

UNIVERSIDADE ESTADUAL DO PIAUÍ – UESPI
CAMPUS PROF. ALEXANDRE ALVES DE OLIVEIRA
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

EVANGELINA DE FREITAS VERAS

UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS NA CLASSIFICAÇÃO DE
MICROCALCIFICAÇÕES EM EXAMES DE MAMOGRAFIA

Biblioteca UESPI PMS
Registro, N° _____
CDD 004.65
GUTTER V473u
V _____ EX. 01
Data 28/01/2018
Voto _____

PARNAIBA

2015

**UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS NA CLASSIFICAÇÃO DE
MICROCALCIFICAÇÕES EM EXAMES DE MAMOGRAFIA**

Monografia submetida ao Curso de Bacharelado em
Ciência da Computação da Universidade Estadual do
Piauí, como parte dos requisitos para obtenção do
título de Bacharel em Ciência da Computação.

Orientador: Prof. Dario Brito Calçada

PARNAIBA

2015

V473u

Veras, Evangelina de Freitas

Utilização de redes neurais artificiais na classificação de microcalcificações em exames de mamografia/ Evangelina de Freitas Veras. – Parnaíba: UESPI, 2015.

44 f. : il.

Orientador: MSc. Dario Brito Calçada

Monografia (Graduação em Ciência da Computação) - Universidade Estadual do Piauí, 2015.

1. Introdução 2. Mamografia e o Câncer de Mama 3. Redes Neurais Artificiais 4. Materiais e Métodos 5. Resultados e Discussões 6. Considerações Finais I. Calçada, Dario Brito II. Universidade Estadual do Piauí III. Título

CDD-4.65

EVANGELINA DE FREITAS VERAS

**UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS NA CLASSIFICAÇÃO DE
MICROCALCIFICAÇÃO EM EXAMES DE MAMOGRAFIA**

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual do Piauí – UESPI, Campus Prof. Alexandre Alves de Oliveira, como parte das exigências da disciplina de Estágio Supervisionado, requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Orientador: M.Sc. Dario Brito Calçada

Monografia Aprovada em: 24 de julho de 2015.

Banca Examinadora:



Prof. M.Sc. Dario Brito Calçada
UESPI/Parnaíba – Orientador



Prof. M.Sc. Eyder Franco Sousa Rios
UESPI/Parnaíba – Avaliador



Prof. M.Sc. Danilo Borges da Silva
UFC/Fortaleza – Avaliador

Dedico a meus pais e avós, que foram meus incentivadores desde o primeiro ano de faculdade. Aos animais de estimação que me acompanham nos trabalhos caseiros e, sobretudo a Deus, incentivador espiritual que a todo o momento esteve comigo.

AGRADECIMENTOS

Ao professor orientador Dario Calçada, meus sinceros agradecimentos, pois incentivou muito a realização deste trabalho, também pela amizade e confiança nesse período de trabalho.

Aos colaboradores Renan Fialho, Adrian Chaves, Aline Marcia, Danielle Arruda e Fabiano do Nascimento, que são amigos pessoais e foram coadjuvantes no processo de pesquisa e atuantes do grupo IC-BIOTEC no qual proporcionaram frutos de pesquisa e extensão.

A coordenação do curso de Ciência da Computação da UESPI, pela colaboração nas datas de entrega dos trabalhos e compreensão das atividades curriculares.

Aos professores efetivos do ultimo bloco de curso, pelo incentivo e colaboração das atividades acadêmicas.

“É muito melhor lançar-se em busca de conquistas grandiosas, mesmo expondo-se ao fracasso, do que alinhar-se com os pobres de espírito, que nem gozam muito nem sofrem muito, porque vivem numa penumbra cinzenta, onde não conhecem nem vitória, nem derrota.”
(Theodore Roosevelt)

RESUMO

O câncer de mama está em segundo lugar com maior incidência mundial. A melhor forma de prevenção e controle do câncer de mama 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 caracterização de exames mamográficos é tarefa complexa. O sistema *Breast Image Reporting and Data System* (BI-RADS) normatiza a descrição e a finalidade dos achados mamográficos entre os especialistas da área médica. Este trabalho fundamenta-se em técnicas de Processamento Digital de Imagens (PDI) para realizar a extração de informações morfológicas das calcificações e uma Rede Neural Artificial (RNA) *Perceptron* Múltiplas Camadas (MLP) para classificar as calcificações analisadas de acordo com a classificação morfológica definida pelo BI-RADS. O desempenho da RNA MLP foi avaliado em um teste de aprendizado com calcificações reais estudadas, e validadas, extraídas de mamografias do *Digital Database for Screening Mammography* (DDSM). Em seguida foi treinada a RNA MLP utilizando o algoritmo *Backpropagation* cujo melhor resultado de treinamento dada a generalização do conhecimento, permitindo uma convergência de até 91% de acerto. Os resultados obtidos caracterizam que a presente proposta identificou um processo satisfatório de análise das calcificações mamárias por meio do uso das técnicas de inteligência computacional e PDI.

PALAVRAS-CHAVES: Mamografia. BI-RADS. Calcificações Mamárias. Rede Neural Artificial. Câncer de Mama. Inteligência Computacional

ABSTRACT

Breast cancer is second with greater global impact. The best form of prevention and control of breast cancer is still early diagnosis. Mammography is considered the main method of early detection of breast cancer. Calcifications (calcium accumulation in breast regions) are often found in breast screening tests. The characterization of mammograms is a complex task. The system Breast Image Reporting and Data System (BI-RADS) regulates the description and purpose of mammographic findings among medical experts. This work is based on Digital Image Processing (PDI) techniques to perform the extraction of morphological information of calcifications and an Artificial Neural Network (ANN) Multilayer Perceptron (MLP) to classify analyzed calcifications according to defined morphological classification BI-RADS. The performance of the MLP RNA was evaluated in a learning test with real calcifications studied and validated, extracted from the Digital Database for Screening Mammography (DDSM). Then, the RNA MLP was trained using the backpropagation algorithm whose best result of training given the generalization of knowledge, allowing a convergence of up to 91% accuracy. The results obtained characterize this proposal identified a satisfactory process of analysis of breast calcifications through the use of techniques of computational intelligence and PDI.

KEY-WORDS: Mammography. BI-RADS. Breast calcifications. Artificial Neural Network. Breast cancer. Computational Intelligence

LISTA DE FIGURAS

Figura 1 – Anatomia da mama	16
Figura 2 – Aglomerações de calcificações (A) e Nódulos de forma irregular (B).....	18
Figura 3 – Formas de um carcinoma.....	18
Figura 4 – Margens de nódulos	19
Figura 5 – Neurônio Biológico.....	20
Figura 6 – Neurônio Artificial.....	21
Figura 7 – Rede neural artificial de múltiplas camadas	23
Figura 8 – Rede Perceptron Multicamadas	24
Figura 9 – Etapas aplicadas a metodologia	27
Figura 10 – Imagens de mamografias com seus respectivos ROIs	28
Figura 11 – Descrição do exame de mamografia	28
Figura 12 – Limiarização dos ROIs.....	30
Figura 13 – <i>Interface</i> da ferramenta Matlab® para treinamento da rede neural MLP	31
Figura 14 – Resultados do treinamento da rede neural, contando com 90 amostras de treinamento, apresentou um índice de acerto de 91%.....	33
Figura 15 – Resultado de treinamento, validação e testes da Rede Neural.....	34
Figura 16 – Configuração da Rede Neural MLP treinada.....	35

LISTA DE TABELAS

Tabela 1 – Padrões de densidade da mama estabelecida por BI-RADS.....	16
Tabela 2 – Funções de Ativação.....	24
Tabela 3 – Descrição da forma dos achados na mama	29
Tabela 4 – Descrição das Margens dos achados na mama.....	29
Tabela 5 – Distribuição de Dados.....	32

SUMÁRIO

1 INTRODUÇÃO	11
2 MAMOGRAFIA E O CÂNCER DE MAMA	14
2.1 A MORFOLOGIA DA MAMA	15
2.2 CALCIFICAÇÕES E NÓDULOS.....	17
2.2.1 Descrições de Microcalcificações	18
2.2.1.1 Forma	18
2.2.1.2 Margem	18
2.3 CATEGORIAS E AVALIAÇÕES EM EXAMES CLÍNICOS MAMOGRÁFICOS.....	19
3 REDES NEURAIIS ARTIFICIAIS	19
3.1 REDES PERCEPTRON MULTICAMADAS	22
3.2 FUNÇÕES DE ATIVAÇÃO.....	24
3.3 PROCESSOS DE TREINAMENTO	25
3.3.1 Treinamento supervisionado	26
3.3.2 Treinamento não-supervisionado	26
3.3.3 Treinamento com reforço	26
4 MATERIAIS E MÉTODOS	26
4.1 AQUISIÇÃO DE IMAGENS.....	27
4.2 EXTRAÇÃO DE CARACTERÍSTICAS.....	29
4.3 ORGANIZAÇÃO DOS DADOS	31
4.4 TREINAMENTO DA REDE NEURAL MLP	32
5 ANÁLISE E DISCUSSÃO DOS RESULTADOS	33
5.1 TREINAMENTO DA REDE NEURAL.....	33
5.2 RESULTADOS.....	34
5.3 TOPOLOGIA DA REDE NEURAL.....	35
6 CONSIDERAÇÕES FINAIS	36
REFERÊNCIAS	38
ANEXO A – Dados utilizados na extração de imagens mamográficas	40

1 INTRODUÇÃO

O câncer de mama é o câncer mais frequente entre as mulheres. De acordo com o Instituto Nacional do Câncer (INCA), a estimativa é que 22% de novos casos surjam a cada ano. No Brasil, os dados computados pelo Ministério da Saúde são de 57 casos a cada 100 mil mulheres. A quantidade de registros de mortes resultantes de câncer de mama continua elevada, principalmente porque o diagnóstico da doença é feito em estágio avançado (INCA, 2015).

As causas para o aparecimento do câncer de mama ainda não são totalmente conhecidas. Sabe-se que a mutação celular, hábitos alimentares e influências externas, como o meio ambiente podem levar a um mau funcionamento do organismo, afetando os únicos genes, que são causadores do câncer, ou nos genes supressores de tumor, responsáveis por evitar qualquer desordem nas células e fazer conter crescimento desordenado das células. A melhor medida de prevenção é visitar frequentemente um mastologista, e fazer regularmente o autoexame doméstico, assim como exames clínicos, tais como: mamografia, ultrassonografia, ressonância magnética, punção por agulha fina e biópsia. Quanto mais precoce for detectado a anomalia, mais chances a paciente têm de cura (FOGAÇA e GARROTE, 2004).

A forma mais eficaz para detecção de câncer de mama é a mamografia. O exame é feito a partir de um aparelho denominado mamógrafo, é realizado posicionando a mama no equipamento, sobre uma espécie de placa acrílica dupla. A principal característica do mamógrafo é capacidade de detecção de micro calcificações a partir de 1 milímetro (mm), aumentando as chances de cura em até 95% (FORSYTH, 2003).

A mamografia é considerada o exame de maior confiabilidade para o rastreamento do câncer de mama devido a sua sensibilidade. Porém, a interpretação de lesões numa mamografia é uma tarefa complexa para os especialistas, de cuja experiência depende um diagnóstico preciso. Distorções na interpretação e classificação de lesões suspeitas por especialistas implicam um número maior de biópsias desnecessárias. Entre 65% a 85% das biópsias de mama são realizadas em lesões benignas (ZHENG *et al.*, 2004).

As principais lesões identificadas na mamografia são: as microcalcificações, um dos primeiros indícios de formação de tumores com grau elevado de suspeita de malignidade, e os nódulos, que são responsáveis por grande parte dos casos de câncer de mama e devem ser analisados de acordo com o tamanho, contorno, limites e densidade. As bordas da massa, segundo a *American College of Radiology*, podem ser divididas em cinco contornos: circunscrita, obscurecida, microlobulada, mal definida e espiculada (KOPANS, 2000).

Nas últimas duas décadas diversos trabalhos utilizando técnicas de Processamento de Imagens, Reconhecimento de Padrões, Visão Computacional e Inteligência Artificial vêm sendo desenvolvidos, como parte de esquemas *Computer Aided Diagnosis* (CAD), os quais podem funcionar como uma segunda opinião aos radiologistas quanto à localização e à natureza destas estruturas (DRUKKER *et al.*, 2010).

O laudo dos exames são classificados pelos especialistas em um sistema padrão chamado BI-RADS (em inglês, *Breast Imaging Reporting and Data System*) que foi desenvolvido em 1993 para minimizar as divergências nas interpretações das imagens, e facilitar o controle de resultados (ACR, 2005), este varia de I a V, sendo que I representa risco pequeno e V grande risco. O diagnóstico feito pelos especialistas por meio de exames de mamografia, e baseado em BI-RADS é bem subjetivo, e depende bastante da qualidade das imagens, em que as áreas suspeitas são previamente identificadas como ROI's (*Region of Interests*), que caracterizam as áreas de interesse das imagens.

A complexidade e subjetividade deve-se a qualidade da imagem e característica morfológicas da paciente, ocorrendo muitos falso-positivos em exames clínicos e levando a biópsias desnecessárias. Alterações visuais advindas de fatores externos (mamógrafo, subjetividade, etc) podem alterar os parênquimas da leitura das imagens (LOUVEIRA *et al.*, 2003; KEMP *et al.*, 2001; YAP *et al.*, 2006).

A Inteligência Artificial (IA) destaca-se no âmbito de técnicas de computação descrita em sua literatura como ferramenta inovadora (STAVROS, 2009; JASMINE *et al.*, 2009; KRAUS, 2009; etc.), estudo que vem sendo difundido e vem mostrando mais eficaz que a programação convencional. As RNA's encontram-se como uma área da Inteligência Computacional capaz de classificação de padrões usada em casos de usos e reutilização de seus parâmetros previamente estabelecidos com o objetivo de aprender por intermédio de exemplos (SILVA, 2010).

O objetivo da pesquisa é discriminar os achados nas imagens mamográficas digitalizadas submetidas aos testes utilizando a classificação BI-RADS, potencializando e caracterizando as massas tumorais, utilizando Inteligência Computacional. Consiste em Processamento Digital de Imagem (PDI) com o intuito de extrair as regiões de interesse (ROI's), utilizando a categoria BI-RADS mais aplicável, fazendo uso da Rede Neural Artificial MPL. A pesquisa usou como base de dados o MMSD (*Digital Database for Screening Mammography*) que dispôs de uma vasta gama de imagens mamográficas digitalizadas devidamente classificadas. A finalidade do trabalho seguiu as seguintes disposições:

- Identificar os algoritmos de IA determinantes na pesquisa;
- Caracterizar os pontos das imagens necessárias para extrair os parâmetros;
- Identificar o método mais eficiente para tratamento de imagens digitalizadas;
- Ajustar os parâmetros relevantes para a Rede Neural Artificial;
- Caracterizar os pontos positivos e negativos dos testes.

Metodologicamente, adotou-se uma pesquisa bibliográfica baseada nos estudos morfológicos e clínicos sobre o câncer de mama, além de identificar os problemas relacionados aos exames específicos de mamografia. A problemática partiu inicialmente da necessidade de potencializar a classificação de calcificações nas imagens de mamografia, sendo isso o ponto culminante para a pesquisa. O estudo foi aplicado com a técnica de IA de Redes Neurais Artificiais com o algoritmo MLP, e a partir de um levantamento de requisitos baseado na estrutura morfológica de exames digitalizados extraindo os parâmetros necessários investigados e declarados pela pesquisa literária e de casos específicos. Adotou-se uma segmentação de imagem baseado nas funções do MatLab®, o que resultou nos números de tratamento para a Rede Neural. Os resultados foram classificados de acordo com um índice de acertos de treinamento, teste e validação, comparando os resultados previamente estabelecidos pelo padrão internacional BI-RADS.

O presente trabalho de conclusão de curso (monografia) está organizado da seguinte forma:

Capítulo 2: Mamografia e o Câncer de Mama. Detalham as causas e os estudos sobre as microcalcificações mamárias, métodos de prevenção exames mamografia.

Capítulo 3: Redes Neurais Artificiais - Aborda a teoria aplicada às redes neurais artificiais, explorando as funções e arquitetura da MLP.

Capítulo 4: Materiais e Métodos - Relata a metodologia empregada na execução da pesquisa e as formas como foram utilizadas as ferramentas usadas no tratamento das imagens.

Capítulo 5: Análise e Discussão do resultados. Apresenta os principais resultados obtidos na pesquisa, relatando os pontos de relevância nos resultados.

Capítulo 6: Classifica os principais objetivos alcançados e apresenta sugestões para possíveis trabalhos futuros.

Referências - Fontes de pesquisas que auxiliaram a pesquisa.

2 MAMOGRAFIA E O CÂNCER DE MAMA

O câncer de mama se apresenta mais frequentemente em mulheres e raramente em homens, sendo um dos principais causadores de morte entre as mulheres, segundo Instituto Nacional do Câncer (INCa). De acordo com Costa (2009) é uma anomalia que apresenta crescimento desordenado das células da mama, originando em uma metástase, que se caracteriza por um nódulo palpável e de consistência endurecida, que na maioria das vezes é indolor. Para Brito (2008), as causas são multifatoriais, e acredita ser pouco provável erradicar totalmente o surgimento dessa doença, logo a melhor forma de tratamento são os cuidados periódicos com saúde, e tendo em até 40% de chances de cura se diagnosticado no início.

Os tumores de mama geralmente são descobertos pela própria mulher apalpando eventualmente os seios. Ainda é grande o número de mulheres que vão a um especialista para fazer exames quando os nódulos são detectados em um estágio avançado para responder positivamente ao tratamento (INCa, 2015). As reações ao câncer de mama dependem diretamente da natureza da anomalia e resposta do paciente, sendo que quanto mais tarde o diagnóstico for feito, mais difícil são as chances de cura (BIFFI, 2003). A ocorrência de câncer em mulheres jovens é bastante incomum, segundo os dados do Instituto Nacional do Câncer, são fortemente acometidos em mulheres acima de 40 anos.

As causas oriundas do câncer de mama ainda são bem indefinidas. Sabe-se que o hormônio estrógeno está diretamente ligado ao surgimento de tumores, contudo, são estudados diversos fatores que podem desencadear anomalias mamárias (GUERRA, 2009; SILVA e RIUL, 2011).

Os fatores familiares apresentam históricos comprovados que a ocorrência do câncer de mama em pacientes que tenham precedência na família é recorrente. O grau de parentesco influi consideravelmente probabilidade do surgimento de tumores. Estudos analisaram que há possibilidade maior se dois ou mais parentes de primeiro grau tiverem sido afetados (GUERRA, 2009).

As mulheres entre 40 e 69 são as mais acometidas pelo câncer, isso porque a exposição do hormônio estrogênio está no auge a partir do 50 anos. Estudos corroboram que a incidência de casos entre 40 e 65 anos são mais frequentes, já a ocorrência em mulheres mais jovens resultam em uma fisiopatologia mais agressiva (CASTIGLIONE e AEBIS, 2006).

Excesso de peso é um indício de alerta, pois o tecido gorduroso atua com nova fábrica de estrógeno, isso porque a gordura serve como fonte de produção do hormônio

estrógeno, que em grande quantidade ocasiona uma propagação desordenada de células (SILVA e RIUL, 2011). Há estudos que apontam um dobro de chances de desenvolvimento da doença em mulheres obesas após a menopausa (CAMPOS, 2006).

Pesquisas detectaram que o tabagismo não há relação com o estadismo de carnificina da mama, que faz alguma relação com consumo de bebidas alcoólicas (HAMAJIMA *et al.*, 2003).

Estudos mais detalhados afirmam que o câncer de mama encontra-se relacionada ao processo de industrialização, com risco de enfermidade associada ao elevado status socioeconômico e ao aumento da expectativa de vida, uma vez que a incidência dobra a cada dez anos a mais vividos até a menopausa. Além da idade, vários outros fatores de risco têm sido levantados, entretanto, não existem medidas práticas específicas de prevenção primária do câncer de mama aplicáveis à população. Estudos observacionais têm sugerido que a prevenção do tabagismo, do alcoolismo, da obesidade e do sedentarismo e o incentivo ao aleitamento materno reduzem o risco de câncer de mama. No entanto, os avanços tecnológicos mais importantes têm sido direcionados para o diagnóstico precoce e para o tratamento (PARKIN; BRAY; DEVESSA, 2001; FLETCHER; ELMORE, 2003).

Para a detecção precoce do câncer de mama, utiliza-se o exame de mamografia realizado pelo aparelho mamógrafo, que por sua vez é considerado de alta sensibilidade, podendo identificar carcinomas de 0.5 mm de tamanho. As características das imagens são analisadas pela diferença de contraste entre tecidos, geralmente as camadas de gordura apresentam tonalidades levemente mais claras, facilitando a leitura das mesmas. Os achados aparecem em tonalidades nitidamente claras, identificando uma possível anomalia (BOYD, 1995).

A mamografia é indicada para fins de diagnóstico, quando há alguma suspeita de anormalidade por alguma queixa de alterações na aparência da mama, e frequentemente recomendada para mulheres acima de 40 anos, por apresentarem altas taxas de incidências de câncer de mama (INCA, 2015).

2.1 A MORFOLOGIA DA MAMA

As mamas femininas são estruturas glandulares pares situadas na parede anterior e superior do tórax, que derivam de glândulas sudoríparas modificadas (ROUVIÈRE e DELMAS, 2005). A análise das glândulas mamárias é o passo inicial para entender como discorre as anormalidades da mama. Ao longo da vida a mulher obtém alterações na mama

devido a diversos fatores (modificações fisiológicas decorrente da idade, período de lactação, uso de medicamentos de reposição hormonal, alteração de peso, etc.) (BRITO, 2008).

A glândula mamária tem origem dérmica e localizam-se na pele por uma fásia, as glândulas são do tipo tubulo-alveolar constituídos por lobos, que podem ser canais de ductos, lóbulos e tecidos conectivos. Os lóbulos são formados por ácidos e são drenados pelo canal (ductos) lactífero. Os tecidos são divididos em adiposo (tecido entre glândula e pele, serve para facilitar o deslizamento da mama sobre os músculos), fibroso (localizado entre os lobos e os canais lactíferos, ajudando a sustentar os lobos) e glandular (prolongamento mamários que envolvem o súpero-externo até o axilar de maior e menor extensão) (GRAY, 2005). A Figura 1 mostra como está constituída a anatomia mamária.

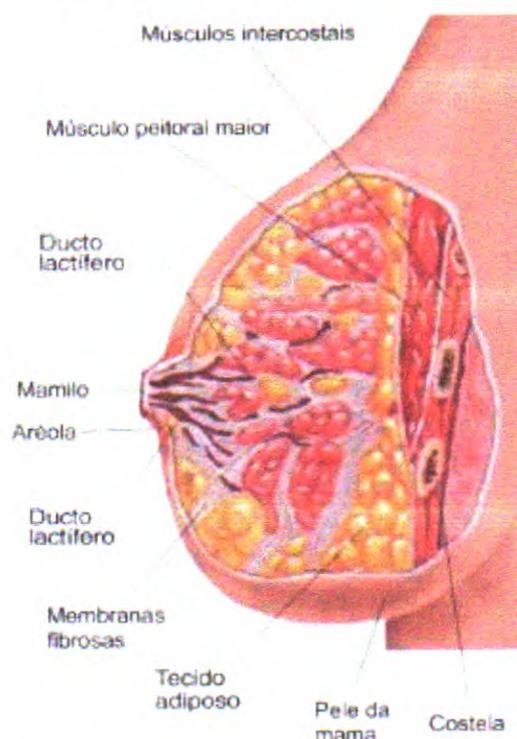


Figura 1 – Anatomia da mama Fonte: CLINIMATER, 2009

Os parênquimas mamários apresentam variações naturais de densidade. A tabela 1 apresenta um padrão que foi estabelecido pelo BI-RADS (*Breast Imaging Reporting and Data System*) (ACR, 2005).

Tabela 2- Padrões de densidade da mama estabelecida por BI-RADS (ACR, 2005)

DENSIDADE MAMÁRIA	DESCRIÇÃO
Mama liposubstituída	Tecido glandular < 25%
Mama fibroglandular	Tecido glandular entre 25% e 50%
Mama heterogeneamente densa	Tecido glandular entre 51% e 71%
Mama acentuadamente densa	Tecido glandular > 75%

Estudos comprovam que a densidade das mamas dificultam a leitura de imagens em exames clínicos, um fator que pode dificultar a primeira instância a detecção de tumores na mama. Normalmente as mulheres jovens tem um índice de densidade maior que as mulheres senis. As mulheres com densidades mais altas é um fator que implica em risco de câncer de mama em comparação com as de baixa densidade (ACR, 2005; MACHETTI e MARANA, 2007).

2.2 CALCIFICAÇÕES E NÓDULOS

Estudos de calcificações que descrevem a origem e a morfologia de calcificações, tendo destaque para Raul Leborg, que identificou que o câncer de mama originara de calcificação, entretanto a nomenclatura mais utilizada é a BI-RADS, que se destaca pela forma mais concisa de discriminação dos nódulos, sendo assinalada quanto a distribuição, a localização, a estabilidade e os achados associados (ACR, 2005).

As microcalcificações são calcificações que podem apresentar-se em agrupamentos, e ficam situadas no tecido adiposo ou fibroglandular. As microcalcificações ocupam áreas pequenas e normalmente não apresentam sombra, facilitando a identificação, podendo ser disseminadas ou agrupadas (ACR, 2005).

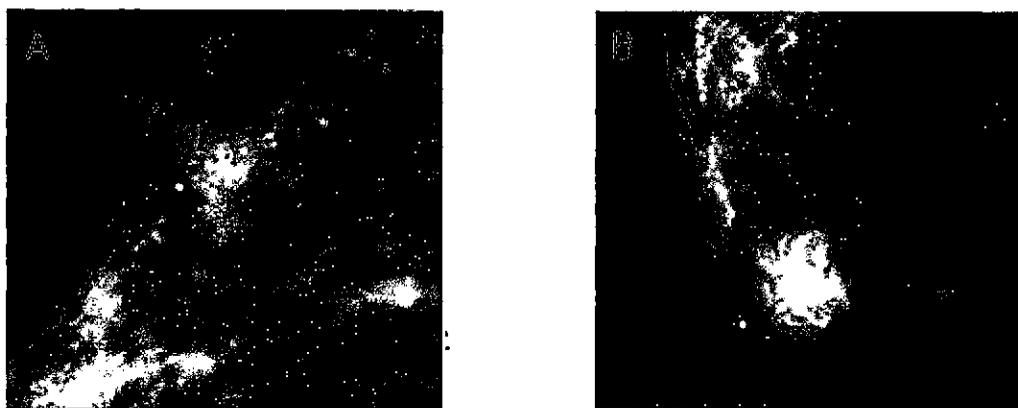


Figura 2 – Aglomerações de calcificações (A) e Nódulos de forma irregular (B) FONTE: DDSM – Programa de pesquisa da Universidade do Sul da Flórida

As suspeitas de malignidade das calcificações estão diretamente ligada ao nível de regularidade ou deformações dos achados (Figura 2). As investigações dos achados em imagens de mamografia foram padronizadas pela organização BI-RADS, e tem o objetivo de

uniformizar os resultados dos exames, para estabelecer uma conduta para cada tipo específico de caso. A descrição dos laudos médicos são baseados de acordo com o tamanho, forma, margem e tipo (aglomerado, disseminado ou nódulo único) de achado suspeito, podendo se estender a uma investigação mais detalhada.

2.2.1 Descrições de Microcalcificações

2.2.1.1 Forma

A forma é uma característica relevante a ser investigado, em caso de um achado suspeito. Para KOPANS (2000), a malignidade de um nódulo é definida pela sua regularidade, e quanto mais irregular se apresentar, mais suspeito de malignidade torna-se o nódulo.



Figura 3 – Formas de um carcinoma Fonte: Descrição de Formas em imagens mamográficas (KOPANS, 2000).

2.2.1.2 Margem

A margem é a borda dos achados, representa a configuração dos limites dos achados, portanto, limites mal definidos caracterizam achados muito suspeitos a malignidade que limites bem definidos (KOPANS, 2000).

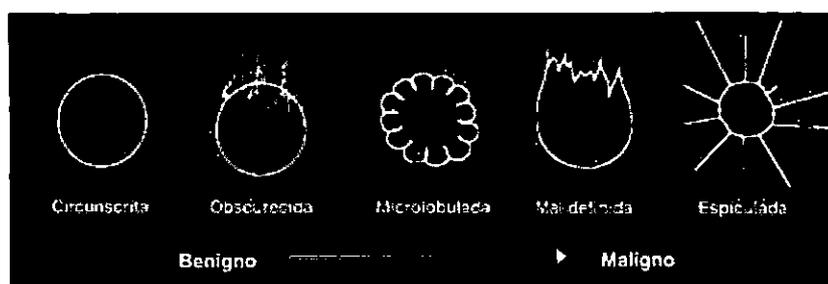


Figura 4 – Margens de nódulos Fonte: Caracterização de margens (KOPANS, 2000).

2.3 CATEGORIAS E AVALIAÇÕES EM EXAMES CLÍNICOS MAMOGRÁFICOS

Os laudos de exames clínicos concedidos de acordo com a padronização BI-RADS, classificados em uma escala que utiliza valores que variam de 1 a 5 (ACR, 2005).

- Categoria 0: Necessita de avaliação adicional
 - Para esta categoria, significa que foi encontrado algo suspeito e precisa fazer um outro exame (ultrassonografia) para investigar mais detalhadamente o caso;
- Categoria 1: Negativa
 - Exame não encontrou nenhuma anomalia, tais como nódulos, distorções de arquitetura, espessamento da pele ou microcalcificações.
- Categoria 2: Achados Benignos
 - Laudo que aponta negativo para malignidade, e apresenta prováveis alterações advindas de alterações pós-cirúrgicas, ou prováveis fibroadenomas.
- Categoria 3: Achado Provavelmente Benigno – Sugerido Seguimento em Curto Intervalo
 - Podem apresentar nódulos sólidos com margens circunscrita, forma ovóide, com a probabilidade de ser achado maligno de 2%. Cistos não palpáveis ou agrupados entram na categoria, sendo sugerido um acompanhamento a curto prazo.
- Categoria 4: Anormalidade Suspeita – Biópsia deve ser Considerada
 - Lesão dessa categoria tem uma probabilidade de câncer de 3 a 94%. Para esta categoria, o recomendado é uma amostra física para maiores investigações.
- Categoria 5: Altamente Sugestivo de Malignidade- Quase certamente maligna
 - Anomalia compreendida desta magnitude tem a probabilidade de risco malignas maiores que 95%.

3. REDES NEURAIAS ARTIFICIAIS

Redes neurais artificiais são sistemas de processamento de informações distribuídas, compostas por vários elementos computacionais simples que interagem por meio de conexões com pesos distintos. Inspiradas na arquitetura do cérebro humano, as redes neurais exibem características como a habilidade de aprender padrões complexos de dados e generalizar a informação aprendida (SAMPAIO, 2009).

Modelo computacional inspirado no neurônio biológico de seres vivos parte da premissa em que o neurônio biológico é o principal elemento, onde ocorrem as reações químicas. Os axônios são os meios de propagação das informações, que se dão a partir dos dendritos (soma). A ligação com outros neurônios é feita por meio de sinapses que estão conectados ao neurônio receptor. De acordo com as excitações, ou inibições que os neurônios receptores adquirem, os elementos são novamente transmitidos pelos axônios para outras células (HAYKIN, 1999).

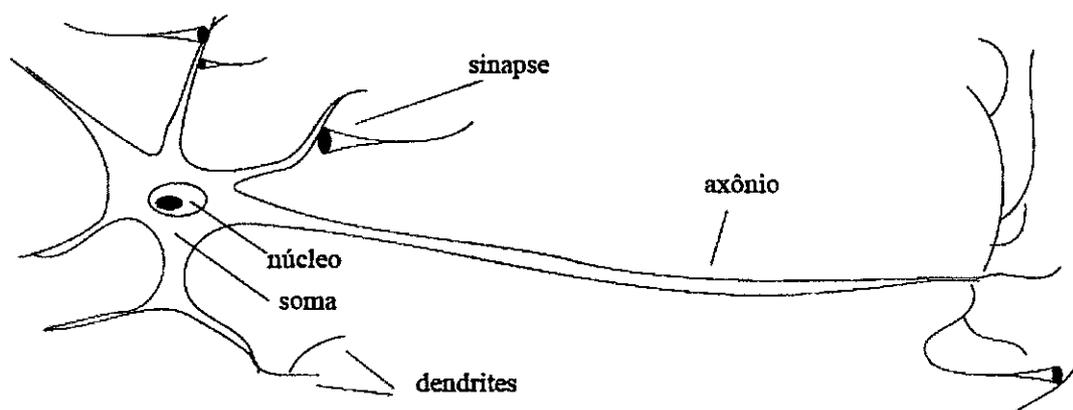


Figura 5 – Neurônio Biológico Fonte: RAUBER, 2015.

A representação do modelo artificial foi proposto pelos pesquisadores McCulloch e Pitts (1943), que engloba as principais características de uma rede de neurônios biológicos, sendo o modelo mais utilizado nas diferentes arquiteturas de redes neurais artificiais.

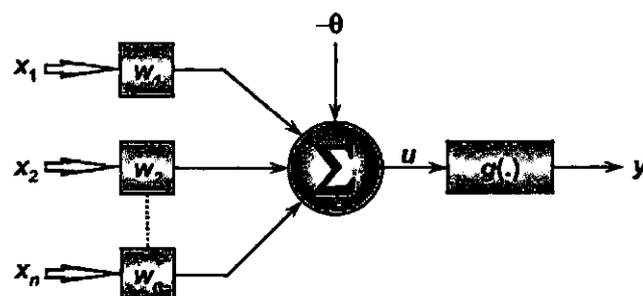


Figura 6 – Neurônio Artificial Fonte: SILVA, 2010.

Na figura 6, observa-se a arquitetura de um neurônio artificial, como ele é construído e seus elementos básicos.

- Sinais de Entrada $\{x_1, x_2, \dots, x_3\}$
 - São sinais de entrada, que se resumem as entradas de parâmetros dos objetos que se pretende observar o comportamento computacional.
- Pesos Sinápticos $\{w_1, w_2, \dots, w_3\}$

- São valores que permite ponderar cada uma das entradas da rede, classificando a relevância de cada valor em relação às funcionalidades dos neurônios.
- Combinador Linear $\{\Sigma\}$
 - O objetivo é fazer a soma das respectivas entradas ponderadas com os pesos sinápticos para produzir um valor com potencial de ativação.
- Limiar de ativação $\{\theta\}$
 - Uma variável que especifica o estado adequado para que a partir do resultado produzido pelo combinador linear possa gerar uma saída por intermédio do neurônio.
- Potencial de ativação $\{u\}$
 - É o resultado da diferença do valor produzido entre o combinador linear e o limiar de ativação, quando o resultado é $u \geq \theta$, ou seja, positivo significa que o neurônio produziu um potencial excitatório, caso contrário, inibitório;
- Função de ativação $\{g\}$
 - A função é limitar a saída do neurônio dentro do intervalo de valores assumidos pela combinação de pesos e entradas atribuídas a uma imagem funcional.
- Sinal de saída $\{y\}$
 - O valor final produzido pelo neurônio, resultado das relações de entradas e pesos, podendo ser aproveitado por outros neurônios.

As expressões seguintes sintetizam os resultados dos neurônios, e os estudos feitos por McCulloch e Pitts (SILVA, 2010):

$$u = \sum_{i=1}^n w_i \cdot x_i - \theta \quad (1)$$

$$y = g(u) \quad (2)$$

As Redes Neurais Artificiais são divididas em três partes, chamadas de camadas, as quais são nomeadas da seguinte forma (SILVA, 2010):

- Camada de Entrada
 - É a camada responsável pelo recebimento de informações, são características predefinidas do meio externo, também denominadas amostras. São normalmente normalizadas em relação as faixas de variações dinâmicas.
- Camadas escondidas, intermediárias, ocultas ou invisíveis

- Composta de neurônios que tem a responsabilidade de extrair as características associadas ao processo ou sistema a ser inferido. Na maioria das vezes os processos são feitos nessa camada.
- Camada de Saída
 - Constituídas de neurônios, é responsável pelos resultados processados pela rede neural, apresentam resultados finais de acordo com os processos advindos das camadas anteriores.

As principais arquiteturas de redes neurais artificiais, consideram-se a maneira como os neurônios estão distribuídos, assim como suas formas de ligações, e a constituição das camadas definem uma topologia a ser empregada para cada tipo de problema que se deseja solucionar.

Na figura 7, tem-se uma rede neural artificial de característica *feedforward*, em que o fluxo de informações segue sempre em uma direção. Composta por uma camada de entradas de n sinais, uma camada de neurônios intermediária de n_1 neurônios, outra camada intermediária com n_2 neurônios e uma camada de saída composta de m neurônios resultando os valores de saída da aplicação.

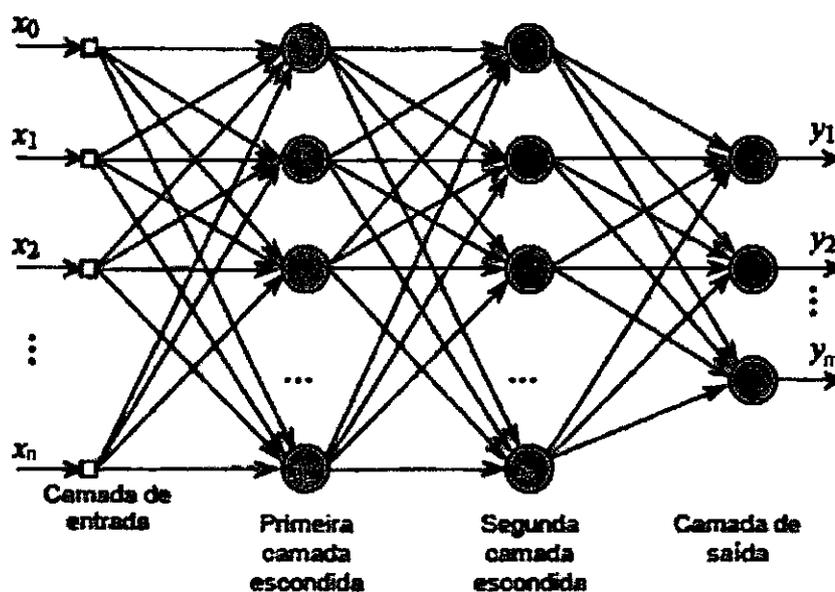


Figura 7 – Rede neural artificial de múltiplas camadas Fonte: Adaptada (OLIVEIRA, 2009).

3.1 REDES PERCEPTRON MULTICAMADAS

As redes perceptron multicamadas, ou MLP (Multi-Layer-Perceptron), são caracterizadas por possuírem camadas escondidas de neurônios, localizadas entre a camada de entrada e camada de saída (SILVA, 2010). A MLP tem uma vasta gama de utilização, sendo

uma arquitetura capaz de solucionar problemas em diversas áreas de conhecimento, tais como os seguintes destaques:

- Aproximação de Funções;
- Reconhecimento de padrões;
- Controle de processos;
- Previsões temporais;
- Otimização de sistemas.

A rede MLP baseia-se no método de camadas múltiplas, e sua forma de treinamento é supervisionado. A figura 8 mostra a estrutura da rede, onde as informações iniciam-se na camada de entrada passando pela camada intermediária chegando a camada de saída. A arquitetura foi proporcionada pelo algoritmo *backpropagation*, que permitiram a implementação nos processos de treinamento da rede e baseia-se na correção de erros, estudo publicado no livro *Parallel Distributed Processing* (RUMELHART, 1986).

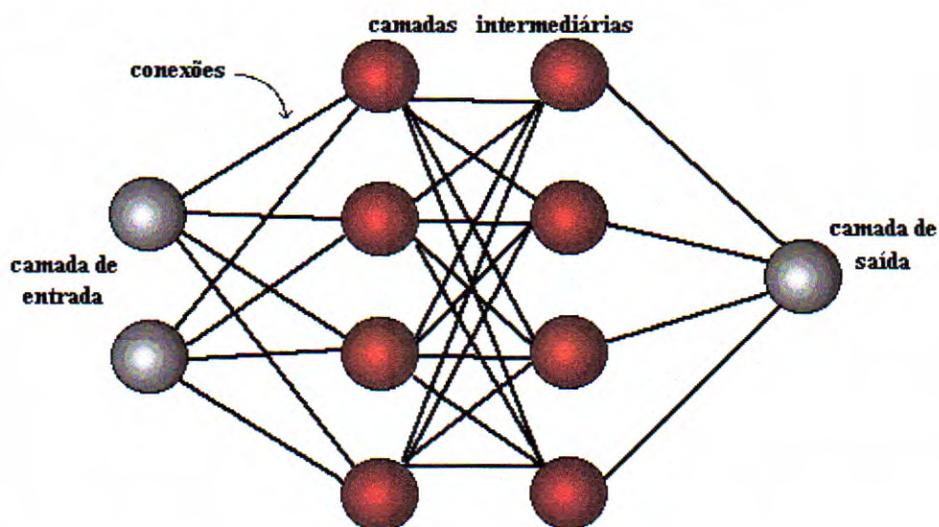


Figura 8 – Rede Perceptron Multicamadas Fonte: ICM – Imagem criada por Prico do repositório da USP

O conjunto de informações é distribuído por todos os neurônios da rede, de entrada a saída. Os estímulos são injetados pela camada inicial, propagando para a camada intermediária, que codificam as informações por meio dos pesos sinápticos e limiares dos neurônios, os neurônios da camada de saída recebem os estímulos da camada intermediária, definindo uma generalização, padrão que será exibido pela rede (BRAGA *et al.*, 2000).

A topologia da rede para cada problema que se deseja aplicar depende de diversos fatores, pois cada classe de aplicação tem características particulares, e volumes de informações diversos, no qual é representado por uma matriz de valores, e também matrizes

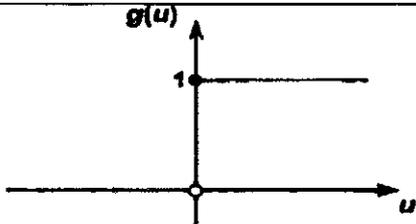
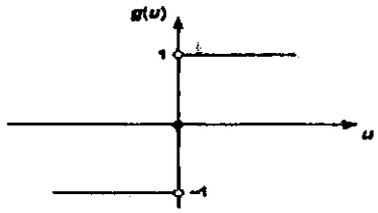
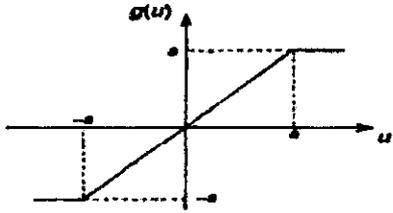
de pesos, estes fatores influenciam na quantidade de neurônios necessários para se obter um bom resultado na resolução de problemas (BRAGA *et al.*, 2000).

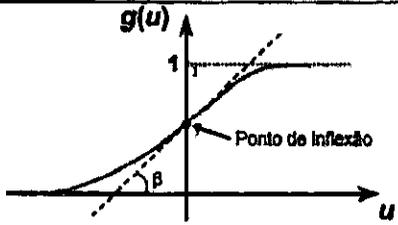
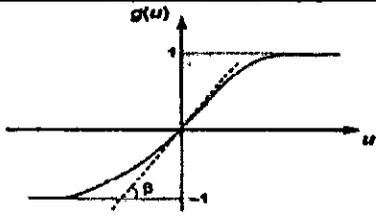
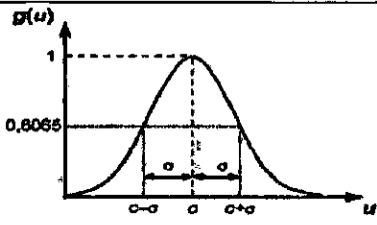
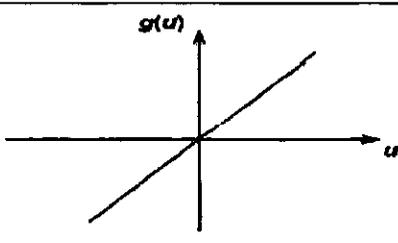
3.2 FUNÇÕES DE ATIVAÇÃO

O modelo de McCulloch e Pitts, os neurônios artificiais possuem uma função que calcula a saída do mesmo, o seu objetivo é limitar a saída do neurônio dentro de um intervalo de valores razoáveis. A principal característica é a diferenciabilidade de padrões (BRAGA, 2000).

As funções de ativação podem ser divididas em dois grupos principais, isto é, funções parcialmente diferenciáveis e totalmente diferenciáveis, considerando o domínio das mesmas. As funções parcialmente diferenciáveis são aquelas em que as derivadas de primeira ordem não existe, como as funções degrau, degrau bipolar e rampa simétrica. Para as funções totalmente diferenciáveis, as derivadas são conhecidas em todos os pontos de seu domínio de definição, são empregadas quatro principais: função logística, tangente hiperbólica, gaussiana e linear. A Tabela 2 mostra as principais funções de ativação utilizadas nas redes neurais (SILVA, 2010).

Tabela 2 - Funções de Ativação

Nome	Representação	Função Matemática
Função Degrau		$g(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases}$
Função degrau bipolar ou sinal		$g(u) = \begin{cases} 1, & \text{se } u > 0 \\ 0, & \text{se } u = 0 \\ -1, & \text{se } u < 0 \end{cases}$
Função rampa simétrica		$g(u) = \begin{cases} a, & \text{se } u > a \\ u, & \text{se } -a \leq u \leq a \\ -a, & \text{se } u < -a \end{cases}$

Função logística		$g(u) = \frac{1}{1 + e^{-\beta u}}$
Função tangente hiperbólica		$g(u) = \frac{1 - e^{-\beta u}}{1 + e^{-\beta u}}$
Função Gaussiana		$g(u) = e^{-\frac{(u-c)^2}{2\sigma^2}}$
Função linear		$g(u) = u$

Fonte: Adaptada (SILVA, 2010).

3.3 PROCESSOS DE TREINAMENTO

A grande relevância das redes neurais artificiais é a capacidade de aprendizagem a partir de um conjunto de características em forma de amostras (padrões), que são inseridas no sistema, sendo observado o comportamento em relação às entradas e saídas, tendo a capacidade de generalizar as configurações de um problema proposto, dando soluções padronizadas (SILVA, 2010).

Os conjuntos de parâmetros que devem ser inseridos na rede são divididos em dois subconjuntos; uma quantidade destinada a treinamento e outra para testes. Os conjuntos separados para testes servem para verificar se a generalização da rede neural está no intervalo aceitável, validando a topologia da rede aplicada (BRAGA *et al.*, 2000).

O processo de treinamento das redes neurais artificiais se dá até o momento em que se encontram as taxas de aprendizagem aceitáveis, acompanhando os ajustes de pesos e

limiares de ativação, que são organizados de acordo com as épocas de treinamento (SILVA, 2010; BRAGA *et al*, 2000).

3.3.1 Treinamento supervisionado

Os pesos e limiares são ajustados por meio das ações comparativas, executadas pelos algoritmos de aprendizagem, que supervisionam os valores de saída em relação aos valores desejados. A rede será finalizada quando os valores de saída da rede estiverem dentro dos parâmetros aceitáveis da generalização (SILVA, 2010).

3.3.2 Treinamento não-supervisionado

Para o treinamento não-supervisionado, há a aplicação dos algoritmos de aprendizagem, no entanto as saídas desejadas são inexistentes. Uma peculiaridade é a reorganização da rede em relação aos elementos componentes do conjunto total de amostras. Os pesos sinápticos e limiares dos neurônios da rede devem ser ajustados de acordo com o algoritmo de aprendizagem (SILVA, 2010).

3.3.3 Treinamento com reforço

O processo de treinamento com reforço é realizado por tentativa de erro, tendo a única resposta se esta for satisfatória ou não. Os pesos sinápticos e limiares são alterados de acordo com a resposta condicional do sistema, visando compensar os estados comportamentais dos processos (SILVA, 2010).

4 MATERIAIS E MÉTODOS

A pesquisa consiste em aplicar a técnica de inteligência computacional, utilizando redes neurais artificiais MLP para predição de diagnósticos de mamografias baseados em laudos médicos definidos pela literatura BI-RADS para avaliar exames clínicos. É uma 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 constituída por treinamento, testes e acompanhamento dos resultados.

A pesquisa, ao longo do desenvolvimento seguiu etapas concomitantes e sequenciais descritas visualmente na figura 9.

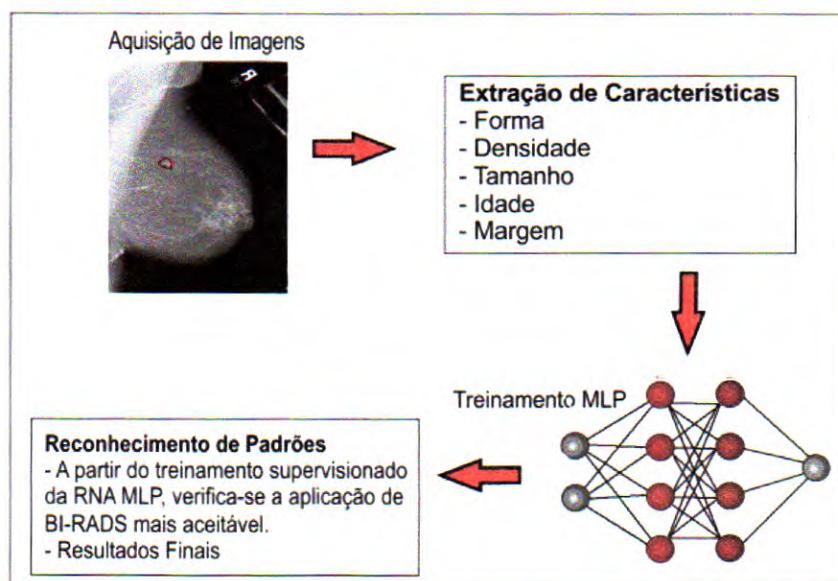


Figura 9 - Etapas aplicadas a metodologia. Fonte: Autor

4.1 AQUISIÇÃO DE IMAGENS

As imagens foram adquiridas de um repositório digital de Imagens desenvolvido pela Universidade do Sul da Florida, que tem em seu acervo uma vasta gama de imagens digitalizadas e devidamente caracterizadas chamada *Digital Database for Screening Mammography* (DDSM). *University of South Florida Digital Mammography* fornece uma pesquisa disponibilizada em uma estrutura de banco de imagens, padronizadas no tamanho de 230 x 350, discriminando as características de segundo as margens, contornos, e classificação BI-RADS de cada uma, identificando nas imagens os pontos de interesses, também conhecido como ROIs (*Region of Interest*).

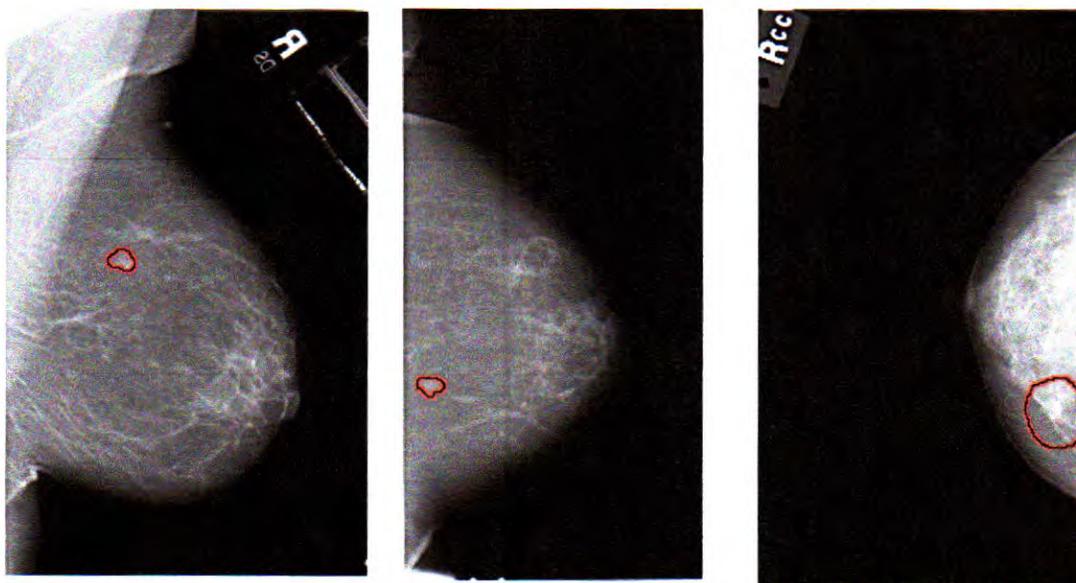


Figura 10 - Imagens de mamografias com seus respectivos ROIs Fonte: USF - Universidade da Florida, banco de imagens (DDSM).

As imagens da figura 10 mostram três casos de calcificações encontradas nos exames de mamografia de diferentes pacientes, o qual foram pré-estabelecidos pela fonte de pesquisa uma identificação dos focos de áreas suspeitas. As áreas consideradas suspeitas circulas em vermelho foram o ponto de estudo e classificação da pesquisa.

A figura 11 mostra as principais características usadas na descrição dos exames mamográficos, que influenciam diretamente no laudo dos exames, dado em BI-RADS. Os parâmetros dos laudos utilizados foram as margens (*margins*), forma (*shape*), classificação (*assessment*), seguido da idade e densidade que se encontram em um arquivo do tipo .txt separado.

```

FILE: A_1111_1.RIGHT_MLO.OVERLAY
TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE MASS SHAPE ROUND MARGINS ILL_DEFINED
ASSESSMENT 4
SUBTLETY 3
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 1
BOUNDARY
  
```

Figura 11 - Descrição do exame de mamografia Fonte: USF - Universidade da Florida, banco de imagens (DDSM).

4.2 EXTRAÇÃO DE CARACTERÍSTICAS

A definição das principais características foram tomadas mediante a um estudo de literatura baseado em KOPANS (2000), que descreve quanto a forma, e margem os pontos importantes para definição dos achados na mama. As classificações dos achados nos exames definiu-se uma escala de 1 a 5 para tipo de forma e margem.

Tabela 3 – Descrição da forma dos achados na mama

Forma	Numeração
Redonda	1
Ovulada	2
Lobulada	3
Irregular	4
Distorção de arquitetura	5

Tabela 4 – Descrição das Margens dos achados na mama

Margem	Numeração
Circunscrita	1
Obscurecida	2
Microlobulada	3
Mal definida	4
Espiculada	5

A margem e a forma são dois parâmetros tirados das descrições de cada exame, descrita pelo estudo de cada caso individual disponibilizado pelo banco de imagens digital da Universidade da Florida (DDSM) (Tabelas 3 e 4).

Além da margem e forma, também foram alvos de estudo a densidade, tamanho e idade da paciente (ACR, 2005; MARTINS, BARRA e LUCENA, 2010). A densidade é explicitada em cada descrição individual de cada imagem, assim como a idade da paciente.

A caracterização do ROI que se pretendia extrair, foi tirado manualmente pela ferramenta CorelDraw®, o qual serviu para identificar as bordas dos achados. Em seguida, utilizou a ferramenta Matlab®, inicialmente para extrair o tamanho (cm²) dos ROI's em cada imagem. As imagens foram padronizadas nas dimensões 230 x 350, estabelecidas pelo DDSM. A figura 12 ilustra os processos adquiridos na etapa de limiarização.

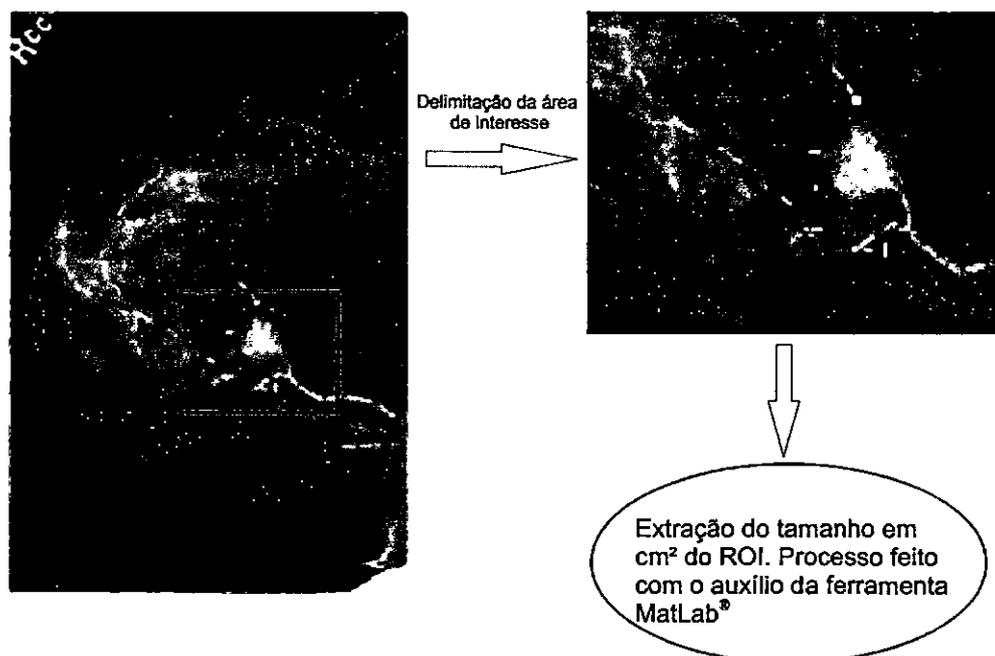


Figura 12 - Limiarização dos ROIs. Fonte: AUTOR

Ao término do procedimento de identificações de ROIs referente as calcificações individuais delimitando a área de interesse, foram reservadas 150 amostras as quais foram distribuídas dentre as 5 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 150 ROIs obteve-se a seguinte amostra por tipo morfológico: 30 BI-RADS 1, 30 BI-RADS 2, 30 BI-RADS 3, 30 BI-RADS 4, 30 BI-RADS 5. As amostras BI-RADS foram retiradas previamente dos laudos dos arquivos disponibilizados pelo DDSM, e configura-se na classificação para cada imagem mamográfica nos parâmetros identificados de acordo com o *assessment* (classificação em BI-RADS), como mostra a figura 11.

4.3 ORGANIZAÇÃO DOS DADOS

Após a definição da quantidade de amostras necessárias para o estudo, estabeleceu-se uma tabela contendo todas as amostras em que foram extraídas as características das imagens (Anexo A). Os parâmetros forma, margem, densidade, idade e tamanho são as entradas da MLP, e o parâmetro BI-RADS é a classificação alvo aplicada a MLP.

4.4 TREINAMENTO DA REDE NEURAL MLP

Para criar, treinar, validar e simular as RNAs, foi utilizado o *software MATLAB*[®] (Figura13), na versão R2013a (8.1.0.604). A RNA adotada para indicar a classificação morfológica (ACR, 2005) das ROIs consiste na função do modelo MLP. Inicialmente foi empregada uma RNA MLP com a seguinte característica: algoritmo de aprendizado *Backpropagation*, 5 (cinco) parâmetros de entrada que correspondem a quantidade de informações morfológicas (previamente descritas na seção 2) extraídas das ROIs, com exceção do parâmetro de BI-RADS, o qual é aplicado apenas na análise de resultados das calcificações.

Para o se obter o processo de treinamento da MLP foi utilizado uma máquina de configuração: processador da Intel® Pentium de 2.0 GHz, memória de 2GB com um sistema de 32 Bits. O custo de tempo necessário para se treinar a rede foi aproximadamente 136 horas, com acompanhamento supervisionado da rede neural.

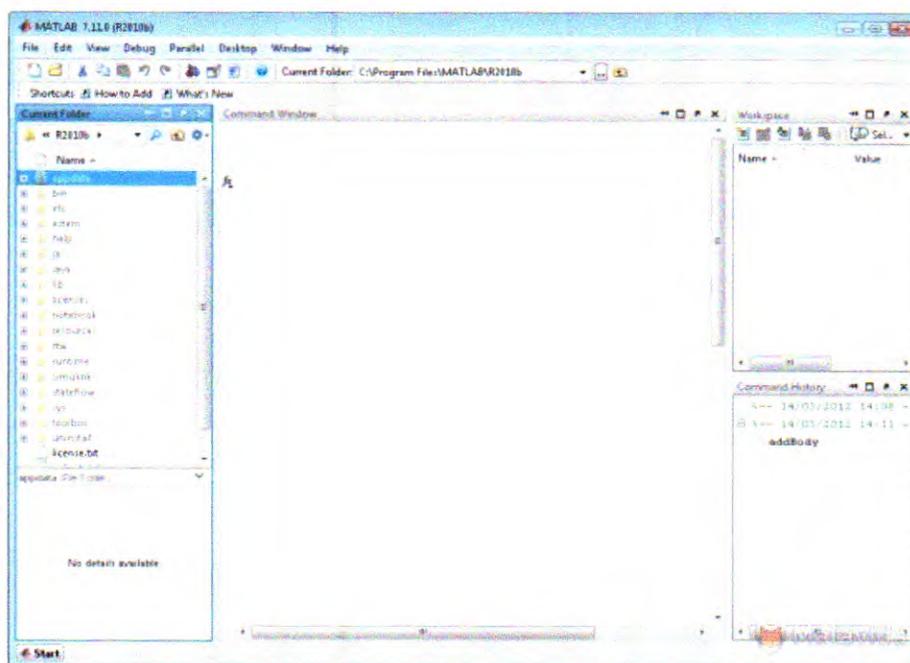


Figura 13 – Interface da ferramenta Matlab[®] para treinamento da rede neural MLP. Fonte: AUTOR

Os elementos foram separados em três grupos no momento em que se inicia o algoritmo de aprendizagem *backpropagation* que tem a finalidade de adquirir resultados para cada tipo de classificação, a fim de verificar a eficiência da distribuição dos elementos na RNA.

Tabela 3 - Distribuição de Dados

Tipos de Classificação	Quantidade de Dados (%)
Treinamento	60
Validação	20
Teste	20

Algumas configurações de treinamento foram adotadas, para o bom funcionamento da rede neural. Foram adotadas no máximo 200 épocas de treinamento, 15 validações cruzadas para cada topologia da rede, sendo que para cada validação foram feitos 10 testes de treinamento de no máximo 200 épocas. Essa configuração permitiu que quando encontrado um resultado bom, fosse armazenado em um conjunto de variáveis, onde o objetivo foi encontrar os valores dos pesos sinápticos e o índices de erro e acertos, caso fosse encontrado outro resultado melhor, o resultado anterior era substituído pelo resultado atual. O treinamento da rede foi supervisionado.

5 ANÁLISE E DISCUSSÃO DOS RESULTADOS

5.1 TREINAMENTO DA REDE NEURAL

O treinamento da rede neural seguiu supervisionada em um tempo aproximado de 136 horas. A quantidade de amostras destinada a treinamento foram de 60% de todas as amostras, correspondendo a 90 amostras. Obteve-se um índice de acertos de 91%. A figura 14 mostra uma linha linear tracejada, o que corresponde a condição ideal para ser alcançada. A linha vermelha paralela a linha tracejada mostra o quão próximo o resultado se aproxima da condição ideal. A fig.14 mostra o gráfico das saídas em relação aos resultados desejados.

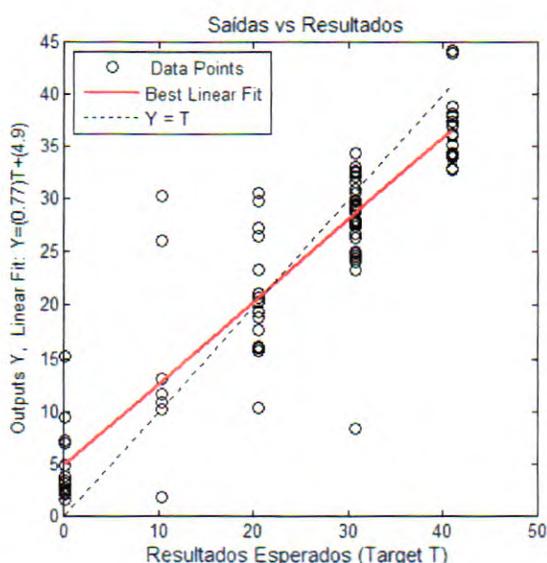


Figura 14 – Resultados do treinamento da rede neural, contando com 90 amostras de treinamento, apresentou um índice de acerto de 91%.

Por meio da análise de parâmetros identificados nesta fase de treinamento, observou-se que a influência de taxas de acertos esteve diretamente ligada a quantidade de amostras, pois quanto mais amostras fossem destinadas a treinamento, mais chances teriam a rede neural para um bom resultado.

5.2 RESULTADOS

O treinamento da rede neural, cujo objetivo era a convergência da função linear capaz de generalizar as classificações de entrada (tamanho, densidade, forma, idade e margem) da rede, foi bem sucedida para a classificação BI-RADS.

A validação cruzada possibilitou a determinação da melhor RNA pelo cálculo do menor erro quadrático médio comparando os resultados obtidos na rede com os resultados esperados e modificando os pesos sinápticos dos neurônios artificiais da RNA conforme as épocas (variável indicadora de etapa de treinamento concluída). Os pesos sinápticos são modificados conforme a necessidade de cada conjunto de entradas da rede, formando uma equação de ajuste linear.

A fim de avaliar os resultados, foram utilizadas duas métricas diferentes: a correlação coeficiente (valor r) e o erro médio absoluto percentual (MAPE). O primeiro indica a intensidade e direção de uma relação linear entre duas variáveis aleatórias, em outras palavras, como as variações dos dados de saída são explicadas pelo modelo neural. A segunda medida indica o quão perto os resultados preditos são dos dados reais.

Os valores referentes ao treinamento obtiveram o melhor resultado de correlação em relação aos valores de validação e testes, sendo de 0,9116, 0,8191, 0,5407 (Figura 15). Os valores acima de 0,9 são considerados excelentes para um treinamento de várias possibilidades de saídas. O conjunto que teve o maior índice de aproveitamento foi o de treinamento, pois para se obter uma boa generalização de função, necessitava de uma boa quantidade amostras. Para os conjuntos de teste e validação, na fase de teste a função generalizadora mostrou-se não tão eficiente quanto a validação, embora tenha-se destinada a mesma quantidade de amostras para ambas.

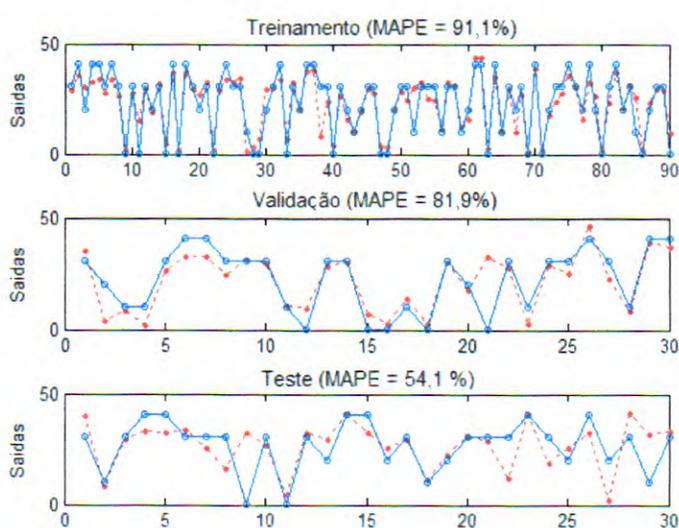


Figura 15 - Resultado de treinamento, validação e testes da Rede Neural.

5.3 TOPOLOGIA DA REDE NEURAL

Com o desenvolvimento do algoritmo da RNA obteve-se um sistema de inteligência computacional capaz de realizar interações entre todas as variáveis de entrada e, calcular a saída referente comparando ao resultado desejado, caracterizando assim, o aprendizado da rede. O sistema foi subdividido de modo a operar testes de validação cruzada a fim de garantir a generalização, minimizando a especialização da rede e conseguir a melhor RNA para a classificação de BI-RADS de imagens mamográficas.

As RNA geradas para este sistema foram criadas variando-se o número de neurônios das camadas escondidas (intermediárias), uma vez que a camada de entrada criada continha cinco elementos caracterizados pelo número de variáveis (tamanho (cm²), margem, forma, idade e densidade) e apenas um neurônio na camada de saída (classificação BI-RADS). A

primeira camada intermediária teve a quantidade de neurônios variada de dois a onze (2-11), já a segunda camada de zero (nenhum) a 5 neurônios (0-5).

Por meio dos testes realizados, a RNA detectada com o maior índice de acerto, ou seja, o menor erro quadrático médio possui topologia 8:3:1 ou seja, oito neurônios na primeira camada escondida, três na segunda camada escondida e um neurônio na camada de saída (Fig. 16).

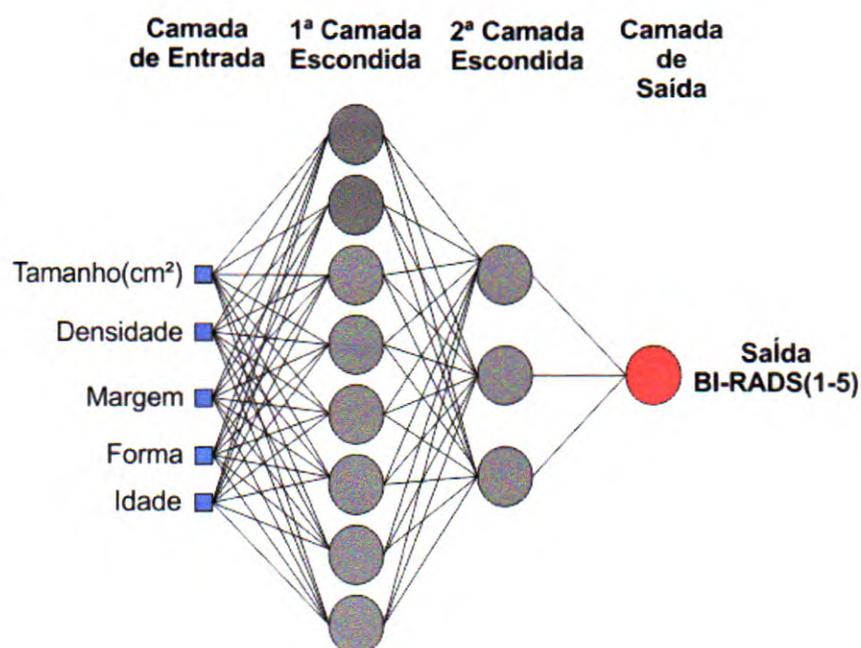


Figura 16 - Configuração da Rede Neural MLP treinada. Fonte: AUTOR.

6 CONSIDERAÇÕES FINAIS

As RNAs foram utilizadas para o treinamento de amostras extraídas de imagens mamográficas, que permitiu o estudo da capacidade de generalização da rede neural MLP para a definição de resultados de exames clínicos em BI-RADS. Os resultados obtidos comprovam que a eficiência de treinamento da rede neural mostrou-se satisfatória para a fase de treinamento tendo como índice de acerto de 91%. Em contrapartida os dados de testes não indicaram um grau de acerto considerável, questão em que o treinamento pode ser melhorado com o possível acréscimo de amostras, em um novo estudo relacionado ao efeito individual de cada parâmetro.

A continuidade deste trabalho pode ser aprofundada em pesquisas no âmbito de aprimorar o esquema de classificação desenvolvido. Para isto, apresentam-se algumas sugestões:

- *Inclusão de parâmetros de instabilidade das calcificações* – Foram estudadas cinco parâmetros relacionadas as calcificações mamárias (Forma, margem, densidade, tamanho e idade). O parâmetro tamanho pode ser estudado mais a fundo, utilizando técnicas automatizadas para extração desse parâmetro, assim como a inclusão de outras características relevantes.
- *Automatização na extração de ROIs* – Os parâmetros foram extraídos com o auxílio da ferramenta Matlab®, e classificados de acordo com os laudos de cada exame, quanto a forma (1-5) e margem (1-5). Tendo os resultados de treinamento da rede neural, pode ser aplicado a um sistema de automatização para laudos no padrão BI-RADS.
- *Melhorar o treinamento da Rede Neural MLP* – Aumentar o conjunto de imagens, principalmente envolvendo exames clínicos de mamografia. A partir dos novos dados, realizar treinamento, garantindo a confiabilidade na caracterização de imagens reais.
- *Tratamento de textura* – Investigar técnicas que eliminem e aplique o tratamento de sombras e reforços acústicos, esses problemas acabam que inibindo as imagens tornando difícil a leitura dos exames.

REFERÊNCIAS

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

BAUER, W.; IGOT, J.P.; LE, G.Y.: Chronologic Du câncer mammaire Utilisantun Modele de Croissance de Gompertz. *Ann AnatPathol* 25:39-56, 1980.

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.

BOYD, N.F.; BYNG, J.W.; JONG, R.A.; FISHELL, E.K.; LITTLE, L.E.; MILLER, A.B.; LOCKWOOD, G.A.; TRITCHLER, D.L.; YAFFE, M.J. Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study. *Journal of the National Cancer Institute*, v. 87, p. 670-675, 1995.

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.

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.

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.

CASTIGLIONE M., AEBI S. The enigma of young age. *Ann Oncol*; 2006; 17: 1475-7.

CLINIMATER, 2009. Disponível em: <<http://www.clinimater.com.br/anatomia.htm>> Acesso em: 10 fev. 2015.

CURI H. A. L., GOTLIEB S.L.D., LATORRE M.R.D.O. Evolução da mortalidade por neoplasias malignas no Rio Grande do Sul, 1979-1995. *Rev Bras Epidemiol* 2001; 4:169-8.

DRUKKER, K; PESCE, L. L; GIGER, M. L. Repeatability and classifier and classifier in computer-aided diagnosis for breast ultrasound. *Medical Physics*, v 37, n. 6, p. 2659-2669, 2010.

ELPÍDIO, F. G. I. Sistema de auxílio na avaliação de calcificações mamárias por processamento digital de imagens e inteligência artificial. 2012. 156 f. Dissertação de pós-graduação stricto sensu em engenharia biomédica, Faculdade Gama, Universidade de Brasília, DF. 2012.

FLETCHER S.W., ELMORE J.G. Clinical practice: mammographic screening for breast cancer. *N Engl J Med*. 2003; 348:1672-80. 21. Schwartzmann G. Breastcancer in South

FOGAÇA, E. I. C.; GARROTE, L. F.; Câncer de mama: atenção primária e detecção precoce. *Arquivos de Ciência da Saúde*, São José do Rio Preto, v. 11, n. 3, p. 179-181, jul-set. 2004.

FORSYTH, D. A.; PONCE, J. *Computer: A modern approach*. New Jersey: Prentice Hall, 2003.

GUERRA, MR.; et al. Sobrevida de cinco anos e fatores prognósticos em coorte de pacientes com câncer de mama assistidas em Juiz de Fora, Minas Gerais, Brasil. *Cad Saúde Pública* 2009; 25(11): 2455-66.

GRAY H. *Gray's Anatomy*. 39ª ed. Edinburgh: Elsevier; 2005.

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.

INCA, Instituto Nacional do Câncer. Endereço na internet: <<http://www.inca.gov.br/>> Acessado em 23 março 2015.

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.

KEMP, C.; ELIAS, S.; BORELLI, K.; NARVAIZA, D. G.; SCHOR, A. P. Punção aspirativa por agulha fina orientada por ultra-sonografia em lesões não-palpáveis. *Revista Brasileira de Ginecologista e Obstetrícia*: Rio de Janeiro, v. 23, n. 5, p. 321-327, jun. 2001.

KOPANS, D.B., *Imagem da Mama*: Rio de Janeiro, MEDSI, p.852, 2000.

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.

LOUVEIRA, M, H.; SOUZA, L. R. M. F.; CASTRO, I. M.; ELIAS, S.; SZEJNFELD, J.; KEMP, C. Avaliação e classificação ultra-sonografia dos nódulos mamários: revisão da literatura e aspectos atuais. *Revista Imagem*, v. 23, n. 2, p. 123-128, 2003.

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.

MARCONNI, K. D. Aplicação de modelos de redes neurais artificiais na segmentação e classificação de nódulos em imagens de ultrassonografia de mama. 2013. 131 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia Elétrica e Área de Concentração em Processamento de Sinais e Instrumentação, Escola de Engenharia de São Carlos, São Paulo. 2013.

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.

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

OLIVEIRA, A. C. S. Aplicação de redes neurais artificiais na previsão da produção de álcool. Universidade Federal de Lavras – UFLA. Minas Gerais, 15 jun. 2009. Disponível em: <<http://www.scielo.br/pdf/cagro/v34n2/02.pdf>>. Acesso em: 24 mai. 2015.

PARKIN, DM; BRAY, FI; DEVESSA, S.S. Cancer burden in the year 2000. The global picture. Eur J Cancer; 2001; 37 Suppl 8:S4-66.

RAUBER, T. Redes Neurais Artificiais. Disponível em: <<http://www.inf.ufes.br/~thomas>> acesso em: 22 mai. 2015.

Redes Neurais Artificiais. Disponível em: <<http://www.icmc.usp.br/~andre/body.htm>> acesso em: 21 mai. 2015

ROUVIÈRE H., DELMAS A. Anatomia Humana descritiva, topologia e funcional. 11ª ed. Barcelona: Masson; 2005.

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.

SILVA, PA.; RIUL, SS. Câncer de mama: fatores de risco e detecção precoce. Ver Bras Enferm, Brasília; 2011; 64(6): 1016-21.

SILVA, I.; SPATTI, D; FLAUZINO. R. Redes Neurais para engenharia e ciências aplicadas. São Paulo: Artliber, 2010.

STRAVOS, A. T.; New advances in breast ultrasound: computer-aided detection. Ultrasound Clinics, v. 4, n. 3, p. 285-290, jul. 2009.

YAP, M. M.; EDIRISINGHE, E. A.; BEZ, H. E. Object boundary detection in ultrasound images. IEEE 3^o Canadian Conference on Computer and Robot Vision 2006. Quebec, Canadá. Proceeding of CRV06, p. 53-58, jun. 2006.

ZHENG, D.; LU, G. Review of shape representation and description techniques. Pattern Recognition, v. 37, p. 1-19, 2004.

ANEXO A – Dados utilizados na extração de imagens mamográficas

Área (cm ²)	Margem	Forma	Densidade	Idade	BI-RADS
0.04926	5	4	2	72	4
0.13193	5	4	2	72	4
0.00588	5	4	2	46	5
0.08934	5	4	2	61	5
0.04001	5	4	2	59	5
0.09106	1	3	1	71	5
0.02147	3	4	1	83	4
0.01335	3	4	1	83	4
0.01631	5	4	3	52	5
0.03131	4	3	4	73	4
0.01292	2	2	1	72	1
0.05021	5	2	2	65	2
0.01795	3	3	3	35	4
0.06751	5	4	3	53	4
0.01975	4	3	2	48	4
0.01386	4	3	1	58	4
0.02743	2	3	2	60	4
0.01328	2	3	3	59	3
0.03612	5	5	1	80	5
0.03318	4	4	3	42	3
0.01781	5	5	3	75	4
0.02484	4	1	3	62	4
0.01795	2	3	3	60	4
0.03863	3	4	2	55	3
0.03835	4	4	2	55	4
0.00926	3	1	2	38	4
0.01321	3	2	1	47	3
0.01163	2	1	2	67	3
0.00576	4	4	3	48	4
0.03353	4	3	2	67	3
0.00919	1	2	2	85	4
0.00904	2	2	3	58	4
0.30939	4	4	2	61	4
0.00193	2	2	3	41	4
0.01637	2	1	3	42	4
0.01536	2	2	3	69	4
0.03203	1	2	3	72	2
0.03152	4	3	2	74	4
0.04007	2	1	4	68	4
0.06276	2	4	4	54	3
0.01185	2	2	2	49	3

0.03375	4	2	4	63	5
0.01034	4	3	3	59	4
0.09695	4	3	2	75	4
0.01364	5	4	2	73	5
0.05422	5	4	2	73	5
0.02305	4	4	2	83	5
0.05594	4	4	2	82	5
0.01723	3	3	1	63	4
0.01165	5	5	2	85	5
0.04682	5	5	2	80	5
0.03239	4	4	4	48	4
0.01659	4	4	2	63	4
0.08158	4	4	4	50	4
0.08998	4	4	4	56	4
0.01457	1	3	3	78	4
0.10501	5	5	3	71	5
0.04294	5	4	4	69	5
0.07167	1	3	3	37	1
0.04021	1	2	2	75	1
0.01831	1	2	1	53	1
0.03021	2	4	1	52	1
0.03125	2	2	2	49	1
0.02341	1	2	3	35	1
0.01035	1	3	1	54	1
0.02346	4	4	1	42	4
0.03024	4	2	2	37	4
0.04523	1	3	2	45	2
0.00423	4	4	2	54	4
0.02032	3	3	2	52	4
0.04026	5	4	3	48	4
0.05031	5	1	2	50	4
0.02681	1	2	1	35	3
0.00403	1	1	2	52	2
0.04065	2	2	3	51	2
0.00802	1	2	3	60	2
0.08203	5	4	3	65	1
0.04613	1	4	1	54	1
0.03915	3	3	3	44	3
0.02723	1	2	3	40	1
0.03512	4	4	2	65	2
0.03852	5	5	3	40	5
0.08246	5	4	4	58	5
0.04856	5	4	3	74	5
0.03241	4	2	2	64	4
0.04352	1	1	3	50	1

0.01701	5	5	2	59	4
0.02382	4	1	4	55	4
0.01895	2	3	2	75	4
0.04873	3	4	1	80	3
0.03605	4	4	3	37	4
0.01926	3	1	3	38	4
0.01851	3	2	2	58	3
0.04263	2	1	1	75	3
0.01576	4	4	4	39	4
0.04353	4	3	1	70	3
0.01919	1	2	2	85	4
0.01204	2	2	3	58	4
0.09939	4	4	2	61	4
0.02193	2	2	2	41	4
0.02437	2	1	4	42	4
0.03346	2	2	2	69	4
0.01831	1	2	2	53	1
0.03021	2	4	3	52	1
0.03105	2	2	2	49	1
0.03341	1	2	4	37	1
0.01535	1	3	2	54	1
0.04683	1	3	3	42	2
0.01013	1	1	2	52	2
0.03965	2	2	4	51	2
0.01202	1	2	1	65	2
0.04501	5	4	2	59	5
0.07996	1	3	1	71	5
0.01391	5	4	3	52	5
0.02438	2	3	3	59	3
0.04112	5	5	1	80	5
0.04118	4	4	3	42	3
0.02681	5	5	3	65	4
0.03584	4	1	3	52	4
0.03341	1	2	3	35	1
0.02035	1	3	4	54	1
0.04381	1	2	4	35	3
0.01043	1	1	3	52	2
0.04265	2	2	3	51	2
0.01802	1	2	3	60	2
0.08223	5	4	3	65	1
0.04003	1	4	2	54	1
0.03465	3	3	3	44	3
0.02403	1	2	3	40	1
0.02522	4	4	1	65	2
0.05332	1	1	3	50	1

0.02051	3	2	3	58	3
0.04343	2	1	1	75	3
0.02165	5	5	1	85	5
0.04582	5	5	1	80	5
0.03129	4	4	3	48	4
0.02159	4	4	1	70	4
0.07958	4	4	3	49	4
0.02601	5	5	2	51	4
0.03582	4	1	4	53	4
0.02495	2	3	1	74	4
0.05473	3	4	1	82	3
0.02307	4	4	3	40	4
0.09501	5	5	3	71	5
0.04374	5	4	3	69	5
0.06737	1	3	4	37	1
0.05223	1	3	4	45	2
0.07926	5	4	4	58	5
0.02898	5	4	3	74	5
0.03887	3	4	1	80	3