da RNA foi classificado em: nenhum objeto rotulado na imagem, ruído, ovo e agrupamento de ovos. A entrada da RNA foi extraída do histograma da sub-imagem, originadas na etapa III do processo de segmentação.

O conjunto de exemplos utilizados (resultante da etapa de segmentação das 5 imagens da etapa I, do processo de segmentação) foi de 217 imagens organizadas em sete bancos de dados. Seis bancos de dados foram utilizados no treinamento e um para teste. Esta estratégia proporcionou sete possibilidades de validação do sistema, com a variação do banco de dados designado para teste.

4. Análise dos Resultados

A RNA PMC utilizada foi testada para diferentes valores das seguintes variáveis: taxa de aprendizagem, número de neurônios nas camadas, termo *momentum* e função de ativação. A função sigmóide e a função tangente hiperbólica foram empregadas em diferentes testes para analisar o desempenho da RNA. Os melhores resultados foram obtidos com a seguinte topologia: uma taxa de aprendizagem de 0,95 e um termo *momentum* com o valor de 0,01, duas camadas ocultas contendo 16 e 50 neurônios, respectivamente. A camada de saída foi estruturada com quatro neurônios que correspondem a formatação das classificações para o valor desejado.

A estratégia de permutação dos bancos de dados empregados no treinamento e no teste, em sete possibilidades, permitiu validar que a RNA obtinha melhores resultados nos bancos em que os dados dos exemplos estavam organizados mais aleatoriamente e quando esses exemplos apresentavam uma maior abrangência dos exemplos que seriam validados no teste. A Tabela 1 apresenta os dados obtidos em cada um dos testes.

Tabela 1 – Dados obtidos no treinamento

Matriz de Teste	NIST	EMQ	CGC	CCC	PCCC
B	23	0,0414	Vermelho	21	91,30%
C	41	0,0475	Azul	37	90,24%
D	23	0,0497	Cian	23	100 %
E	33	0,0425	Rosa	30	90,91%
F	35	0,0387	Preta	33	94,29%
G	35	0,042	Amarela	33	94,29%
H	27	0,046	Azul claro	26	96,3%

Legenda da Tabela:
NIST - Número de Imagens submetidas para teste.
EMQ - Erro Médio Quadrático.
CGC - Curva no Gráfico Comparativo.
CCC - Casos Classificados Corretamente.
PCCC - Porcentagem de Casos Classificados Corretamente

A Figura 1 ilustra o gráfico do erro médio quadrático em cada um dos testes:

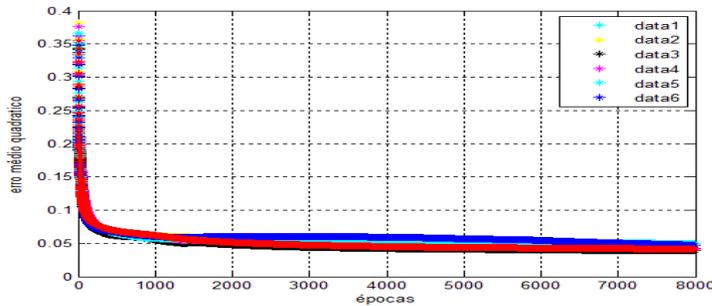


Figura 1. Variação do erro médio quadrático e número de épocas.

Os melhores resultados obtidos foram com a função sigmóide, que permitiu atingir um erro médio quadrático de aproximadamente 0.057 e com cerca de 91% de casos classificados corretamente.

5. Conclusão

O uso da rede PMC apresentou resultados significativos para o problema em questão. Foi observado que a carência de exemplos em alguns bancos influenciou no resultado da classificação de alguns padrões. Futuros trabalhos podem ser realizados viabilizando até a contagem dos ovos em cada placa. Para isto, é necessário aprofundar o conhecimento da rede PMC de forma que ela consiga identificar o limite de um ovo para outro em aglomerados.

7. Referências

- [1] Brasil. Ministério da Saúde. Programa Nacional de Controle da Dengue. Disponível em URL: http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=23614. Acesso em : 21/06/2009.
- [2] Braga I. A., Valle D. (2007) *Aedes aegypti*: vigilância, monitoramento da resistência e alternativas de controle no Brasil. *Epidemiol. Serv. Saúde*, Brasília, v.16, n.4, p.295-302.
- [3] Bonat W. H. et al. (2009) Investigando Fatores Associados a Contagens de ovos de *Aedes Aegypti* Coletados em ovitrampas em Recife/P. *Rev. Bras. Biom.*, São Paulo, v.27, n.4, p.519-537.
- [4] Haykin S.(2001) *Redes Neurais. Princípios e práticas*, 2a. ed, Ed. Bookman.
- [5] Gonzalez R.C., Woods R. (2008), *Digital Image Processing*, Upper Saddle River, New Jersey: Prentice-Hall, 3rd edition.
- [6] Mello C.A.B., Wellington P.S., Rodrigues M.A.B., Candeias A.L.B., Gusmão C.M.G. (2008), *Image Segmentation of Ovitrap for Automatic Counting of Aedes Aegypti Eggs*. In: *30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada*, August 20-24.
- [7] Fernandes, F. C. (2007), *Sistema Neuro-Fuzzy para Classificação de Calcificações em Mamogramas*. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação) - Universidade Católica de Brasília.

ANEXO 3: ARTIGO PUBLICADO NO ANAIS DO PAHCE-2011

Contagem automática de ovos de *Aedes Aegypti* depositados em Ovitrapas através de algoritmo de Processamento Digital de Imagens e Rede Neural Artificial

F. G. G. Elpídio¹, L. F. R. Costa², M. M. Andrade³, E. A. Costa Jr.⁴, L. M. Brasil⁵, M. A. B. Rodrigues⁶
1, 5 Pós-Graduação em Engenharia Biomédica, Faculdade Gama, Universidade de Brasília, Gama, Brasil

2,3,4 Engenharia Eletrônica, Faculdade Gama, Universidade de Brasília, Gama, Brasil

According to World Health Organization, Dengue is identified as one of the main pandemic viral disease in tropical and semi-tropical countries in the world. A lot of researches have divulged that the use of ovitraps has itself shown a simple and inexpensive alternative for monitoring and controlling Dengue vector, *Aedes Aegypti*, but generally the process of counting Dengue mosquito eggs in ovitraps is performed manually. This paper describes a proposal for automatic counting of *Aedes Aegypti* eggs deposited in ovitraps by Digital Image Processing techniques associated to an Artificial Neural Network.

Keywords — Dengue, Digital Image Processing, Ovitrapas, Artificial Neural Network

As ovitrapas (espécie de armadilha utilizada para capturar os ovos do mosquito *Aedes Aegypti*) interrompem o ciclo de evolução do mosquito e contribuem na investigação da proliferação da Dengue através da contagem dos ovos [1]. Atualmente o processo de contagem dos ovos do mosquito *Aedes* presente em ovitrapas é realizada manualmente por um técnico treinado empregando uma lupa ou microscópio.

O algoritmo desenvolvido para a Contagem Automática dos Ovos em Imagem Digitalizada de Ovitrapa (CAOIDO) foi elaborado com a versão 7.6 do software Matlab e pode ser dividido em três etapas básicas. Primeiro é realizado a binarização da Imagem colorida Digitalizada da Ovitrapa (IDO), em seguida, à identificação e classificação das regiões de ocorrência de ovos e por último ocorre à estimativa do total de ovos em cada imagem. Esse algoritmo é uma evolução de [2], atualmente está em fase de teste à amostragem estatística dos exemplos, o uso do método de binarização com limiar adaptativo [3], e uso de uma Rede Neural Artificial (RNA) Perceptron Múltiplas Camadas (PMC) [4] para classificar os ovos.

Nesse estudo foram empregadas 30 amostras de ovitrapas típicas, previamente digitalizadas. As ovitrapas tinham um tamanho típico de 2300x6000 *pixels* e foram submetidas a um processo de segmentação para gerar um banco de sub-imagens com 512x512 *pixels* (visando facilitar a contagem manual). O algoritmo realiza o processamento individual das amostras (similar ao realizado em [2], porém com as adaptações descritas anteriormente) para obter o total de ovos por imagem, posteriormente é realizado um somatório da estimativa de ovos por banco de imagens. O resultado do algoritmo CAOIDO na contagem de ovos do mosquito da Dengue nas IDO foi comparado com a contagem manual de técnicos treinados, para validar o desempenho do algoritmo. O teste estatístico de hipótese [5] apresentou um nível de significância de 99%, confirmando uma forte correlação entre os métodos de contagem.

A binarização com limiar adaptativo atribui um tratamento uniforme para imagens digitalizadas de ovitrapas, pois a maioria é corrompida por fatores ambientais e falhas no processo de digitalização. A RNA PMC aprimora o processo de rotulação e classificação dos ovos do mosquito da Dengue, na IDO, e atua como um mecanismo de decisão, considerando fatores como o tamanho médio do ovo, área do objeto rotulado e dentre outros fatores para realizar a classificação. Em testes desenvolvidos com o RNA PMC, foi possível atingir com ela um erro quadrático médio de aproximadamente 0,09 e cerca de 94% de casos classificados corretamente [6]. A forte correlação entre a contagem manual e o algoritmo CAOIDO, indica o potencial da técnica desenvolvida. Essa pesquisa é parte de um projeto de extensão maior intitulado *Sistema de Aquisição e Processamento de Imagens de Ovitrapas (SAPIO)* desenvolvido pela *Universidade de Brasília (UnB) - Faculdade Gama (FGA)* e *Universidade Federal de Pernambuco (UFPE)*.

REFERÊNCIAS

- [1] W. H. Bonat, et al “Investigando Fatores Associados a Contagens de ovos de *Aedes Aegypti* Coletados em ovitrampas em Recife,” *Revista Brasileira de Biometria*, vol.27, n°. 4, pp. 519-537, 2009, “*impresso*”.
- [2] F. G. G. Elpídio, et al “Automatic Identification of *Aedes Aegypti* Eggs Deposited in Ovitrap Slides Using Image Processing Techniques”. *XXII Brazilian Congress of Biomedical Engineering (CBEB, in Portuguese)*, Rio: SBEB, vol. 1, Nov. 2010, “*impresso*”.
- [3] W. Niblack “An Introduction to Digital Image Processing” *IEEE Trans. Biomed.* NJ: Prentice Hall, 1986.
- [4] S. Haykin *Neural Networks: A Comprehensive Foundation*, Ed. Upper Saddle River, NJ: Prentice Hall, 1999.
- [5] J. Levin *J. Estatística Aplicada a Ciências Humanas*, Brasil: HARBRA, pp.143-169, 1987.
- [6] F. G. G. Elpídio, L. M. Brasil, M. M. Andrade. “Identificação de Ovos de *Aedes Aegypti* Presentes em Lâminas de Ovitrapas Através de Redes Neurais Artificiais”. Encontro de Ciência e Tecnologia do Gama (ECT Gama), Brasília, 2010, “*inédito*”.

ANEXO 4: ARTIGO PUBLICADO NO ANAIS DO ISEM-2011

Using Artificial Neural Networks to Identify *Aedes Aegypti* Eggs in Ovitrap

Fátima G. G. ELPÍDIO¹, Lourdes M. BRASIL² and Ana P. B. SILVA³

^{1,2} *Post-Graduation Course in Biomedical Engineering, University of Brasília, Gama College, Brazil*

Tel.: +61 3107-8201; Fax: +61 3107-8219;

E-mail: fgizele@hotmail.com, lmbrasil@gmail.com

³ *Electrical Engineering Department, University of Brasília, Brasília, Brazil*

Tel.: +61 3107-8201; E-mail: anapaula@ucb.br

Abstract. According to the World Health Organization (WHO), Dengue is one of the most widely spread viral diseases transmitted by mosquitoes. Many studies have promoted the use of traps as a simple and cheap alternative for monitoring and controlling the vector of the dengue virus, the *Aedes Aegypti*. However, the analysis and counting of the mosquito eggs in ovitraps, is completed manually. This article describes the proposal for the automation of the process of identifying Dengue mosquito eggs in ovitraps through the Artificial Neural Networks (ANN). In tests completed with ANN developed it was possible to achieve a correct classification rate of 94% cases.

Keywords: Dengue, *Aedes Aegypti*, Ovitrap, Artificial Neural Networks (ANN).

Introduction

Dengue is identified as one of the main viral pandemics in tropical and semi-tropical areas in the world. Annually approximately 50 million people are affected by the Dengue virus resulting in hospitalization or death [1].

The mosquito that transmits Dengue, *Aedes Aegypti*, lives inside or near urban habitations, reproducing in places where there is stagnant water (lakes, bottles, tires, etc.). Many studies demonstrate that the cycle and evolution of the illness can be minimized with measures that focus on raising awareness and education of the population, besides investments in basic sanitation [2]. Meanwhile, the application of poisons in the air and water is one of the most frequently adopted methods for combating the Dengue vector. Various studies have shown that these measures can compromise human health in virtue of the neurotoxic and allergenic effect of the substances used [3].

Facing this situation, studies have demonstrated the efficiency of ovitraps (a type of trap to capture the *Aedes Aegypti* eggs) interrupting the evolution cycle of the mosquito and by estimating the proliferation of Dengue by containing the *Aedes Aegypti* mosquito eggs, in ovitraps [2]. Generally the counting is completed manually by trained technicians who complete the counting with the help of a microscope or magnifying glass.

The Artificial Neural Network (ANN) has been frequently used in problems of recognizing and classifying patterns due to its capacity to learn through example [4]. This capacity comes from the learning process, where a set of training procedures is interactively presented to the network, which starting from an algorithm of learning constructs an internal representation of the problem, completing a mapping, generally non-linear, of the entrances and respective exits. In tasks of recognition of visual

patterns, a set of training procedures is formed by images. These are submitted to pre-processing steps [5] aimed at reducing noise, improving quality and enhancing areas of interest to better the quality of the training procedure set, once it is directly related to the learning capacity and ANN generalization [4].

The present proposal addresses the use of an ANN Multi Layer Perceptron (MLP) [4] to automate the process of identifying *Aedes Aegypti* eggs in areas of Lamina of the Ovitrap. This study addresses a new computational paradigm besides complementing the work developed in the algorithm described in earlier works [6-8]. This work is part of a wider ranging project (System of Acquisition and Processing of Ovitrap Images—SAPIO) developed by the University of Brasilia (UnB)—Gama College (FGA) and the Federal University of Pernambuco (FPE) [6-8].

2. Materials and Methods

2.1. Specification of the algorithm developed and samples used

The algorithm developed to automatically identify the *Aedes Aegypti* eggs in digitalized images of ovitraps was elaborated with Matlab software, version 7.6. In this study, 5 samples of typical ovitraps, previously digitalized were used. The images measure, on average, 23000X6000 pixels and were submitted to a subdivision process to generate a bank of sub-samples of original images to test and validate the ANN. The sub-sampling of the original images has as an objective, to help the in the training of the ANN and later to compare the results obtained with the manual analysis completed by trained technicians.

Segmentation process and ROI extraction of the digitalized Ovitrap

In the process of segmentation the Regions of Interest (ROI) were extracted [5], resulting in 217 sub-images, organized in 7 data banks, being that 6 were used in training and one for testing. This strategy resulted in 7 possibilities of system validation with the variation of the bank designed for the test. Figure 1 illustrates the process of sub-sampling and ROI extraction of the digitalized images of the ovitraps.

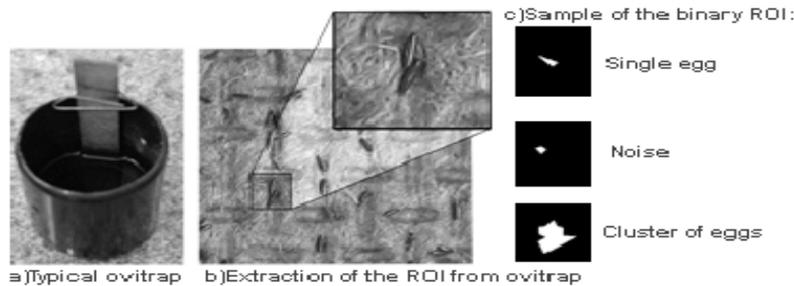


Fig.1. Example of extraction and classification of ROI obtained from a typical ovitrap.

Digital processing of the Ovitrap images and formation of the bank of entrances ANN PMC.

Before the classification process of the *Aedes Aegypti* eggs with the ANN PMC, each sub-image was submitted to a pre-processing process, where they were binarized by tonality of the ROI. An average filter was applied and the morphological operations of opening and closing to minimize noise effects [5]. These operations are useful in the identification and extraction of image characteristics, besides helping in the identification and minimization of noise that interferes in the form of objects contained in the image. In general, the application of the morphological operations simplifies the objects of interest and enhances essential characteristics of the same. Therefore, the objective of the interactive application of the opening and closing operations is the elimination of artifacts of the image that interfere in the identification of the ROI without distorting the global geometry of the characteristics that it intends to preserve.

Finally, the labeling of the objects present in the binary image, whose result was used as a desired value in the classification of ANN, was completed. The desired value was classified in: no labeled object, noise, single egg or cluster of eggs. The classification of the desired value considered factors such as average size of the egg, area of the labeled object, among others, to complete classification. The entrance of the ANN was extracted from the histogram of the binarized sub-images, originated after the process of segmentation.

Results and Discussions

The ANN MLP perfects the process of identification of Dengue mosquito eggs, on the digitalized image of the ovitraps, and acts as a deciding mechanism to facilitate classification of ROI. The ANN PMC used was tested with various configurations and the best results were obtained with the following architecture: a rate of learning from 0.95 and one term momentum with the value of 0.01, two hidden layers containing 16 and 50 neurons, respectively. The exit layer was structured with 4 neurons that correspond to the possibilities of classifications of desired value.

In tests developed with the ANN PMC, it was possible to achieve an average error quadratic [4] of approximately 0.09 and about 94% of cases classified correctly. During the classification, the performance of the ANN was analyzed with the hyperbolic functions tangent and sigmoid [4]. It was observed that the ANN PMC obtained better performance during the classification with the use of the Sigmoid function. Figure 2 presents the results of the training with both functions and the consequential performance of the ANN PMC.

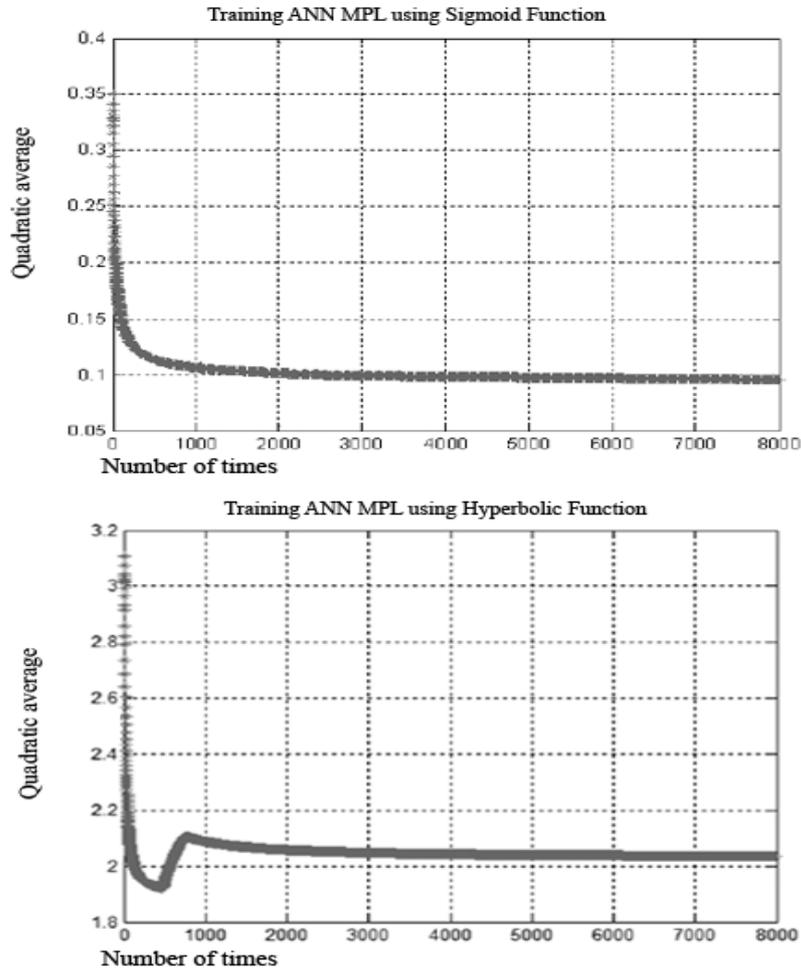


Fig.2. Performance of the ANN PMC (variation of average quadratic error) according to the type of function.

Table 1 presents the set of data resulting in the classification of the ANN PMC according to the set of image banks.

Table 1 – Results of the ANN PMC Classification

Image database	Number of images submitted to testing	Average Error Quadratic	Correctly classified cases	Percentage of correctly classified cases
DATABASE 1	23	0,0414	21	91,30%
DATABASE 2	41	0,0475	37	90,24%
DATABASE 3	23	0,0497	23	100 %
DATABASE 4	33	0,0425	30	90,91%
DATABASE 5	35	0,0387	33	94,29%
DATABASE 6	35	0,042	33	94,29%
DATABASE 7	27	0,046	26	96,3%

It was found that the major part of the ANN classification resulted in the cluster of *Aedes Aegypti* mosquito eggs. Even though the quantifying of egg clusters has already been considered in earlier works [6-8], as the objective of this approach is to provide a new strategies to validate the algorithm developed, besides making improvements in its performance, currently the efforts made are being invested in broadening knowledge of the ANN PMC to identify the limit between eggs in clusters and making the automated counting of eggs in ovitraps viable.

Acknowledgements

This research was partially sponsored by FINEP-Brazil. The authors would like to thank Aggeu Magalhães Research Center for the concession of the ovitraps used in this study. The first author would like to thank the FINEP-Brazil for a financial research grant while completing her Master's degree, and the second author would like to thank the DPP-UnB and FAP-DF for financial support to present this work.

References

- [1] WHO. *Dengue*. <http://www.who.int/denguecontrol/en/index.html>, 2010.
- [2] Braga I. A., Valle D.: *Aedes aegypti: surveillance, monitoring of resistance and alternative control in Brazil*. In Portuguese: *Aedes aegypti: vigilância, monitoramento da resistência e alternativas de controle no Brasil*. Epidemiol. Serv. Health, Brasília, vol.16, n.4, 2007.
- [3] Câmara Neto, H. F., Augusto, L. G. S.: *Sanitary Conditions of the Urban Environment and the use of household pesticides: implications for health*. In Portuguese: *Condições Sanitárias do Ambiente Urbano e o uso de Pesticidas Domésticos: implicações para a saúde*. In: XXVII International Congress of Sanitary and Environmental Engineering, Porto Alegre Book of Abstracts: ABES-AIDIS, vol.1, 2000.
- [4] S. Haykin., *Neural Networks: A Comprehensive Foundation*, 2nd. Ed., USA, Prentice Hall, 1999.
- [5] R. C. Gonzalez and R. Woods, *Digital Image Processing*. Upper Saddle River, New Jersey: Prentice-Hall, Edinburgh, 2008.
- [6] F.G.G. Elpídio et al., *Automatic Identification of Aedes Aegypti Eggs Deposited in Ovitrap Slides Using Image Processing Techniques*. In: XXII Brazilian Congress of Biomedical Engineering, 2010.
- [7] F.G.G. Elpídio et al., *Automatic Counting of Aedes Aegypti Eggs Deposited in Ovitrap by Algorithm of Digital Image Processing and Artificial Neural Network*. In: Pan American Health Care Exchanges, Rio de Janeiro. Pan American Health Care Exchanges (PAHCE), 2011.
- [8] L.F.R. Costa et al., *Processing Images of Aedes Aegypti Eggs to Control Dengue Disease*. In: Congress of Numerical Methods in Engineering, Coimbra. Proceedings of the Congresso de Métodos Numéricos em Engenharia - CMNE 2011, 2011.

ANEXO 5: ARTIGO PUBLICADO NO ANAIS DO WC2012

AUXILIARY SYSTEM FOR COMPUTER-AIDED EVALUATION OF BREAST CALCIFICATIONS BASED ON DIGITAL IMAGE PROCESSING AND ARTIFICIAL NEURAL NETWORKS

F.G.G. Elpídio¹, L.M.Brasil¹, J.M.Lamas² and C.J.Miosso¹

¹ University of Brasília at Gama (UnB /FGA) / Post-Graduation Program of Biomedical Engineering, Gama, Brazil

² Janice Lamas Clinical Radiology / Brasília, Brazil

Abstract— Early diagnosis still represents the best approach in the prevention and control of breast cancer, the second most frequent form of cancer worldwide. In this context, mammography has been largely used as a major method for disease early detection, as it aids the early identification of calcification clusters that can be or can later become tumors. This paper discusses the use of Wavelet Transforms to help highlight calcification areas in combination with digital image morphological techniques for feature extraction from regions of interest in digital mammography. The results obtained in this work are being upgraded to a next step, which will use a MultiLayer Perceptron artificial neural network to classify calcifications according to the most applicable category in the Breast Imaging Reporting and Data System (BIRADS).

Keywords— Breast cancer, Wavelet Transform, Artificial Neural Network Multilayer Perceptron, Breast Imaging Reporting and Data System (BIRADS).

INTRODUCTION

Breast cancer primarily affects the female population. Among nine women who live at least 85 years, one is expected to develop breast cancer [1]. Also, the number of breast cancer occurrences has increased in recent decades, and, according to the World Health Organization (WHO), it can reach the 15 million mark in 2020 [2].

Calcifications (calcium build-ups) in breast regions are frequently found in breast screening examinations [3]. Although most changes characterize benign formations [4], there is a large proportion of cancers that start with clusters of calcifications [5]. It is estimated that 90% of nonpalpable intraductal cancers are detectable morphologically by the presence of calcifications. They are associated with approximately 40% of all tumors [6].

Classifying calcification clusters as benign or malignant is a complex task, often requiring a biopsy (removal of a fragment of tissue for microscopic analysis) before a

definitive conclusion. According to [7], half of the biopsies are performed in order to investigate suspicious breast calcifications. Despite the high probability of detecting cancer in calcification regions that are suspicious for malignancy, some patients undergo this procedure just to find later that no cancer was present. In such cases, they face a number of problems [8] (psychological, financial, familial, social, etc), that could be avoided if there was a more precise analysis of the examination before surgery [9]. Mammography, one of the most sensitive tests for breast cancer screening in asymptomatic cases, enables the detection of breast cancer in an early stage and not diagnosed by clinical examination [10]. However, the detection of breast cancer in asymptomatic women depends on multiple factors, including the quality and form of acquisition of mammography, the way images are processed and the experience of the professional responsible for the interpretation of mammographic images [11].

The Breast Imaging Reporting and Data System (BIRADS) aims at standardizing mammographic reporting in order to reduce differences in the subjective interpretation of mammographic images and to facilitate the control of the results [12].

In medicine, the use of computational techniques Computer-Aided Diagnosis (CAD) is widespread [13, 14]. Because they are computational tools, the CAD systems are not affected by fatigue and pre-conception, factors that are intrinsic to human beings and which can influence the results [15].

Many studies in the literature [16-18] have reported the effectiveness of Wavelet Transform (WT) use [19] to detect breast calcifications. In fact, after the wavelet decomposition a mammographic image into components associated to different frequency bands, one observes that the calcifications generally concentrate on the coefficients corresponding to particular bands. In parallel, the Artificial Neural Networks (ANNs) have been frequently used in

pattern recognition and classification problems of due to their ability to learn by example [20-22].

This paper addresses the use of WTs combined with ANNs – specifically, the MultiLayer Perceptron (MLP) – in order to identify and classify calcifications according to the BIRADS categories. This study is a collaboration between the University of Brasília at Gama (FGA/UnB) and Janice Lamas Clinical Radiology (JLCR). It is also, part of a greater extension project entitled "3D Anatomical Atlas Applied to Breast", with the National Laboratory for Scientific Computing of Rio de Janeiro (NLSC/RJ).

MATERIALS AND METHODS

The algorithms developed for aid in the evaluation of breast calcification was implemented with Matlab software version 7.6 and PHP language version 5.3.8 with Apache service version 2.2.21. The sample used in this work consists of 16 mammograms containing calcifications cases previously diagnosed at JLCR. Each exam consists of about 5 different images, including a Skull projection Flow (SF), the projection a Middle Lateral Oblique (MLO) and a Spot with selective compression and expansion [10].

In this work, the analysis and classification of the calcifications follow the flowchart shown in Figure 1.

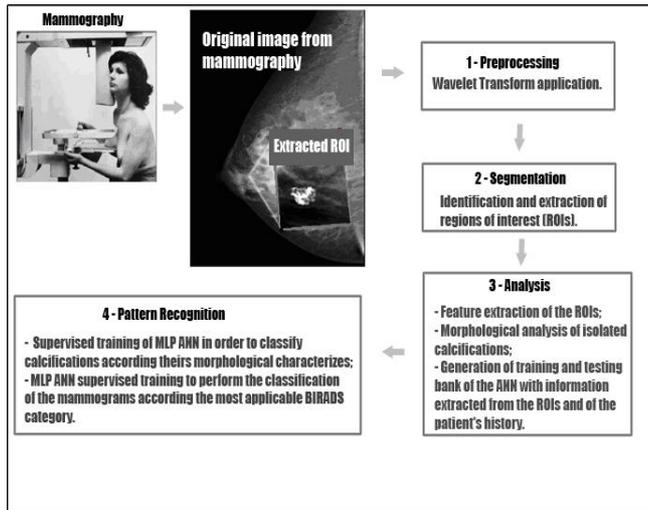


Fig. 1 Flowchart of the steps used in the calcifications analysis and classification based on mammograms.

Note that, after the images are acquired, the pre-processing stage applies the WT in order to automatically detect and highlight areas. The Fast Wavelet Transform (FWT) algorithm deals with the decomposition of a discrete signal through a set of filters that extract the average characteristics (approximations) of the signal, as well a set

of filters, based on wavelets functions, responsible for feature extraction of high frequency (details) of the signal [19]. In practice, this decomposition process consists of applying high-pass filters and low-pass through multiple decompositions. Figure 2 illustrates how these concepts were adopted in this experiment.

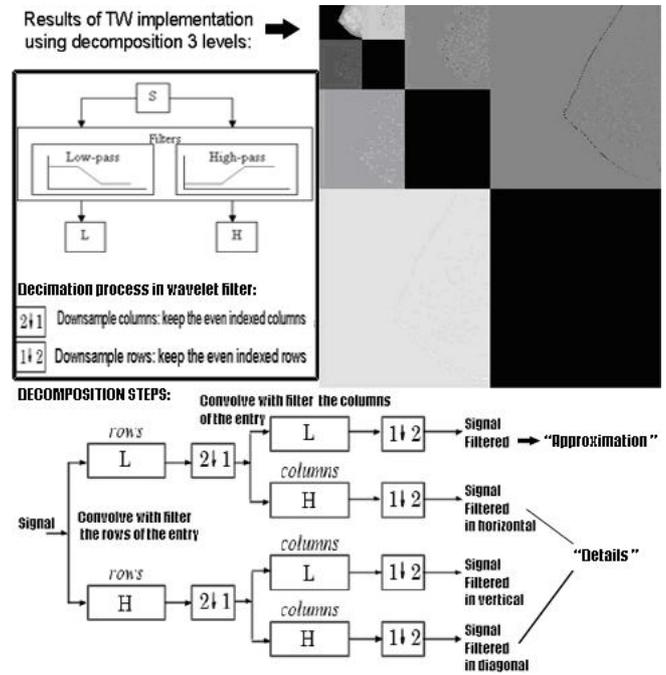


Fig. 2 Computation of the WT of a mammogram. The WT is later used to aid the detection of calcification regions.

After applying the WT to aid the identification of Regions of Interest (ROIs) [20, 21], where the calcifications are concentrated, the next phase (step 2) shown in Figure 1 performs information extraction and segmentation of the ROIs. The objective of the step 2 is to generate a database that will be used by the MLP to perform the ANN BIRADS classification.

At this stage we developed an algorithm to generate images that simulate benign and malignant calcifications in the image. The objective is to create a controlled test environment that consists of images simulating ROIs of real mammograms. This set of images was divided into two subsets; the first includes ROIs similar to calcifications suspicious for malignancy, whereas the second characterizes benign calcifications. The test ROIs were designed simulating intermediate calcifications, which are difficult to characterize as benign or malignant.

One of the criteria used to differentiate the two groups is that the ROIs characterized as benign in the experiment were distributed more openly and sparsely, while the group

characterized as suspicious of malignancy was generated in more dense groups. A feature extraction stage analyzes the ROIs in order to distinguish them and divide them into two groups. The information extracted from the ROIs are:

The normalized image histogram: Used to extract estimates of the ROI [23].

The image sample variance: Allows one to calculate the level of dispersion (proximity or distance) between objects that simulate calcifications in the image [23].

The number of elements: The number of objects that simulate calcifications in the image.

The Euler number: A measure of image topology that allows us to analyze the total number of objects in the image and the level of discontinuity between them, information related to the type of clusters in the picture [24].

In order to analyze ROIs of real mammographic images, we also intend to use, in addition to the information above, extracted from the ROIs, a set of data extracted from the history of patients, such as age, race, lifestyle, emotional state, age at which the menarche occurred, age at which the first pregnancy occurred, age at which the menopause started, correlated treatments (removal of uterus or ovaries, cancer treatment, treatment of depression, use of hormone therapy), family history related to breast cancer and smoking history. These data comprise the patient's history and are considered by many experts, including JLCR (partner of this research), to allow for a more effective diagnosis.

Regarding the ANN topology we opted for networks fed directly with a hidden layer. The standard implementations of MLP ANN in this research are feedforward networks with backpropagation learning algorithm [22]. In this experiment they comprise a hidden layer and one output neuron (responsible for assigning the most applicable BIRADS category). The input vector length equals the number of parameters that are extracted from the ROIs.

The identification and extraction of ROIs from the real mammographic images are currently being performed manually. We developed an interface that allows the expert to indicate the ROI before the digital image processing stage.

For the ROIs classification, we use two MLP ANNs. The first ANN uses a set of parameters that seek to distinguish a benign calcification from a suspicion of malignancy. These parameters include:

Type of calcification: Is assigned a value from 0 to 11 corresponding to the 12 possibilities in the first classification in [12]: skin, blood vessels, popcorn-shaped, rod-shaped, acinar on eggshells, calcium in milk, dystrophic, sutures, punctate, heterogeneous or branched.

Form: Is assigned the value 0 corresponding to a regular form or 1, corresponding to an irregular form.

Distribution: Is assigned a value from 0 to 5, corresponding to the six possibilities in the second classification in [12]: grouped, linear, segmental, diffuse, regional or multiple groups.

Size: Is assigned the value 0, for small, 1, for medium, or 3, for large.

The second ANN MLP will be responsible for classifying the ROI as a whole. The idea is to initially pre-process the ROI by applying the WT. Next, the system will use the above calcification categories, together with other parameters (distribution of most of the calcifications, histogram of the ROI, the ROI level of stability and patient's lifestyle) to indicate the most applicable BIRADS category. The level of stability variation indicates the degree at which the calcifications change during follow-up examinations (if any). This parameter is translated into a MPL input according to three possibilities: 0 - no change, 1 - small changes or 2 - big changes. The lifestyle parameter, on the other hand, is extracted from the patient's medical history and is represented in the ANN by 1 (normal) or 2 (conflicting).

DISCUSSION

The information mentioned above, extracted from the ROIs, generates a database used to train an MLP ANN in order to perform automatic classification of ROIs. Currently, efforts are invested in developments of this experiment, in order to consider morphological information that characterizes the shape and size of the and hence to approximate the test conditions to real calcifications.

In tests regarding the application of the WT in digitized mammograms, it was established, as reported in previous studies [16, 17], that since the calcifications are related to high frequencies they become more evident in the sub-bands corresponding to high frequency components. Therefore, it becomes easier to analyzer and to segment a mammographic image showing the ROI after the implementation of the WT. From this result, is in testing a computer code that was developed to segment the image digital mammography. With this algorithm is possible to divide the image by the histogram in frequency to show the ROI.

CONCLUSIONS

This proposal is intended to provide a system for aiding the evaluation of breast calcifications. This methodology is being developed based on digital signal processing

techniques (wavelet transforms and image morphology) and artificial neural networks. Currently the results are being validated in a controlled test that simulates real calcifications. The next steps aim to consolidate the research methodologies to increase their effectiveness in real digital mammography.

ACKNOWLEDGMENT

The project is being supported by edict number 15/2008 - MCT/CNPq/FNDCT/CAPES/FAPEMIG/FAPERJ/FAPES, Institutos Nacionais de Ciência e Tecnologia (National Institutes of Science and Technology) and INCT/MACC. The first author would like to thank the CAPES-Brazil for a financial research grant while completing her Master's degree and the second author acknowledges the research support of the CNPq by PQ 2F scholarship. The authors are also grateful to JLCR for providing the mammograms used in this research.

REFERENCES

- Kopans, D. B. (2000) Imagem da mama. 2a ed. Rio de Janeiro: Revinter Medsi.
- WHO (2011) Breast cancer. Available at: <http://www.who.int>.
- Nappi, J. et al. (1999) Computerized diagnosis of breast calcifications using specimen radiography and simulated calcifications. In: SPIE Conference on Image Processing, v.3661, pp 1351-1360.
- Leung, J. W.; Sickles, E. A. (2007) The probably benign assessment. *Radiol Clin North Am.* 45(5):773-89.
- Libertman, L. et al. (1998) The Breast imaging report and data system: positive predict value of mammographic features and final assement categories. *AJR*, vol. 171, pp 35-40.
- Lamas, J. M. (1998) Prevalencia e fatores de risco de câncer de mama e de lesões pré-malignas em mulheres assintomáticas no Distrito Federal., Dissertação (Mestrado), Curso de Pós Graduação em Medicina, Universidade de Brasília, Distrito Federal, pp 27-29.
- Kouskos, E. et al. (2006) Wire localisation biopsy of non-palpable breast lesions: reasons of unsuccessful excision. *Eur J Gynaecol Oncol.* 27(3):262-6.
- Biffi, R. G. (2003) A dinâmica familiar de um grupo de mulheres com câncer de mama. Tese de Doutorado, Escola de Enfermagem de Ribeirão Preto, USP.
- Martins, L. A. L. M.; Barra, A. A.; Lucena, C. E. M. (2010) Microcalcificações Mamárias Suspeitas de Malignidade. vol.56(2): 251-258, *Revista Brasileira de Cancerologia*.
- Queiroz, R. Y.; Hummel, A. C.; (1995) O valor das microcalcificações agrupadas no diagnóstico precoce do câncer de mama. *Revista brasileira de cancerologia*. vol. 41(1): 29-31.
- Lamas, J. M. (2000) Avaliação dos resultados de exames mamográficos para detecção precoce do câncer de mama, no Distrito Federal, segundo indicadores de qualidade. Tese de Doutorado, Faculdade de Medicina do Rio de Janeiro, Rio de Janeiro.
- ACR (2005). BI-RADS: Sistema de laudos e registro de dados de imagem da mama. 1. ed. São Paulo: American College of Radiology.
- Turgay, A. et al.(2010) Computer-aided diagnostic models in breast cancer screening. *Imaging Med.* vol.2(3): 313–323, June.
- Cheng, H. D. et al. (2009) Automated breast cancer detection and classification using ultrasound images: A survey. *ACM*, Available online 30 May.
- Zheng, B. et al. (2004) Method to test the reproducibility and to improve performance of computer-aided detection schemes for digitized mammograms. *Medical Physics*, vol. (31):11, 2964-2972.
- Rezai-Rad, G. et al. (2005) Detecting microcalcification clusters in digital mammograms using combination of wavelet and neural network. *Computer Graphics, Imaging and Vision: New Trends*.
- Jasmine, J. S. L. et al. (2009) Microcalcification detection in digital mammograms based on wavelet analysis and neural networks. *Control, Automation, Communication and Energy Conservation, INCACEC 2009, Perundurai, Tamilnadu*.
- Lashkari, A. (2010) Full automatic micro calcification detection in mammogram images using artificial neural network and Gabor wavelets. *Machine Vision and Image Processing (MVIP)*, 2010 6th Iranian.
- Mallat, S. A (2009) *Wavelet tour of signal processing*. 2ed. Elsevier.
- Elpídio, F.G.G, et al. (2011) Automatic Counting of Aedes Aegypti Eggs Deposited in Ovitrap by Algorithm of Digital Image Processing and Artificial Neural Network. In: Pan American Health Care Exchanges, Rio de Janeiro. Pan American Health Care Exchanges (PAHCE).
- Elpídio, F.G.G.; Brasil, L.M.; Silva, A.P.B. (2011) Using Artificial Neural Networks to Identify Aedes Aegypti Eggs in Ovitrap. 15th International Symposium on Applied Electromagnetics and Mechanics (ISEM 2011), Setember, Napoli, Itália.
- Haykin, S. (2009) *Neural Networks: A Comprehensive Foundation*, 2nd. Ed., USA, Prentice Hall.
- Gonzalez R.C., Woods R. (2008), *Digital Image Processing*, Upper Saddle River, New Jersey: Prentice-Hall, 3rd edition.
- Nivaldo, B. F. (1997) Algoritmos de afinamento 3-d: uma análise teórica e prática. Dissertação (Mestrado), Instituto de Computação, Universidade Estadual de Campinas, São Paulo, pp 7-11.

Author: Fátima Gisele Gomes Elpídio
Institute: University of Brasília (UnB)
Street: Campus Universitário Faculdade Gama
City: Gama
Country: Brazil
Email: fgizele@hotmail.com

ANEXO 6: ARTIGO PUBLICADO NO ANAIS DO PAHCE-2012

Abstract — Calcifications (calcium build-ups) in breast regions are frequently found in breast screening examinations. It is

Morphological analysis for feature extraction and classification of breast calcifications

F.G.G. Elpídio¹, L.M.Brasil¹, J.M.Lamas², C.J.Miosso¹

¹University of Brasília at Gama (UnB /FGA), Post-Graduation Program of Biomedical Engineering, Gama, Brazil

estimated that 90% of nonpalpable intraductal breast cancers are detectable morphologically by the presence of calcifications. This work performs a morphological analysis of regions of interest related to breast calcifications by extraction of a set of features that help the health care provider in the automatic classification of these calcifications. The results are being validated in a controlled test that simulates calcifications in order to consider morphological information that characterizes the geometrical characteristics. This allows one to assign an object signature and hence to approximate the test conditions to real calcifications.

Keywords — Breast Cancer, Calcifications, Breast Cancer, Morphological analysis

I. INTRODUCTION

Breast cancer is considered a major cause of female mortality [1]. Early diagnosis of the disease significantly increases the chances of successful treatment. For this control it is crucial to determine if an existing lesion is benign or malignant [2].

Mammography, one of the most sensitive tests for breast cancer screening in asymptomatic cases, enables the detection of breast cancer in an early stage, when it is still not diagnosed by clinical examination [3]. However, the detection of breast cancer in asymptomatic women depends on multiple factors, including the quality and type of mammography acquisition, the way images are processed and the experience of the professional responsible for the interpretation of mammographic images [4].

Classifying calcification as benign or malignant is a complex task, often requiring a biopsy (removal of a fragment of tissue for microscopic analysis) before a definitive conclusion. According to [5], half of the biopsies are performed in order to investigate suspicious breast calcifications. Despite the high probability of detecting cancer in calcification regions that are suspicious for malignancy, some patients undergo this procedure just to find later that no cancer was present. In such cases, they face a number of problems [6] (psychological, financial, familial, social, etc), that could be avoided if there was a more precise analysis of the examination before surgery [7].

In medicine, the use of computational techniques – Computer-Aided Diagnosis (CAD) – is widespread [8, 9]. Because they are computational tools, the CAD systems are not affected by fatigue and pre-conception, factors that are intrinsic to human beings and which can influence the results [10].

This paper discusses the extraction of morphological characteristics of mammography, which are relevant to the diagnosis [11, 12] of breast calcifications, in a database controlled trial to enable the future classification of real breast calcifications. This study is a collaboration between the University of Brasília at Gama (FGA/UnB) and Janice Lamas Clinical Radiology (JLCR). It is also, part of a greater extension project entitled “3D Anatomical Atlas Applied to Breast,” with the National Laboratory for Scientific Computing of Rio de Janeiro (NLSC/RJ).

II. METHODOLOGY

Using the development platform of Matlab (version 7.6), we created an algorithm that generates images simulating real breast calcifications, which sought to synthesize a set of characteristics that distinguish malignant from benign calcifications [13], as can be seen in Table 1.

TABLE I

MORPHOLOGICAL CHARACTERISTICS OF BENIGN AND MALIGNANT BREAST CALCIFICATIONS

Features	Benign	Malignant
Density	uniformly dense	thin
Spatial arrangement	Diffuse, without pattern	Grouped in a restricted

From the images that simulate calcifications we tested the efficiency of extraction of morphological features [11-12, 14]. These features are used to analyze and discriminate the images by dividing them into two subsets, which represents a ROI (Region Of Interest) similar to calcifications suspicious for malignancy and the other characterizing benign calcifications. The parameters used to perform the classification were:

Compactness (C) [11-12]: it combines the perimeter (p) and area (A) to establish a measure of compression according to

$$C = \frac{p^2}{4\pi A}. \quad (1)$$

This value is a minimum for a circle and increases as the degree of irregularity of the border.

Fractional concavity (FC) [11-13]: measures the relationship between the cumulative length of the concavity of the contour and the contour length. This value is usually low for benign calcifications because they tend to have a rounded or oval shape, as in malignant calcifications, where the outline is elliptical or lobulated estimated to be located several hollows.

Spiculation index (SI) [12] – is calculated based on a polygonal model in conjunction with the contour length of the segment, the base widths and angles of possible spicules. Calcifications suspicious for malignancy usually tend to have a more irregular shape [15] resulting in a high SI value.

Eccentricity (E) [11-12] – defines the spatial distribution between the axes of the object and is calculated by the ratio between the smallest and largest horizontal or vertical axis, according to

$$E = \frac{(\mu_{02} - \mu_{20}) + 4\mu_{11}}{A}, \quad (2)$$

where A is the area of the object and μ_{pq} refers to the central moments given by

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q, \quad (3)$$

with $p+q > 1$ and where (\bar{x}, \bar{y}) are the center of gravity of the object under study, and M and N are respectively the width and height of the object.

Circularity (CC) [11-12] – defines how close the shape of the object approaches the shape of a circle. It is calculated by evaluating the region encompassing the object with no peaks or contour defects, using

$$CC = \frac{4\pi A}{p^2}, \quad (4)$$

where A is the area of the object and p is the convex perimeter. If the calculation has an approximately circular shape, the CC is closer to 1.

Circular density (D_c) [11-12] – defines the percentage of intersection between the object being analyzed and a circle with the same area and same center of mass, according to

$$D_c = \frac{n}{A}, \quad (5)$$

where A is the object's area and n is the total number of points belonging to the object and which have the same center of mass. The D_c displays values close to zero for elongated objects and values close to 1 for circular objects.

Acutance (A) [11-12] – measures the change in density along the edge's contour. The calculation is defined as a line of pixels perpendicular to each pixel of the contour, in the case of a hollow structure with spiculated or limited, is employed as many pixels as possible to avoid the line of pixels along the perpendicular to the contour does not cross more than once.

III. RESULTS

Figure 1-a presents a set of skin calcifications [15] where 1-b and 1-c are extended fictional calcifications that were generated in this study. As they are in a controlled test, then they can be used in tests to represent both malignant as benign calcifications, for that will change the parameters listed in Table 1 according to the objective of the experiment.

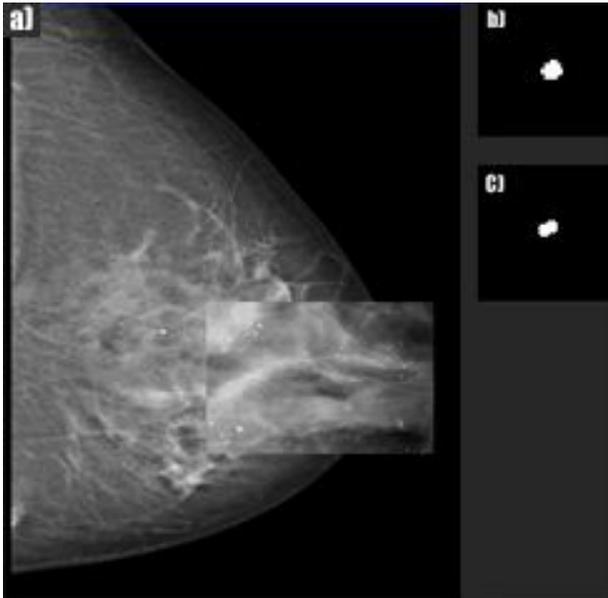


Fig. 1. 1-a is an example of skin calcifications (picture provided by Janice Lamas Clinical Radiology). 1-b and 1-c are examples of calcifications generated by the algorithm developed in this experiment.

Figure 2 presents a set of calcifications referred to as "popcorn-shaped" [15]. To represent this kind of shape it was generated the image 2-b, which may also represent other types of calcifications by changing of parameters like SI, CC and FC.

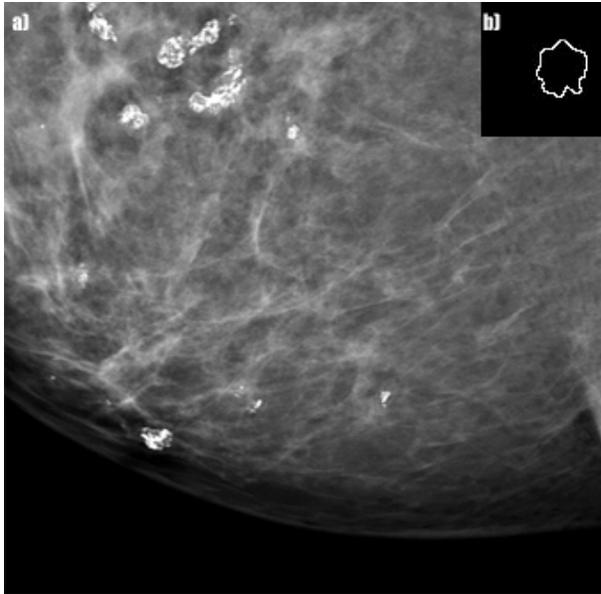


Fig. 2-a is an example of calcifications "popcorn-shaped" (picture provided by Janice Lamas Clinical Radiology). 2-b is an example of calcification generated by the algorithm developed in this experiment to represents this kind of calcification.

Figure 3-a illustrates a set of amorphous calcifications suggestive of malignancy where some of this calcifications are "rod-shaped" [15] and Figure 3-c and 3-b are calcifications generated by the algorithm developed in this research to illustrate some of this type of calcification. Similarly to the previous examples, the fictitious calcifications originated in this example are sufficiently versatile to be used in other tests representing other types of calcifications, for example, vascular calcifications, sutures calcifications, duct calcifications, etc.

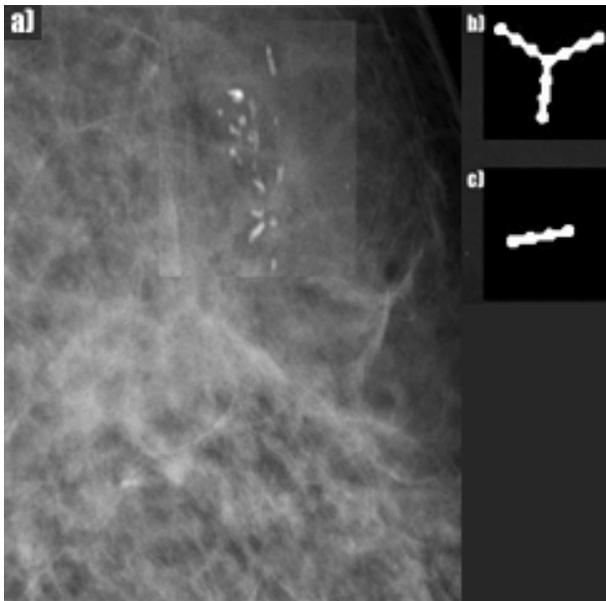


Fig. 3-a is an example of amorphous calcifications suggestive of malignancy (picture provided by Janice Lamas Clinical Radiology). 3-b and 3-c are examples of calcification generated by the algorithm developed in this experiment to represents some of this kind of calcification.

IV. DISCUSSION

The algorithm of image database generation to represent real calcifications allows one to characterize a series of controlled tests prior to application in real cases. With the five basic shapes that were presented in this work it is possible to represent a vast amount of real calcifications by changing geometrical shapes such as C, FC, SI, CC and Dc. One advantage of artificial calcifications originated in this work is that they don't have noise or artifacts that are common in real images.

This work addresses the individual characteristics that are used to distinguish and characterize calcifications as malignant or benign. However for effective analysis in mammography it is necessary to perform a complete ROI analysis that has calcifications [3, 5]. The information mentioned above, extracted from the ROIs, is currently being used to generate a database that will be used to train an Artificial Neural Network (ANN) Multilayer Perceptron (MLP) [16-18] in order to perform the automatic classification of ROIs. For the processing of ROIs extracted of real mammograms we will adopt, in addition to the information extracted from the ROIs, a set of information extracted from the history of patients, such as age, race, lifestyle, emotional state, age at menarche, age at first pregnancy, age at onset of menopause, treatments correlated (removal of uterus or ovaries, cancer treatment, treatment of depression, use of hormone therapy), family history related to breast cancer and smoking history.

V. CONCLUSION

This proposal is intended to provide a system for aiding the evaluation of breast calcifications. This methodology is being developed based on digital signal processing techniques (image morphology) and artificial neural networks.

The information mentioned above, extracted from the ROIs, generates a database used to train an MLP ANN in order to perform automatic classification of ROIs. Currently, efforts are invested in developments of this experiment, in order to consider morphological information that characterizes the shape and size of the and hence to approximate the test conditions to real calcifications.

Currently the results are being validated in a controlled test that simulates real calcifications. The next steps aim to consolidate the research methodologies to increase their effectiveness in real digital mammography.

ACKNOWLEDGMENT

The first author thanks CAPES-Brazil for a financial research grant while completing her Master's degree. The second author acknowledges the research support PQ 2F. The authors are also grateful to INCT/MACC by the financial grant and to the Janice Lamas Clinical Radiology for providing the mammograms used in this research.

REFERENCES

- [1] WHO (2011, December 17) Breast cancer. Available: <http://www.who.int>
- [2] J. Chakrabarti, A. J. Evans, J. James, I. O. Ellis, S. E. Pinder, and R. D. Macmillan, "Accuracy of mammography in predicting histological extent of carcinoma ductal in situ (DCIS)". *Eur J Surg Oncol.*, 2006 32(10): pp. 1089-92.
- [3] R. Y. Queiroz, and A. C. Hummel, "The value of clustered microcalcifications in the early diagnosis of breast cancer". In portuguese: O valor das microcalcificações agrupadas no diagnóstico precoce do câncer de mama. *Revista brasileira de cancerologia.* vol. 41(1): pp. 29-31, 1995.
- [4] J. M. Lamas, "Evaluation of mamography results for early breast cancer detection in asymptomatic women in Distrito Federal, Brazil". In portuguese: Avaliação dos resultados de exames mamográficos para detecção precoce do câncer de mama, no Distrito Federal, segundo indicadores de qualidade". Ph.D. dissertation, Faculdade de Medicina do Rio de Janeiro, Rio de Janeiro, Brasil, 2000.
- [5] E. Kouskos, G. P. Gui, D. Mantas, K. Revenas, N. Rallis, Z. Antonopoulou, E. Lampadariou, H. Gogas, and C. Markopoulos, "Wire localisation biopsy of non-palpable breast lesions: reasons of unsuccessful excision". *Eur J Gynaecol Oncol.*; 27(3):262-6, 2006.
- [6] G. Biffi, "Family Dynamics of a group of women with breast cancer". In portuguese: A dinâmica familiar de um grupo de mulheres com câncer de mama. Ph.D. dissertation, Escola de Enfermagem de Ribeirão Preto, USP, Brasil, 2003.
- [7] L. A. L. M. Martins, A. A. Barra, and C. E. M. Lucena, "Breast microcalcifications Suspected Malignancy". In portuguese: Microcalcificações Mamárias Suspeitas de Malignidade vol.56(2): p. 251-258, *Revista Brasileira de Cancerologia.* 2010.
- [8] A. Turgay, U. S. A. Mehmet, Ze Xiu Liu, A. Oguzhan, and S. B. Elizabeth, "Computer-aided diagnostic models in breast cancer screening". *Imaging Med.* vol.2(3): p.313-323, June, 2010, in press.
- [9] H. D. Cheng, S. Juan, J. Wen, G. Yanhui, and Z. Ling, "Automated breast cancer detection and classification using ultrasound images: A survey.ACM", Available online 30 May, 2009.
- [10] B. Zheng, D. Gur, W. F. Good, and L. A. Hardesty, "Method to test the reproducibility and to improve performance of computer-aided detection schemes for digitized mammograms". *Medical Physics.* 2004, vol. (31):11, 2964-2972.
- [11] O. L. Mangasarian, "Multi-surface method of pattern separation. Street". W.N., Wolberg W.H. and Mangasarian. O.L., "Nuclear Feature Extraction For Breast Tumor Diagnosis", *International Symposium on Electronic Imaging: Science and Technology*, vol. 1950, pp. 861-870, San Jose, CA, 1993, in press.
- [12] R. M. Rangayyan, N. R. Mundigonda, and J. E. L. Desautels, "Boundary modelling and shape analysis methods for classification of mammographic masses". *Medical and Biological Engineering and Computing*, vol. 38, pp. 487-495, May 2000.

- [13] Montoro. "Prevention and Detection of Breast Cancer". In portuguese: Prevenção e Detecção do Câncer de Mama. McGraw-Hill do Brasil, pp. 147-163, São Paulo, 1979.
- [14] Braz. "Identification of Masses in Mammograms using Texture Geometry and Algorithms for Clustering and Classification". In Portuguese: Identificação de massas em mamografias usando textura, geometria e algoritmos de agrupamento e classificação. Graduation dissertation, Departamento de ciências da computação. Universidade Federal do Maranhão. São Luís, Brasil, 2006.
- [15] ACR. BI-RADS: Breast Image Reporting and Data System. 1. ed. São Paulo: American College of Radiology, 2005.
- [16] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd. Ed., USA, Prentice Hall, 2009.
- [17] F. G. G. Elpídio, L. F. R. Costa, M. M. Andrade, E. A. Costa Jr., L. M. Brasil and M. A. B. Rodrigues, "Automatic Counting of Aedes Aegypti Eggs Deposited in Ovitrap by Algorithm of Digital Image Processing and Artificial Neural Network". In: Pan American Health Care Exchanges, Rio de Janeiro. Pan American Health Care Exchanges (PAHCE), 2011, in press.
- [18] F. G. G. Elpídio, L. M. Brasil, J. M.Lamas and C. J. Miosso, "Auxiliary system for computer-aided evaluation of breast calcifications based on digital image processing and artificial neural networks". In World Congress 2012 Medical Physics and Biomedical Engineering, unpublished.

ANEXO 7: ARTIGO PUBLICADO NO ANAIS DO CBEB2012

CLASSIFICAÇÃO DE CALCIFICAÇÕES MAMÁRIAS BASEADO NA ANÁLISE MORFOLÓGICA DESCRITA POR BIRADS E Le GAL ATRAVÉS DE REDE NEURAL ARTIFICIAL

F. G. G. Elpídio*, L. M. Brasil*, J. M. Lamas** e C. J. Miosso*

*Universidade de Brasília (UnB/Gama) / Programa de Pós-Graduação em Engenharia Biomédica, Gama, Brasil

**Clínica Janice Lamas Radiologia, Brasília, Brasil

fgizele@hotmail.com

Abstract: Mammography is one of the most sensitive tests on early detection of breast cancer. Calcifications (calcium build-ups) in breast regions are frequently found in breast screening examinations. Although most calcifications are benign, there is a large portion of carcinoma that starts with their clusters. As the distinction between clusters of benign and malignant calcification is a complex task, this paper proposes the extraction of morphological features relevant to diagnosis of calcifications previously described by *Le Gal* and the Breast Imaging Reporting and Data System (*BI-RADS*). The morphological features extracted are used to generate a database that uses an Artificial Neural Network to classify breast calcifications in benign or suspicious for malignancy. The results are validated in a controlled test that simulates clusters of calcifications and uses morphological information extracted to characterize and classify the calcifications.

Palavras-chave: Calcificações mamárias, Rede Neural Artificial, *BI-RADS*, *Le Gal*.

Introdução

Segundo pesquisas do INCA há uma estimativa de 52.680 novos casos de câncer de mama para o ano vigente e previsão de aumento significativo nas incidências para os próximos anos [1]. A eficácia da mamografia na detecção precoce do câncer de mama em uma fase incipiente foi comprovada por ensaios aleatórios controlados nos quais foi possível atingir uma redução da mortalidade absoluta, graças a capacidade do exame no rastreamento de carcinomas *ductais in situ* e cânceres infiltrantes de tamanho reduzido e em estágio mais precoce do que em grupos de controle que não passaram pelo exame [2].

As calcificações mamárias são frequentemente encontradas em exames de rastreamento como a mamografia. Estima-se que 90% dos cânceres inductais não palpáveis sejam detectados morfológicamente pela presença de calcificações, elas estão associadas a aproximadamente 40% de todos os tumores [3]. Nesse contexto, a caracterização adequada das calcificações mamárias [4] torna-se fundamental no diagnóstico precoce do câncer mamário. Atualmente existe uma alta taxa de exames falso-positivos, com aumento do custo total do rastreamento em virtude da necessidade de investigações histopatológicas e, conseqüentemente, cirurgias desnecessárias, as quais poderiam ser minimizadas através de técnicas menos invasivas para ampliar a eficiência da análise precoce das calcificações antes da submissão do paciente a biópsias [5].

Le Gal [6] foi um dos pioneiros na análise morfológica das calcificações mamográficas. Em 1984 ele publicou uma proposta de diferenciação das calcificações em 5 tipos, baseado na histopatologia da Região de Interesse (*Region Of Interest - ROI*). Posteriormente, o *Breast Imaging Reporting and Data System* (*BI-RADS*) [2] foi desenvolvido para normatizar a descrição e manejo dos achados mamográficos entre os especialistas da área médica agregando outros critérios, além da morfologia das calcificações, tais como a densidade, distribuição e agrupamento.

As Redes Neurais Artificiais (RNA) [8-10] têm sido freqüentemente utilizadas na literatura em problemas de reconhecimento e classificação de padrões devido a sua capacidade de aprender por meio de exemplos. Este trabalho aborda a extração de características morfológicas [7] das calcificações mamárias em concordância com a

nomenclatura e os critérios de diferenciação adotados pelo *BI-RADS* e *Le Gal* e utiliza uma RNA *Perceptron* Múltiplas Camadas (PMC) [10] para classificar calcificações mamárias em benignas ou suspeitas de malignidade. Esta pesquisa resulta da colaboração entre UnB-GAMA e a Clínica Janice Lamas Radiologia (CJLR), além de integrar o projeto intitulado “Atlas Anatômico 3D Aplicado à Mama” com o Laboratório Nacional de Computação Científica do Ministério de Ciência e Tecnologia (LNCC/MCT) e Instituto Nacional de Ciência e Tecnologia em Medicina Assistida por Computação Científica (MACC), contribuindo no desenvolvimento de um ambiente educacional *web* para o aprendizado prático-morfológico das estruturas da mama.

Materiais e Métodos

Os dados empregados como amostra nessa pesquisa foram cedidos pela clínica de estudo, CJLR. Nesse trabalho foram adotadas 10 mamografias digitais, cada uma contendo cerca de 5 imagens em média.

A primeira etapa do processo de extração de informações das mamografias digitais consiste na identificação e extração de ROIs relacionadas a calcificações. Atualmente, essa etapa está sendo realizada manualmente com o auxílio de um algoritmo desenvolvido para permitir que o especialista indique a ROI antes da segmentação. A ROI selecionada na mamografia digital, cujo formato inicial é o *Digital Imaging and Communications in Medicine* (DICOM) [11] passa por um processo de limiarização por histograma para transformar a imagem em binária (onde os *pixels* assumem valores de 0 ou 1). Em seguida, realiza-se outra subdivisão para gerar ROIs referentes a calcificações individuais, extraídas de agrupamentos. O processo de extração de ROIs é explanado em detalhes na Figura 1.

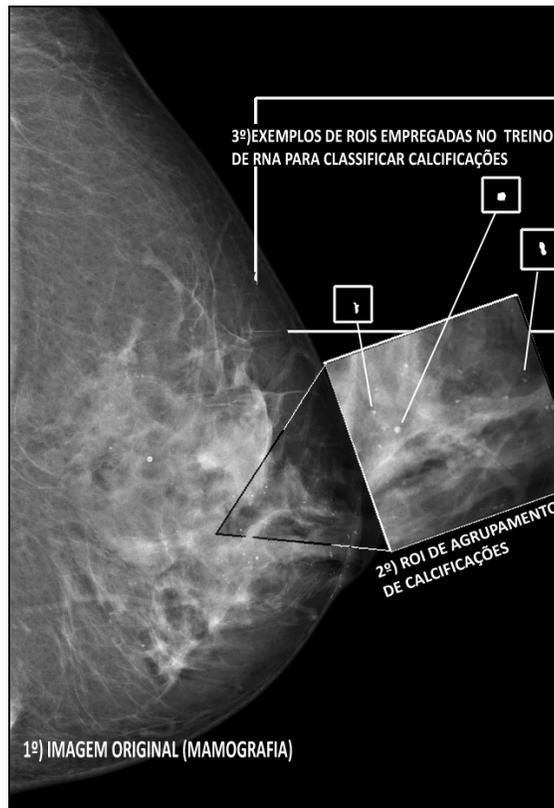


Figura 1: Procedimento para extração de ROIs.

Ao término do procedimento de extração de ROIs referentes a calcificações individuais, foram reservadas 149 amostras as quais foram distribuídas em 4 possibilidades de classificação que estão sendo analisadas individualmente no escopo desse trabalho, conforme a nomenclatura e os critérios de diferenciação do *BI-RADS* [3] e de *Le Gal* [6]. A Figura 2 ilustra o conjunto de classificações adotadas para classificar as calcificações analisadas nesse experimento.

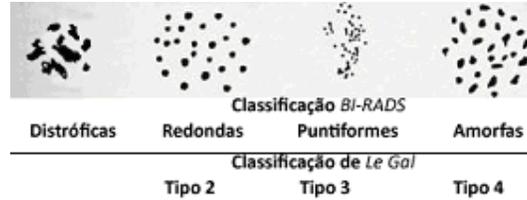


Figura 2: Tipos de classificação de calcificações [6].

Analisando a Figura 2 é possível perceber que a primeira classe de classificação (calcificações distróficas) emprega apenas a metodologia do *BI-RADS*, pois não consta referência direta para a descrição desse tipo no trabalho de *Le Gal*.

Após o procedimento de extração de ROI, cada uma é processada por um algoritmo de extração de características [7] que extrai as seguintes informações:

Área (A) - obtida através da função *bwarea* do Matlab. Ela é útil na análise do tamanho das calcificações. Segundo os estudos [2 e 5], calcificações com tamanho reduzido (menor que 0,5mm) geralmente são associadas a malignidade.

Perímetro (p) das bordas - mensurado através de uma função criada para calcular a soma do resultado da função *bwperim* do Matlab. Este fator é empregado em conjunto com outras características para analisar o aspecto que caracteriza a calcificação (Exemplos: circular, ovular, lobulada, circular).

Variância - calcula o nível de dispersão (proximidade ou distância) entre os objetos que simulam as calcificações na imagem. Esse parâmetro é calculado através da função *find* do Matlab que retorna a disposição dos *pixels* que assumem o valor 1 e 0.

Número de Euler - é uma medida de topologia da imagem que permite analisar o total de objetos presentes na imagem e o nível de descontinuidade entre eles, ou seja, permite analisar agrupamentos de objetos na imagem.

Excentricidade (E) - define a distribuição espacial entre os eixos do objeto. É calculado pela razão entre o menor e maior eixo horizontal ou vertical, conforme a Equação (2):

$$E = \frac{(\mu_{02} - \mu_{20}) + 4\mu_{11}}{A} \quad (2)$$

Nível de compactidade (c) - combina o perímetro (p) e área (A) para estabelecer uma medida de compactação conforme a Equação (3):

$$c = \frac{p^2}{4\pi A} \quad (3)$$

Circularidade (C) - utiliza o perímetro (p) e a área (A) para definir o quão próxima está a forma da ROI com um círculo. É obtida por meio da Equação (4):

$$C = \frac{4\pi A}{(p)^2} \quad (4)$$

Esse conjunto de informações é organizado em um banco de informações que está sendo empregado para gerar uma base de treino de uma RNA PMC [8-9]. Nessa pesquisa será explanado a RNA PMC empregada para classificar as calcificações do tipo Puntiforme/Tipo 3 em benigna ou suspeita de malignidade. Devido a classificação de padrões ser realizada com um único tipo morfológico não houve necessidade de empregar as características relacionadas a forma das calcificações, portanto a RNA foi configurada contendo apenas os

parâmetros que permitem caracterizar a distribuição das partículas de calcificação nos agrupamentos analisados. A Figura 3 apresenta o modelo de RNA PMC adotado nessa pesquisa.

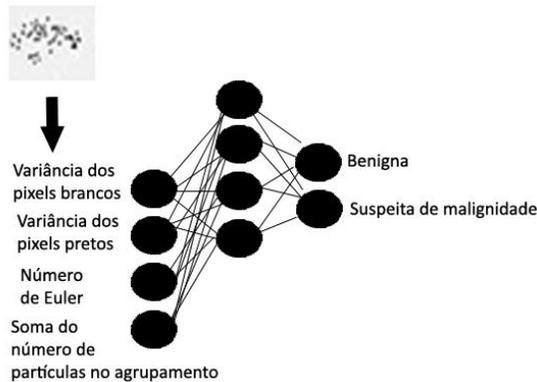


Figura 3: Arquitetura da RNA PMC adotada.

Enquanto a base de treino das imagens reais está sendo consolidada, os testes para discriminar calcificações suspeitas de malignidade de calcificações benignas estão sendo realizados em um banco de testes controlado [7]. Para demonstrar o funcionamento desse experimento foram utilizadas 2.000 ROIs fictícias geradas por um algoritmo [7] desenvolvido para gerar calcificações puntiformes consideradas intermediárias, as quais são difíceis de caracterizar como benignas ou malignas. Um dos critérios adotados para diferenciar os dois grupos é que as ROIs caracterizadas como benignas no experimento foram distribuídas de forma mais espaçadas, enquanto as do grupo caracterizado como suspeitas de malignidade foram geradas de forma mais agrupada, conforme apresentado na Figura 4:



Figura 4: Amostra de imagens geradas no banco de testes controlado desenvolvido para simular ROIs que caracterizam calcificações intermediárias.

Resultados

A RNA PMC utilizada para diferenciar calcificações malignas de benignas foi testada usando o algoritmo *Backpropagation* com diferentes valores das seguintes variáveis: taxa de aprendizagem (η), número de neurônios nas camadas e função de transferência (ϕ) [8]. A Função Sigmóide (FS) e a função Tangente Hiperbólica (TG) foram empregadas em diferentes testes para analisar o desempenho da RNA. Os melhores resultados da RNA PMC foram obtidos com a seguinte topologia: uma η de 0.9, uma camada oculta contendo 4 neurônios e termo *momentum* com valor de 0.9. O número de neurônios da camada de entrada da rede (número de entradas) corresponde ao conjunto de características morfológicas extraídas das ROIs que caracterizam a distribuição das partículas de calcificação nessa pesquisa, portanto a RNA possui 4 neurônios na primeira camada. A camada de saída foi estruturada com 2 neurônios equivalentes a classificação final da ROI (benigna ou maligna). Vale ressaltar que cada neurônio também recebe a influência de um fator *bias* com valor 1. A Tabela 1 apresenta uma amostra da variação dos parâmetros da RNA testada de acordo com a alteração de parâmetros durante o treinamento.

Tabela 1: Variação dos parâmetros da RNA PMC de acordo com topologia empregada durante o treinamento.

η	Φ	Quantidade de épocas	Erro Médio Quadrático	Taxa de acerto
0.7	FS	199	0.01	51%
0.7	TG	127	0.01	47%
0.7	FS	112	0.01	53%
0.9	TG	128	0.01	77%
0.9	FS	500	1.47	79%
0.9	TG	500	1.17	87%
0.9	FS	700	0.32	96%
0.9	TG	700	0.07	98%
0.9	FS	1000	0.25	57%
0.9	TG	1000	0.42	54%

De uma forma geral a RNA apresentou uma convergência rápida (com poucas épocas) durante o treino, porém nos 3 primeiros testes, conforme pode ser observado na Tabela 1, o Erro Médio Quadrático (EMQ) reduzido não foi suficiente para fazer com que a RNA generalizasse seu aprendizado para classificar exemplos distintos dos que foram apresentados no treino. A aplicação de treinamento por um número muito alto de épocas também fez com que a RNA não tivesse um bom desempenho na classificação pós-treinamento. No decorrer do treinamento foi possível avaliar que a RNA apresentou um comportamento melhor com o uso da TG. Os melhores resultados obtidos na classificação foram atingidos com uma η elevada (em torno de 0.9) onde foi possível obter uma convergência rápida da RNA e uma boa generalização do conhecimento, permitindo uma classificação com até 98% de acerto. A Figura 5 ilustra o decaimento do EMQ ao longo do treinamento da RNA PMC com a TG.

Discussão

Esse trabalho aborda a extração de características morfológicas de calcificações individuais. O conjunto de características morfológicas extraídas é utilizado para distinguir e discriminar as calcificações em malignas ou benignas. A etapa de classificação das calcificações em benigna ou suspeita de malignidade é realizada por uma RNA PMC [8-9].

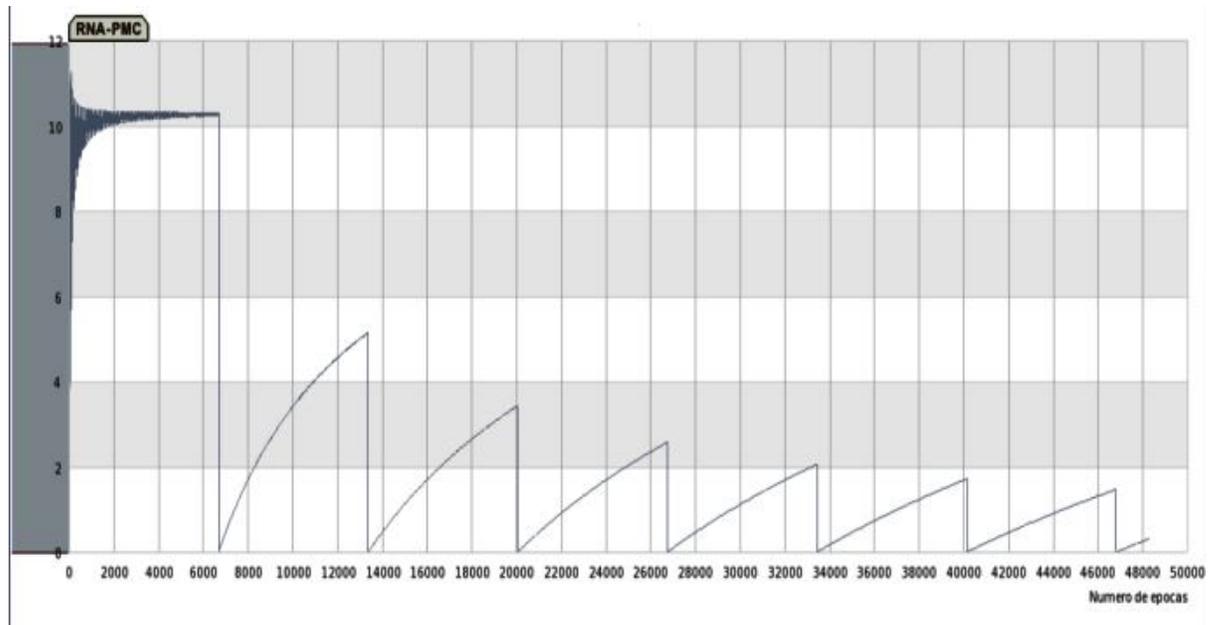


Figura 5: Gráfico do EMQ ao longo do treino da RNA.

O objetivo final da RNA PMC treinada é indicar as calcificações que estão mais relacionadas a processos malignos, distinguindo-as das que estão relacionadas a processos benignos. Dentre o conjunto de classificações analisados, consta na literatura [2;6] que as calcificações puntiformes (Tipo2) e amorfas (Tipo 4) estão mais relacionadas a processos malignos do que os outros tipos analisados nessa pesquisa. A eficiência das informações morfológicas extraídas das ROIS para classificar as calcificações é avaliada através de um banco de dados controlado constituído por calcificações fictícias que simulam as calcificações puntiformes reais [7].

Entretanto, para uma análise efetiva no intuito de estabelecer um diagnóstico da mamografia, torna-se necessário uma análise completa da ROI [9] que apresenta aglomerado de calcificações analisando também o histórico do paciente. Para o processamento das ROIs de imagens mamográficas reais, pretende-se adotar, além das informações extraídas das ROIs, um conjunto de informações extraídas da anamnese das pacientes, tais como: idade, estilo de vida, estado emocional, idade da menarca, idade da primeira gestação, idade de início da menopausa, tratamentos correlacionados (retirada de útero ou ovários, tratamento de câncer, tratamento de depressão, uso de terapia hormonal), histórico familiar relacionado ao câncer de mama e histórico de tabagismo.

O uso de duas metodologias de classificação das calcificações (*Le Gal* e BIRADS) em conjunto estabelece uma classificação de consenso.

Conclusão

Esta proposta visa disponibilizar um sistema para auxiliar na avaliação de calcificações mamárias [9]. A metodologia empregada fundamenta-se em técnicas de processamento digital de imagens para realizar a extração de informações morfológicas das ROIs e uma RNA PMC para classificar os achados em benigno ou suspeito de malignidade. Atualmente, os resultados estão sendo validados em um teste controlado que simula calcificações reais. Em trabalhos futuros pretende-se realizar o mesmo procedimento aplicando os testes nas ROIs das calcificações reais da base de dados que está em desenvolvimento conforme descrito nessa pesquisa.

Agradecimentos

O primeiro autor agradece à CAPES-Brasil pelo financiamento dos estudos realizados no decorrer de seu mestrado. O segundo autor agradece pela bolsa de pesquisa PQ 2F concedida pelo CNPq. Todos os autores são gratos ao LNCC/MCT/MACC pelo suporte financeiro concedido a pesquisa.

Referências

- [1] Brasil (2012), Estimativas do Instituto Nacional de Câncer (INCA). Incidência de câncer no Brasil. Rio de Janeiro. Disponível em <http://www.inca.gov.br>. Acesso em 02 abr. 2012.
- [2] ACR (2005), BI-RADS: Sistema de laudos e registro de dados de imagem da mama. 1. ed. São Paulo: American College of Radiology. ACR (2005), BI-RADS: Sistema de laudos e registro de dados de imagem da mama. 1. ed. São Paulo: American College of Radiology.
- [3] Lamas, J. M. (1998) *Prevalencia e fatores de risco de câncer de mama e de lesões pré-malignas em mulheres assintomáticas no Distrito Federal*, Dissertação (Mestrado), Curso de Pós Graduação em Medicina, Universidade de Brasília, Distrito Federal, p.27-29.
- [4] Leung J. W, Sickles E. A. (2007) The probably benign assessment. *Radiol Clin North Am.*;45(5):773-89.
- [5] Martins, L. A. L. M., Barra, A. A., Lucena, C. E. M. (2010), Microcalcificações Mamárias Suspeitas de Malignidade vol.56(2): p. 251-258, *Revista Brasileira de Cancerologia*.
- [6] Le Gal M, Chavanne G, Pellier D. (1984), Valeur diagnostique des microcalcifications groupées découvertes par mammographies. *Bull Cancer. Paris*, v.71, n(1), p.57-64
- [7] ELPÍDIO, F.G.G. *et al.* (2012) Morphological analysis for feature extraction and classification of breast calcifications. In: *Pan American Health Care Exchanges Conference, PAHCE 2012*, 2012, Florida, v. 39, p. 46-49.
- [8] S. Haykin (2009), *Neural Networks: A Comprehensive Foundation*, 2nd. Ed., USA, Prentice Hall.
- [9] Brasil L. M. *et al.* (2012), Auxiliary System for Computer-Aided Evaluation of Breast. In: *World Congress 2012 Medical Physics and Biomedical Engineering*, 2012, Beijing, p. 876-879.
- [10] ELPÍDIO, F.G.G. *et al.* (2011) Automatic Counting of Aedes Aegypti Eggs Deposited in Ovitrap by Algorithm of Digital Image Processing and Artificial Neural Network. In: *Pan American Health Care Exchanges, PAHCE 2011, Rio de Janeiro*.
- [11] NEMA (2000), *Digital Imaging and Communications in Medicine (DICOM), Part 5: Data Structures and Encoding*. National Electrical Manufacturers Association. EUA.

ANEXO 8: MODELO DE QUESTIONÁRIO PARA AQUISIÇÃO DE CONHECIMENTO DO RBC

O uso de sistemas de apoio à decisão está vinculado a necessidade de auxiliar profissionais a difundir o conhecimento implícito e mapear computacionalmente os processos de raciocínio que levam a tomada de decisão. O RBC modela o conhecimento a partir de uma base de casos que constituem o conhecimento previamente experimentado. Nessa pesquisa está sendo desenvolvido um sistema baseado em conhecimento que se propõe a classificação de calcificações mamárias baseado na análise da mamografia (análise das calcificações, modificadores de distribuição e achados relacionados). O objetivo deste questionário é a classificação de quais características são mais importantes e atribuição de pesos em relação ao sinais/sintomas para a concepção e o desenvolvimento de um Sistema Inteligente baseado em RBC aplicado para auxiliar no diagnóstico e aplicação de categoria BIRADS.

Com o resultado do questionário, o qual o especialista não precisa se identificar, iremos utilizar a média das opiniões dos especialistas para o cálculo de encontrar os casos similares para a resolução do problema e apresentação dos resultados. Classifique as características segundo a ordem de importância, atribuindo um valor de 0 (importância mínima) a 1(importância máxima).

Característica	Valor	Observações
Morfologia das Calcificações		
Distribuição das calcificações		
Achados Associados às calcificações		
Estabilidade das calcificações		
Anamnese da paciente		

Anamnese da paciente (análise dos últimos 3 meses):

Nódulo sólido palpável		
Descarga da papila		
Dor		
Espessamento ou retração da pele		
Grande Linfonodo axilar		
Problema com implante de mama		
Anormalidade na(s) papila(s)		
Presença de outro câncer		
Aumento da Mama		
Redução da mama		
Biopsia		
Tumorectomia de câncer de mama		
Mastectomia		
Possui implante mamário		
Reconstrução da mama		
Retirada de implante mamário		
Submissão a radioterapia		

Anamnese da paciente (implante mamário):

Gel de silicone		
Combinação		
Solução salina		
Pré-peitoral		
Retropeitoral		

Anamnese da paciente (terapia hormonal):

Contraceptivo		
Reposição hormonal		
Apenas estrogênio		
Apenas progesterona		
Estrogênio+progesterona		
Tamoxifeno ou raloxifeno		

Anamnese da paciente (Mestruação):

Não menstrua		
Está menstruando		
Está na menopausa		
Teve menstruação precoce		
Teve menopausa tardia		

Anamnese da paciente (achados associados):

Fibroadenomas calcificados involuídos		
Múltiplas calcificações secretórias		
Lesões com conteúdo de gordura (cistos oleosos, lipomas, galactoceles e hamartomas de densidades mistas)		
Lifonodos intramamários		
Calcificações vasculares		
Implantes		
Distorção arquitetural claramente relacionada a cirurgia anterior		
Nódulo sólido circunscrito não calcificado		
Assimetria focal		
Agrupamento de calcificações redondas (puntiformes)		

Anamnese da paciente (histórico pessoal do câncer de mama/ovariano):

Pessoal		
Irmã		
Mãe		
Filha		
Parente de sangue mais distante		
Idade de ocorrência antes dos 50		
Idade de ocorrência depois dos 50		
Hiperplasia anterior comprovada por biopsia com atipia celular ou carcinoma lobular in situ		

Anamnese da paciente:

Acima de 50 anos de idade		
Estilo de vida/estado emocional		
Nuliparidade		
Não amamentação		
Primiparidade tardia		
Obesidade		
Histórico pessoal de doença mamária benigna		
Histórico pessoal de câncer mamário		
Consumo de álcool		
Tabagismo		
Dieta rica em gordura animal		
Histórico familiar de câncer mamário em um ou mais parentes de primeiro grau		
Exposição a altas doses de radiação		

Tipo de mama da paciente:

Mama lipossubstituída (tecido glandular < 25%)		
Tecido glandular entre 25% e 50%		
Tecido glandular aproximadamente entre 51% e 75%		
tecido glandular > 75%		

Tipo de calcificação:

Cutânea		
Vascular		
Grosseira ou semelhante a pipoca		
Redondas		
Puntiformes		
Centro radiotransparente		
Casca de ovo ou em anel		
Leite de cálcio		
Fios de sutura		
Distróficas		
Amorfas ou indistintas		
Heterogêneas Grosseiras		
Pleomórficas finas		
Finas Lineares ou Lineares ramificadas		

Tipo de distribuição da calcificação:

Agrupada		
Linear		
Segmentar		
Difusa		
Regional		
Grupos múltiplos		

ANEXO 9: MODELO DE PROPOSIÇÃO DE HEURÍSTICAS PARA AQUISIÇÃO E MODELAGEM DO CONHECIMENTO NO SISTEMA DE RBC

Qual a classificação BIRADS aplicável? (Corrigir e/ou colocar observações sobre a classificação BI-RADS mais aplicável segundo alteração dos atributos de morfologia, distribuição, formato, tamanho, achados associados e composição da mama).

Legenda da composição mama	
Atributo	Índice que representa o atributo
Tecido glandular < 25%	I
Tecido glandular entre 25% e 50%	II
Tecido glandular aproximadamente entre 51% e 75%	III
Tecido glandular > 75%	IV

Legenda da distribuição da calcificação na mama	
Atributo	Índice que representa o atributo
Difusas homogêneas	I
Difusas não homogêneas	II
Regionais envolvendo a maior parte ou mais que um quadrante	III
Regionais envolvendo uma porção delimitada do quadrante	IV
Agrupadas	V
Grupos Múltiplos	VI
Segmentares	VII
Lineares	VIII

Legenda do tamanho da calcificação na mama	
Atributo	Índice que representa o atributo
> 2-3mm de diâmetro	I
1cm < x < 1mm	II
x > 0.5mm ou >=1mm	III
x < 1mm	IV
x < 0.5 mm	V

Legenda do formato da calcificação na mama	
Atributo	Índice que representa o atributo
Redondo grande, anelar ou meia-lua	I

Oval	II
Grosso	III
Linear, fino, redondo pequeno	IV
Amorfo, heterogêneo/pleomórfico	V
Redondo grande, anelar ou meia-lua	VI

Legenda de achados associados à calcificação na mama	
Atributo	Índice que representa o atributo
Nenhum	I
Distorção arquitetural claramente relacionada a cirurgia anterior ou trauma	II
Distorção arquitetural	III
Nódulo sólido circunscrito não calcificado	IV
Assimetria focal	V
Lifonodos intramamários	VI
Lesões com conteúdo de gordura	VII
Fibroadenomas calcificados involuídos	VIII

Legenda de classificação morfológica das calcificações na mama		
Atributo	Exemplos	Índice que representa o atributo
Tipicamente Benignas	Cutânea	I
	Vasculares	
	Grosseiras ou semelhantes a pipoca	
	Semelhantes a bastonetes	
	Redondas	
	Centro radiotransparente	
	Distróficas	
	Casca de ovo ou em anel	
	Fios de sutura	
	Leite de cálcio	
Intermediárias	Puntiformes	II
	Amorfas ou indistintas	
	Heterogêneas grosseiras	
Altamente suspeitas	Pleomórficas finas	III
	Finas lineares ou finas lineares ramificadas	

De acordo com as legendas dos atributos acima indique a classificação BI-RADS mais aplicável.

A tabela faz uma análise combinatória de todos os atributos mencionados anteriormente e aplica a classificação BI-RADS conforme a combinação desses atributos e os pesos das características

atribuídas pelo especialista (consulte o Anexo 8 para obter maiores informações dos pesos das características).

Morfologia	Distribuição	Formato	Tamanho	Achados associados	Composição da mama	Classificação BI-RADS
Análise das calcificações benignas						
I	I	I	I	I	I	II
	II					II
	III					II
	IV					
	V					
	VI					
	VII					
	VIII					
	I	II				
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	III				
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	IV				
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	V				
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	VI				

Yellow	II	Orange	I	Blue	Grey	
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I					I
	II		II			
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I					III
II						
III						
IV						
V						
VI						
VII						
VIII						
I	IV					
II						
III						
IV						
V						
VI						
VII						
VIII						
I	V					
II						
III						
IV						
V						
VI						
VII						
VIII						

	I	I	I	I	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	II	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	III	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	IV	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	V	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	VI	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					

	VIII					
	I	I	I	VII	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	VIII	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	I	II	
	II					
	III					
	IV					
	V					
	VI					
	VII					
VIII						
I	I	I	I	III		
II						
III						
IV						
V						
VI						
VII						
VIII						
I	I	I	I	IV		
II						
III						
IV						
V						
VI						
VII						
VIII						
Análise das intermediárias						
II	I	I	I	I	I	
	II					

	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	II			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	III			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	IV			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	V			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	VI			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				

	I	I	I		
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	II			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	III			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	IV			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	V			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	I	
	II				
	III				
	IV				
	V				

	VI				
	VII				
	VIII				
	I	I	I	II	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	III	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	IV	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	V	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	VI	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	VII	I
	II				
	III				
	IV				

	V					
	VI					
	VII					
	VIII					
	I	I	I	VIII	I	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	I	II	
	II					
	III					
	IV					
	V					
	VI					
	VII					
	VIII					
	I	I	I	I	III	
	II					
	III					
	IV					
V						
VI						
VII						
VIII						
I	I	I	I	IV		
II						
III						
IV						
V						
VI						
VII						
VIII						
Análise das calcificações suspeitas						
III	I	I	I	I	I	
	II					
	III					
	IV					
	V					
	VI					

	VII				
	VIII				
	I	II			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	III			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	IV			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	V			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	VI			
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I		
	II	I			
	III				

	IV				
	V				
	VI				
	VII				
	VIII				
	I		II		
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I		III		
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I		IV		
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I		V		
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	I	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	II	I
	II				

	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	III	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	IV	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	V	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	VI	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	VII	I
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	VIII	I

	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	I	II
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	I	III
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				
	I	I	I	I	IV
	II				
	III				
	IV				
	V				
	VI				
	VII				
	VIII				