

**RENORBIO-REDE NORDESTE DE BIOTECNOLOGIA
UNIVERSIDADE FEDERAL DO PIAUÍ
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**

**MÉTODO DE REDES NEURAIIS PARA CLASSIFICAR NÓDULOS
MAMÁRIOS BASEADO NO PADRÃO BI-RADS® NAS CATEGORIAS 4 E 5**

ELMO DE JESUS NERY JÚNIOR

**TERESINA
2022**

ELMO DE JESUS NERY JÚNIOR

**MÉTODO DE REDES NEURAIIS PARA CLASSIFICAR NÓDULOS
MAMÁRIOS BASEADO NO PADRÃO BI-RADS® NAS CATEGORIAS 4 E 5**

Tese de Doutorado apresentada ao Programa de Pós-Graduação da Rede Nordeste de Biotecnologia (RENORBIO), Ponto Focal Piauí como requisito para obtenção do título de Doutor.

Área de Concentração: Biotecnologia em Saúde

Linha de Pesquisa: Desenvolvimento de agentes profiláticos, terapêuticos e testes diagnósticos.

Orientador: Prof. Dr. Francisco das Chagas Alves Lima

Coorientador: Prof. Dr. Otilio Paulo da Silva Neto

Coorientador: Prof. Dr. Francisco Adelson Alves Ribeiro

**TERESINA
2022**

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Serviço de Processos Técnicos

N455m Nery Júnior, Elmo de Jesus.
Método de redes neurais para classificar nódulos mamários baseado no padrão BI-RADS® nas categorias 4 e 5 / Elmo de Jesus Nery Júnior. -- 2022.
94 f.

Tese (Doutorado) – Universidade Federal do Piauí, Rede Nordeste de Biotecnologia - RENORBIO, Programa de Pós-Graduação em Biotecnologia, Teresina, 2022.
“Orientador: Prof. Dr. Francisco das Chagas Alves Lima”.
“Coorientador: Prof. Dr. Otilio Paulo da Silva Neto”.

1. Redes neurais (Computação). 2. Câncer de mama. 3. BI-RADS. I. Ferreira, Maria Rosa Gomes Rocha. II. Título.

CDD 006.3

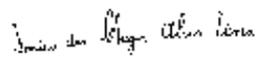
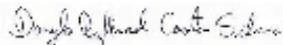
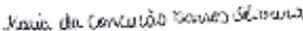
Bibliotecária: Thais Vieira de Sousa Trindade - CRB3/1282

FOLHA DE APROVAÇÃO – DEFESA DE TESE

ALUNO: ELMO DE JESUS NERY JUNIOR

TÍTULO DO PROJETO: “MÉTODO DE REDES NEURAIS PARA CLASSIFICAR NÓDULOS MAMÁRIOS BASEADO NO PADRÃO BI-RADS® NAS CATEGORIAS 4 E 5.”

PROFESSOR ORIENTADOR: Prof. Dr. Francisco das Chagas Alves Lima - UESPI

BANCA EXAMINADORA:	CONCEITO	ASSINATURA
Prof. Dr. Francisco das Chagas Alves Lima - UESPI (Presidente)	Satisfatório	
Prof. Dr. Carlos Giovanni Nunes de Carvalho - UESPI (Examinador)	Satisfatório	
Prof. Dr. Danylo Rafael Costa Silva - UESPI (Examinador)	Satisfatório	
Prof. Dr. Francisco Adelson Alves Ribeiro - UFPI (Examinador)	Satisfatório	
Profa. Dra. Maria da Conceição Barros Oliveira - UFPI (Examinadora)	Satisfatório	

DATA DA AVALIAÇÃO: 10 de Junho de 2022

HORÁRIO: 14:00h

LOCAL: Plataforma virtual

DEDICATÓRIA

Dedico esta Tese primeiramente a **Deus** por ser minha fonte de proteção e fortaleza diante dos desafios que surgiram ao longo desta caminhada.

Dedico à minha esposa e também companheira **Isabel Cristina de Almeida Brito Nery**, por sempre estar presente na minha vida, pelo amor, carinho, paciência, apoio, companheirismo e incentivo durante todas as etapas desta conquista. E também dedico aos meus filhos **Samuel de Jesus Almeida Nery**, **Davi de Jesus Almeida Nery** e **Enzo de Jesus Almeida Nery**.

Dedico aos meus pais, **Elmo de Jesus Nery (in memoriam)** e **Bibiana Maria Pereira Nery**, os principais incentivadores e guias da minha carreira acadêmica, os frutos de hoje são consequências da semente plantada por eles nos meus primeiros anos de escola, eles são fundamentais em todos os aspectos da minha vida, muito obrigado pelo carinho, amor e respeito que a mim dedicaram. Também dedico: ao minha irmã **Naiana Bianelly Nery** pela companhia e incentivo e aos meus sogros **Rosa Maria Nunes de Almeida** e **Luiz Nunes de Almeida (in memoriam)**.

DEDICATÓRIA ESPECIAL

Ao **Prof. Dr. Benedito Borges da Silva (in memoriam)**, orientador desta tese, mestre na verdadeira acepção da palavra, marco em minha formação como pesquisador, me ensinou a importância da pesquisa e seus benefícios para a sociedade, pela sua brilhante orientação deste trabalho, por ter me acolhido em seu grupo de pesquisa, pelo incentivo, amizade e confiança, a minha eterna gratidão (IN Memoria).

Ao **Prof. Dr. Francisco das Chagas Alves Lima**, meu novo orientador pelo acolhimento e incentivo, a sua ajuda foi primordial para a conclusão da tese. Sua escolha foi fundamental e o aprendizado adquirido com seu ensinamento foram de grande valia para conclusão da tese, minha eterna gratidão

Ao Prof. Dr. **Francisco Adelson Alves Ribeiro**, Coorientador desta tese, Professor do Instituto Federal do Maranhão (IFMA – Codó/MA), pelo incentivo e sua brilhante coorientação o meu muito obrigado.

Ao Prof. Dr. **Otílio Paulo da Silva Neto**, Coorientador desta tese, Professor do Instituto Federal do Piauí (IFPI – Teresina/PI), pelo incentivo e sua brilhante coorientação a sua ajuda foi primordial para conclusão desta tese o meu muito obrigado.

À **Universidade Federal do Piauí** pela oportunidade de crescimento intelectual e profissional.

Ao Magnífico Reitor, **Prof. Dr. Gildásio Guedes Fernandes**, e Vice-Reitor, **Prof. Dr. Viriato Campelo**, pela existência desta pós-graduação stricto sensu e grande incentivo à pesquisa em nossa Instituição.

Ao **Prof. Dr. Daniel Dias Rufino Arcanjo**, coordenador da RENORBIO Ponto Focal do Piauí, pelo apoio e incentivo à pesquisa stricto sensu no Piauí.

A Direção do **HGV**, pela acolhida no laboratório da Coordenação de Ginecologia durante a realização deste estudo.

Aos **professores da RENORBIO** pelos ricos ensinamentos e preciosas contribuições.

Às secretárias do RENORBIO, **Deusilene** e **Eliana**, pela ótima recepção.

Ao **Prof. Dr. Otílio Paulo da Silva Neto**, pela amizade, palavras de encorajamento, sempre disposto a ajudar no desenvolvimento do software e escutar minhas preocupações, enfim, amigo para todas as horas.

Aos amigos do **Grupo de pesquisa do professor Dr. Benedito Borges**, dentre eles: **Carla Solange, Danylo Rafael, Diego Cipriano, Maria da Conceição, Larysse Maira, Cléciton Braga, Camila Revoredo, Gilmara Péres, Luana Martins, Fabiane Araújo, Renato de Oliveira, Victor Alves**, agradeço pela amizade e incentivo.

Aos **funcionários da Ginecologia e Mastologia do Hospital Getúlio Vargas** em particular a dona Eugênia, pela recepção e amizade.

Aos **colegas de trabalhos do Unifacema**, Patrícia Lemos, Val Sales, Thais Beleza, Maisa Beleza, Daiana, Debora, Laís, Camila, Cristina, Camilo, Humberto, Adilson, Jonathan, e Luciane Lima, agradeço pela amizade e incentivo.

NERY-JUNIOR, ELMO DE JESUS. Método de Redes Neurais para Classificar Nódulos Mamários Baseado no Padrão BI-RADS® Nas categoria 4 e 5. __. **Tese de Doutorado** (Programa de Pós-Graduação em Biotecnologia- RENORBIO - Ponto Focal Piauí) - Universidade Federal do Piauí, Teresina, 2021.

O câncer de mama é uma neoplasia que mais acomete as mulheres em todo mundo, isso ocorre tanto em países de primeiro mundo quanto em países em desenvolvimento. No Brasil, para o biênio 2020 a 2021, foram estimados 66.280 casos novos de câncer de mama e cerca de 15.000 mortes de mulheres pela doença, com um risco estimado de 43,64 casos novos para cada 100 mil mulheres. Entretanto há estratégia para se reduzir a mortalidade por câncer de mama. A mamografia é um exame radiológico que gera uma imagem em tons de coloração cinza da mama, que o especialista analisa e identifica de forma visual a localização da lesão, sendo este tipo de exame o padrão para detecção precoce deste tipo de neoplasia, o que facilita a identificação na sua fase inicial, em particular pela sistematização internacional para avaliação mamária conhecida por Breast Imaging Reporting and Data System (BI-RADS), que é de suma importância na abordagem terapêutica. Todavia, trata-se de um sistema de classificação de imagens de lesões mamográficas que apresenta um resultado de falso negativo em cerca de 15%, taxa que pode ser reduzida por sistemas computacionais utilizando as redes neurais. A presente pesquisa desenvolveu um método de redes neurais para classificar nódulos mamários no padrão BI-RADS nas categorias 4 e 5. Esta tese foi estruturada em dois capítulos: O capítulo I realizou uma busca de anterioridade dos registros de programa de computador que faz uso de redes neurais artificiais aplicadas ao diagnóstico de câncer de mama, por meio de uma prospecção tecnológica, a qual foi realizada em três bases de depósitos de patentes: *European Patent Office* (EPO), *World Intellectual Property Organization* (WIPO) e *United States Patent and Trademark Office's* (USPTO). Utilizando as seguintes combinações de palavras-chave: software AND image processing, software AND breast cancer, software AND BI-RADS, software AND neural networks, software AND image processing AND breast cancer AND BI-RADS AND neural networks. As bases de depósitos de patentes que mais se destacaram foram WIPO e USPTO. Encontrou-se um total de 24 patentes, sendo 22 na WIPO e 2 na USPTO, relacionadas a imagens de câncer mamário e sistema BI-RADS. Ressalta-se que a busca feita na base de dados EPO não retornou registros relacionados ao tema pesquisado, demonstrando a necessidade de desenvolvimento de software para o diagnóstico do câncer de mama. O capítulo II exibe as técnicas de redes neurais para desenvolver uma ferramenta computacional capaz de classificar lesões das categorias 4 e 5 do padrão BI-RADS. Os resultados adquiridos pelo software, observaram que o melhor classificador no que diz respeito à taxa de acerto acurácia, foi o Deep Learning, atingindo um percentual de 82,60%, o *Support Vectors Machine* - SVM teve um percentual de 73,97%. Isto demonstra que as técnicas de redes neurais utilizadas no projeto do software mostram uma eficácia na tarefa de classificação das lesões.

Palavras-chave: Câncer de Mama, BI-RADS, Redes Neurais.

ABSTRACT

NERY-JUNIOR, ELMO DE JESUS. Neural Network Method To Classifying Breast Nodules Based on the BI-RADS® Standard In Category 4 and 5. __. Doctoral Thesis (Postgraduate Program in Biotechnology - RENORBIO - Focal Point Piauí) - Federal University of Piauí, Teresina, 2021.

Breast cancer is a neoplasm that most affects women around the world, it occurs both in first world countries and in developing countries. In Brazil, for the biennium 2020 to 2021, there were an estimated 66,280 new cases of breast cancer and about 15,000 deaths of women from the disease, with an estimated risk of 43.64 new cases for every 100,000 women. However, there is a strategy to reduce mortality from breast cancer. Mammography is a radiological exam that generates a grayscale image of the breast, which the specialist analyzes and visually identifies the location of the lesion. This type of exam is the standard for early detection of this type of neoplasm, which facilitates identification in its initial phase, in particular by the international systematization for breast evaluation known as the Breast Imaging Reporting and Data System (BI-RADS), which is of paramount importance in the therapeutic approach. However, it is a mammographic lesion image classification system that presents a false negative result in about 15%, a rate that can be reduced by computational systems using neural networks. The present research developed a method of neural networks to classify breast lumps in the BI-RADS standard in categories 4 and 5. This thesis was structured in two chapters: Chapter I carried out a search for prior records of a computer program that makes use of artificial neural networks applied to the diagnosis of breast cancer, through a technological prospection, which was carried out in three patent filing bases: European Patent Office (EPO), World Intellectual Property Organization (WIPO) and United States Patent and Trademark Office's (USPTO). Using the following keyword combinations: software AND image processing, software AND breast cancer, software AND BI-RADS, software AND neural networks, software AND image processing AND breast cancer AND BI-RADS AND neural networks. The most prominent patent filing bases were WIPO and USPTO. A total of 24 patents were found, 22 from WIPO and 2 from USPTO, related to breast cancer imaging and the BI-RADS system. It is noteworthy that a search made in the EPO database did not return records related to the researched topic, demonstrating the need to develop software for the diagnosis of breast cancer. Chapter II shows the neural network techniques to develop a computational tool capable of classifying injuries from categories 4 and 5 of the BI-RADS standard. The results acquired by the software, observed that the best classifier with regard to the accuracy rate was the Deep Learning, reaching a percentage of 82.60%, the Support Vectors Machine - SVM had a percentage of 73.97%. This demonstrates that the neural network techniques used in the software design show an effectiveness in the lesion classification task.

Keywords: Breast Cancer, BI-RADS, Neural Networks.

2 REVISÃO DA LITERATURA

Figura 1. Estrutura da mama	29
Figura 2. Mamógrafo	32
Figura 3. Estrutura de uma rede neural	47

CAPÍTULO 1

Figuras 1. Número de patentes por banco	64
Figura 2 Numero de patente por país	64
Figura 3 Número de patente por área de classificação	65
Figura 4 Porcentagem de depositantes	66
Figura 5 Depósito de patentes por área	66
Figura 6 Número de patentes por área de classificação	67

CAPÍTULO 2

Figuras 1 Resultados das etapas do melhoramento	76
Figura 2 Processamento OTSU	77
Figura 3 Resultados dos Clusters	78
Figura 4 Resultados do algoritmo Graph Clustering	79
Figura 5 Nódulo em fundo preto	80
Figura 6 Visualização das dimensões das lesões	81
Figura 7 Base treinamento, validação, teste	81

LISTA DE QUADROS E TABELAS

CAPÍTULO 1

Quadro 1 Quantidade de registros63

CAPÍTULO 2

Quadro 1 Resultados dos classificadores83

ACR *American College of Radiology*

BI-RADS *Breast Imaging Reporting and Data System*

CAD Diagnóstico auxiliado por computador, do inglês: *Computer Aided Diagnosis*

CADe Detecção auxiliada por computador, do inglês: *Computer Aided Detector*

CADx Diagnóstico auxiliado por computador, do inglês: *Computer Aided Diagnosis*

CC Crânio-caudal

CLAHE *Contrast Limited Adaptive Histogram Equalization*

DDSM *Digital Database for Screening Mammography*

FDC Função de Distribuição Cumulativa

IPC Classificação Internacional de Patentes

MLO Médio-Lateral-Oblíqua

MM Morfologia Matemática

PSO *Particle Swarm Optimization*

RFP Redução de Falsos Positivos

ROIs Regiões de Interesse

SVM Máquina de Vetores de Suporte, do inglês: *Support Vector Machines*

SUMÁRIO

1 INTRODUÇÃO	21
2 REVISÃO DA LITERATURA	25
2.1 CANCER DE MAMA	26
2.1.1 TIPOS DE LESÕES	28
2.1.2 MAMAS	29
2.1.3 MAMAS FEMININAS	30
2.2 MAMOGRAFIA	31
2.2.1 RASTREAMENTO MAMOGRÁFICO	34
2.3 CLASSIFICAÇÃO BI-RADS®	37
2.4 SISTEMAS CAdE e CADx	40
2.5 PROCESSAMENTO DIGITAL DE IMAGEM	42
2.5.1 AQUISIÇÃO DE IMAGEM	42
2.5.2 PRÉ-PROCESSAMENTO	43
2.5.3 SEGMENTAÇÃO	44
2.5.4 ALGORITMO OTSU	44
2.5.5 <i>PARTICLE SWARM OPTIMIZATION (PSO)</i>	44
2.5.6 GRAPH CLUSTERING	45
2.5.7 EXTRAÇÃO DE CARACTERISTAS	45
2.5.8 RECONHECIMENTOS DE PADRÕES	46
2.6 REDES NEURAI	46
2.6.1 TIPOS DE APRENDIZAGEM	48
2.6.2 REDES NEURAI CONVOLUCIONAL	48
2.7 MÉTRICAS DE AVALIAÇÃO	49
2.8 JUSTIFICATIVA	52
REFERENCIA	53
3 CAPÍTULOS	56
3.1 CAPÍTULO 1 Contribuição de Redes Neurais em Sistemas Computacionais para o Diagnóstico do Câncer de Mama	58
3.2 CAPÍTULO 2 Classificação de lesões mamárias das categorias 4 e 5 do padrão BI-RADS utilizando redes neurais	71
4 CONSIDERAÇÕES FINAIS	89
ANEXOS	90
APÊNDICES	93

1 INTRODUÇÃO

O câncer de mama é a neoplasia maligna mais incidente em mulheres na maior parte do mundo, caracterizado especificamente pela divisão incorreta e desordenada das células na região dos seios, motivo pelo qual leva este nome, podendo ocorrer tanto no corpo masculino quanto no feminino, no entanto, são as mulheres que têm a massacrante maioria dos casos de câncer mamário. (SANTOS, 2018).

Segundo as estatísticas extraídas pelo Observatório Global do Câncer (GLOBOCAN, 2020), em 2020, foram estimados 2,2 milhões de casos novos de câncer e 690 mil óbitos pela doença em 185 países do globo, comprovando que este tipo de câncer é o que mais afeta a população no mundo atual.

A estimativa para o Brasil, de acordo com Instituto Nacional do Câncer (INCA), para biênio de 2020-2021 é o diagnóstico de 66.280 novos casos de câncer de mama que corresponde a uma taxa 29,7% dos casos na população feminina e cerca de 15.000 mortes pela doença, com um risco de aproximadamente 43,74 para cada 100 mil mulheres (INCA, 2019). Todavia há estratégias para se reduzir a mortalidade por câncer de mama. A mamografia é o exame mais utilizado para a detecção precoce deste tipo de neoplasia, o que facilita a identificação de lesões mamárias em sua fase inicial, fator primordial para o sucesso do tratamento (RIBEIRO, 2017).

A mamografia, também chamada mamograma ou mamografia digital, é um exame realizado que submete as mamas a raios-X para produzir imagens de alta resolução e permitir que o especialista analise e identifique de forma visual onde se encontra a lesão (GIGER, 2000); (MOORE, 2019). De acordo com estudos observacionais, quando realizado o rastreamento por meio da mamografia visando a detecção precoce, reduz-se o índice de mortalidade em mulheres acometidas pelo câncer de mama em cerca de 15% a 25% (SALA, 2021).

O diagnóstico prévio é a melhor estratégia para prevenção desse tipo de câncer, os exames clínico de imagem da mama são os mais recomendados, com destaque a mamografia se torna um dos métodos mais eficazes (GIESS, 2012); (TORTORA, 2016); (PRADO, 2020); (BALAS, et all 2020). Ressalta-se que a

leitura e interpretação das imagens são tarefas difíceis que exigem experiência do radiologista visto que mesmo um profissional bem treinado apresenta uma taxa de acertos entre 65% a 75% (JUSTO, et al 2013).

No entanto, apesar de todo o avanço tecnológico e científico, era comum que houvesse confusão entre médicos que tratam o câncer de mama perante a análise dos diagnósticos e dos laudos, o que causava problemas aos pacientes relativos às abordagens terapêuticas (GUIMARÃES; CHOINIÁK, 2019).

Com base nestas divergências, no ano de 1993, o Colégio Americano de Radiologia, padronizou os achados mamográficos para que os especialistas “falassem a mesma língua” reduzindo a confusão na interpretação dos laudos, foi criado o sistema de classificação BI-RADS - *Breast Image Reporting and Data System*, tem o papel importante na indicação de biópsia em lesões suspeitas ACR (2016). A sua classificação é dividida por categoria variando de 0 a 6, em cuja classificação a categoria 0: incompleta; 1 e 2 achado benigno; 3 achado provavelmente benigno; 4 achado suspeito variando de 2% a 95%; 5 achado maligno e 6 malignidade comprovada por biópsia.

A mamografia apresenta falha na sua interpretação com uma quantidade de falso negativo de 15% e um falso positivo variando de 7% a 10% causando repercussão emocional e física para as pacientes, (RIBEIRO, 2017). No entanto, ferramentas computacionais podem auxiliar os profissionais médicos na interpretação e no diagnóstico dessas imagens, como por exemplo, os sistemas Computer Aided Design – CADs, os quais aumentam a sensibilidade, especificidade e acurácia do câncer da mama. Tais ferramentas aplicadas em conjunto com as técnicas de redes neurais para o processamento de imagens mamográficas, podem melhorar e extrair informações úteis ao diagnóstico preciso do câncer de mama melhorando o índice de acertos médicos (EL ATLAS et al., 2014) (GANESAN et al., 2012).

De acordo com as evidências expostas, a pesquisa tem como objetivo desenvolver um método de redes neurais para classificar nódulos mamários no padrão BI-RADS® nas categorias 4 e 5, para auxiliar o diagnóstico médico e teve como resultado o registro de programa de computador intitulado **SPIBIRADS** no Instituto Nacional de Propriedade Industrial – INPI com o nº **BR512019001754-**

4.

O conteúdo descrito neste estudo está estruturado sob a forma de dois capítulos, conforme segue: O capítulo I trata de um artigo de prospecção que investigou sobre a quantidade de patentes/registros de software relacionado a diagnóstico médico, processamento de imagens mamográficas, classificação BI-RADS e redes neurais que resultou em um manuscrito intitulado “Contribuição de redes neurais em sistemas computacionais para o diagnóstico do câncer de mama”. O capítulo II descreve sobre técnicas de redes neurais utilizado para classificar as lesões mamárias no padrão BI-RADS®, que resultou no artigo intitulado “Classificação de lesões mamárias das categorias 4 e 5 do padrão BI-RADS® utilizando redes neurais”.

2 REVISÃO DE LITERATURA

2.1 CANCER DE MAMA

O câncer de mama é a neoplasia maligna mais incidente em mulheres na maior parte do mundo, caracterizado especificamente pela divisão incorreta e desordenada das células, motivo pelo qual leva este nome, podendo ocorrer tanto o corpo masculino quanto o feminino, no entanto, são as mulheres que tem a maioria dos casos de câncer mamário, conforme explicam (SANTOS; GONZAGA, 2018, p.359):

Todo câncer é caracterizado por um crescimento rápido e desordenado de células, quando as células adquirem características anormais, células dos lobos mamários, células produtoras de leite ou dos ductos por onde é drenado o leite, podem causar uma ou mais mutações no material genético da célula. Esta doença acontece quase exclusivamente em mulheres, contudo existem casos de homens com câncer de mama também.

Segundo as estatísticas extraídas pelo Observatório Global do Câncer (GLOBOCAN, 2020), em 2020, foram estimados 2,2 milhões de casos novos de câncer e 690 mil óbitos pela doença em 185 países do globo, comprovando que esta espécie de câncer é o que mais afeta a população no mundo atual.

Conforme o INCA (2019), o conhecimento da doença em seu processo inicial e o tratamento sistêmico e adequado são os fatores preponderantes para evitar que o câncer de mama alcance seus estágios mais avançados e dificultem o prognóstico, aumentando as expectativas de cura e reduzindo os riscos de metástase, ora, quando as células cancerosas se disseminam do local onde se originou para outros órgãos, complicando a recuperação do paciente, visto que os tratamentos deverão alcançar outros órgãos além da mama (INCA, 2019).

Oliveira (2019) explica o motivo da urgência no diagnóstico do câncer de mama é de que as taxas de sobrevida são mais favoráveis às pessoas que iniciam os estudos quando descobrem a doença em estágio inicial, vez que quanto maior o tamanho inicial do tumor e o comprometimento da região, menor será a sobrevida da paciente, logo, menos tempo para realização do tratamento e menores chances de recuperação sem sequelas maiores.

Ainda, Conforme Oliveira (2019, p. 19):

A neoplasia da mama é constituída por uma grande variabilidade em seu prognóstico, que é determinada por fatores prognósticos associados à sobrevida dos pacientes acometidos pela doença, permitindo analisar o comportamento do tumor Segundo sua evolução. Existem os fatores clássicos, que estão relacionado ao grau histológico e ao estadiamento e existem os subtipos moleculares, que além de direcionarem o tratamento adequado, oferecem a possibilidade de predizer a sobrevida do paciente. O grau histológico é considerado um dos fatores mundialmente aceitos como

indicadores de prognóstico e está associado com a capacidade proliferativa do tumor. [...] o estadiamento do tumor indicará suas características morfológicas ou anatômicas, sua taxa de crescimento, sua extensão e a relação estabelecida com o portador.

Outro fator importante no que tange o tratamento do câncer de mama é a análise do padrão molecular das células cancerosas, pois com tal estudo se torna possível identificar qual será o padrão de resposta deste às terapias específicas e o prognóstico da doença, assim aumentando a assertividade nos tratamentos realizados, aumentando a possibilidade de sucesso no combate às células malignas, reduzindo o tempo de tratamento e também causando menores sofrimentos ao paciente, que antes realizava diversos tratamentos, e muitos deles podendo ser ineficazes frente ao seu tipo específico de câncer (INCA, 2019).

A referida avaliação pode ser realizada através dos métodos imunohistoquímica e microarranjo: o primeiro tem função de quantificar tanto os receptores hormonais e de crescimento do tumor, enquanto o segundo método caracteriza os tumores em subtipos moleculares e os classifica conforme suas peculiaridades na pretensão de padronizar as formas de ação prognóstica. Nessa situação, o plano de estratégias de controle do desenvolvimento celular por meio da detecção prematura se faz iniciado (INCA, 2019).

Quanto mais cedo um tumor é detectado e o tratamento é iniciado, maior a probabilidade de cura. Por esse motivo, várias pesquisas e ações vêm sendo implementadas pelos estudiosos para diagnosticar o câncer nos estágios iniciais. No mesmo aspecto, os pesquisadores da área concentram seus esforços em busca de biomarcadores cuja função é identificar de maneira mais assertiva e específica os tipos de blastomas, seja benigna ou maligna, encaminhando a resposta clínica mais adequada o tratamento mais apropriado para cada paciente (INCA, 2019).

Desta forma, em razão da individualização orgânica e da extrema heterogeneidade tumoral associada à presença de fatores de risco conhecidos e não conhecidos, o câncer de mama é considerado uma doença de comportamento dinâmico, em constante transformação (INCA, 2019).

2.1.1 TIPOS DE LESÕES

Há uma variedade considerável de espécies de carcinomas de mama comuns e incomuns, com classificações histológicas e moleculares. No entanto, merece destaque quanto à forma das lesões, visto que são descritas por três características: massa (também denominada por “nódulo”), microcalcificações, assimetria focal e difusa, dilatação ductal isolada e distorção focal de arquitetura (INCA, 2007). A lesão denominada “massa” é definida segundo sua forma e densidade, se resumindo a qualquer opacidade de forma arredondada; as “microcalcificações” são estruturas cálcicas que se acumulam e são classificadas pela morfologia e distribuição; a “assimetria focal e difusa” representa uma região com massa volumar diferente da mama e similar à densidade do tecido epitelial glandular; a “dilatação ductal isolada” é, de fato, a dilatação das glândulas mamárias; e, a “distorção focal de arquitetura” é quando se verifica uma distorção da arquitetura da ducto mamário, sem um nódulo definido (INCA, 2007).

Conforme estão dispostas as lesões, torna-se possível identificar o grau do tumor, podendo este ser classificado como *in situ* ou invasor, também chamado por infiltrante. O tipo de carcinoma *in situ* é conhecido como tumor benigno, visto que este não ultrapassou a membrana basal, logo não se propagou aos demais órgãos, permanecendo no local em que encontra-se, sendo menos radical o tratamento por limitar-se a uma região, contudo a gravidade deste é contínua, pois pode tornar-se um neoplasma invasivo (ONCOGUIA, 2020).

Sobre o tumor infiltrante, este propaga-se a outros órgãos, tendo em vista que já penetrou a membrana basal e as células cancerosas multiplicaram-se a outros aparelhos, ocasionando a metástase, caso não seja tratado precocemente, o que exige maior cuidado e urgência nos procedimentos a serem tomados para evitar males maiores ao paciente (ONCOGUIA, 2020). Quanto aos casos de lesões incomuns, estas não serão abordadas no presente estudo, no entanto merecem sua citação à título acadêmico, entre os quais estão as: hiperplasiatípica e atípica, hamartoma, lipoma retromamário, tumor filóide benigno e maligno, fibroadenoma atípico, ectasia ductal gigante, doença de Mondor, rabdomyossarcoma, carcinoma adenóide cístico e linfoma (ONCOGUIA, 2020).

Neste aspecto, dados os tipos e especificações das lesões que são

características no que chamamos de câncer de mama, merece um esclarecimento mais assertivo quanto o que é a mama, tanto no corpo masculino quanto no feminino, e sua composição biológica. No entanto, dado que o presente estudo baseia-se no câncer de mama e esta enfermidade é mais frequentemente encontrada no corpo da mulher, será este o foco do próximo subitem.

2.1.2 MAMAS

Moore (2019) descreve as mamas como estruturas superficiais e mais proeminentes na parede anterior do tórax, figura 1, em geral no corpo feminino, vez que nos homens, as mamas não são desenvolvidas. As mamas se formam por tecidos glandular e fibroso para sua sustentação integrados a uma matriz de gordura, junto com vasos sanguíneos, vasos linfáticos e nervosos.

Homens e mulheres, ambos têm mamas, geralmente, elas são bem desenvolvidas no corpo feminino. As mulheres tem as glândulas mamárias como acessórios à reprodução do ser humano, contudo, no corpo masculino, as mamas são rudimentares e não têm finalidade, consistindo apenas em pequenos ductos e embaixa numerosidade ou cordões epiteliais. Em sua maioria, no homem, há pouca gordura na região mamária, motivo pelo qual as glândulas da mama masculina não apresentam desenvolvimento. Por sua vez, nas mulheres, elas fazem parte dos caracteres sexuais secundários e são estruturas superficiais mais acentuadas na parede do tórax anterior e está apoiada sobre o músculo peitoral maior como mostra na Figura 1.

Figura 1: Estrutura da Mama



Fonte: (INCA, 2019)

As glândulas mamárias localizam-se no tecido subcutâneo, elas sobrepõem os músculos maior e menor do peitoral. O tamanho das mamas é determinado segundo a quantidade de gordura que circula o tecido glandular. No ponto mais alto, mais elevado da mama é situada a papila mamária, e está envolvida por uma região de pele em formato oval-circular pigmentada: a aréola (do latim, “área pequena”), cuja coloração se dá de acordo com a cor da pele do indivíduo. A aréola geralmente escurece durante a gravidez e essa pigmentação escurecida persiste (MOORE, 2019).

2.1.3 MAMAS FEMININAS

É durante a puberdade – entre 8 e 15 anos - que as mamas femininas normalmente crescem, em razão do desenvolvimento glandular e da deposição aumentada de gordura. As aréolas e papilas mamarias também aumentam. O tamanho e formato da mama resultam em fatores genéticos, raciais e alimentares. Os ductos lactíferos dão origem a brotamentos que formam 15 a 20 lóbulos de tecido glandular, que constituem a glândula. Cada lóbulo da mama é drenado por um ducto lactífero que, em geral, se abre, independentemente, na papila mamaria. Os ductos convergem para a papila mamária como raios de uma roda de bicicleta. Profundamente à aréola, cada ducto possui uma porção dilatada, o seio lactífero no qual uma pequena gotícula de leite se acumula ou permanece na mãe que amamenta (MOORE, 2019).

Ao longo da vida da mulher, a mama sofre alterações à medida que esta envelhece e conforme outros fatores que podem influenciar neste quesito, como hormonal e gravidez, pois, que quando adolescente, a mama da mulher é constituída majoritariamente por tecido fibroglandular e pouco percentual de gordura. Com o amadurecimento da adolescente, ocorre uma substituição lenta e gradual do tecido fibroglandular por tecido gorduroso, até que, na pós-menopausa, a mama é majoritariamente constituída por tecido gorduroso. Bem como quanto às mulheres multíparas, as mamas frequentemente se tornam grandes e pendentes, enquanto nas mulheres idosas são pequenas e encolhidas devido à redução de gordura e à atrofia do tecido glandular (MOORE, 2019).

Relacionado à anatomia da mama e a neoplasia mamária, ensina Keith Moore (2019, p. 67)

Os carcinomas da mama são tumores malignos, geralmente adenocarcinomas originados nas células epiteliais dos ductos lactíferos nos lóbulos das glândulas mamárias. As células cancerosas metastáticas que entram em um vaso linfático geralmente atravessam dois ou três grupos de linfonodos. A interferência do câncer na drenagem linfática dérmica pode causar linfedema (excesso de líquido na tela subcutânea) na pele da mama que, por sua vez, pode resultar em desvio da papila mamária e aspecto espesso e coriáceo da pele.

Um detalhe importante é que a mama é dividida em quadrantes (quatro partes iguais) para facilitar sua localização e melhorar a compreensão descritivo-anatômica de carcinomas, lesões, cistos ou outras anomalias identificadas na região. Quanto à menor suspeita de lesão na mama, a investigação se faz necessária. O exame das mamas por técnicas de imagem detecta anormalidades mamárias, diferenciando cistos e massas neoplásicas de variações na arquitetura mamária. À primeira vista, a mamografia é o estudo mais indicado em caso de prognose de lesões mamárias, vez que detecta anormalidades mamárias, diferenciando cistos e massas neoplásicas de variações na arquitetura mamária.

2.2 MAMOGRAFIA

A mamografia, também chamada mamograma ou mamografia digital, é um dos exames realizados que submete as mamas a raios-X para produzir imagens de alta resolução e permitir a detecção da lesão na mama de forma mais assertiva (MOORE, 2007). Como mostra na Figura 2:

Figura 2: Mamógrafo Digital



Fonte: (SAFETY, 2018)

Conforme Keith Moore (2007, p. 67):

A mamografia é um estudo radiográfico das mamas, que são achatadas para estender a área que pode ser examinada e para reduzir a espessura das mamas, tornando-as mais uniformes e melhorando a visualização. A mamografia é realizada basicamente para rastreamento de condições ainda não evidentes clinicamente. Muitas vezes o carcinoma apresenta-se como uma área densa grande e irregular na imagem de mamografia. A pele sobre o tumor é espessa e há depressão da papila mamária. Na mamografia convencional as estruturas mais densas aparecem mais claras, enquanto na xeromamografia, as estruturas mais densas (estroma normal, ductos e tumor) aparecem mais escuras.

Além disso, Gerótica, Azevedo e Sanches (2015, n.p.) enfatizam a importância da mamografia:

A mamografia é o exame mais confiável para detectar o início de qualquer alteração das mamas, antes que o paciente ou o médico possam nota-las. Dado a frequência do câncer de mama a mamografia deve ser incluída como exame preventivo de rotina para todos os pacientes principalmente mulheres que façam parte do grupo de risco. Toda mulher acima de 40 anos deve submeter-se à mamografia preventiva pelo menos uma vez ao ano.

O objetivo da mamografia é gerar quatro imagens de alta resolução em tons de cinza das estruturas internas da mama, com o intuito de detectar anomalias. As principais anomalias visualizadas a partir de uma mamografia, em geral, são calcificações e massas (GERÓTICA, 2015), contudo as microcalcificações são as lesões mamárias não palpáveis mais frequentes.

A mamografia deve ser realizada por um aparelho de raios X específico, conhecido como mamógrafo, que pode ser convencional ou digital: o primeiro utiliza o filme para a detecção da imagem após a exposição da mama aos raios X, enquanto o segundo utiliza um detector que transforma os raios X em sinais elétricos. Dessa maneira, no sistema digital existe a possibilidade de usar monitores de alta resolução para a leitura das imagens, técnicas de pós-processamento para melhorar a qualidade da imagem e a capacidade diagnóstica, permitindo o armazenamento de imagens para verificações posteriores (PISANO et al, 2005).

Estudos que compararam a acurácia diagnóstica das duas modalidades. E, geral, não demonstraram diferença significativa nos diagnósticos, entretanto houveram diversidades quando foram comparados grupos específicos, como em pacientes jovens, mulheres com mamas densas e em fases pré e perimenopausa, o mamógrafo digital apresentou superioridade nas imagens e no diagnóstico

comparados aos realizados pela modalidade convencional (TOSTESON et al, 2008).

Este procedimento visa radiografar as duas mamas, por isso na mamografia deve ser realizada duas projeções para cada mama, sendo essas, a médio-lateral oblíqua e a crânio-caudal. A primeira projeção permite visualizar do alto da axila para baixo, incluindo a prega infla-mamária e o músculo peitoral, estendendo-se obliquamente sobre a mama. A segunda projeção permite visualizar a região pósteromedial da mama, complementando a visão da projeção médio-lateral oblíqua (SAMPAIO, 2011).

Ribeiro (2017, p. 32) esclarece a importância das projeções médio-lateral-oblíqua e do crânio caudal da seguinte forma:

No entanto, incidência médio-lateral-oblíqua é a mais eficaz, pois mostra uma quantidade maior de tecido mamário e inclui estruturas mais profundas do quadrante súpero-externo e do prolongamento axilar. A incidência crânio-caudal tem como objetivo incluir todo o material pósteromedial, complementando a médio-lateral-oblíqua. Permite também mais compressão da mama, uma vez que não inclui a axila, resultando em uma definição superior da arquitetura mamária e de lesões. Os radiologistas estudam as incidências crânio-caudais e as médio-laterais aos pares de modo a permitir a comparação de regiões simétricas, pois qualquer assimetria pode ser indício de patologia.

Em virtude do fato de utilizar radiação ionizante, não deve ser realizada em pacientes com idade anterior aos vinte e cinco anos, assim como não devem ser submetidas à radiologia mamográfica as pacientes grávidas ou com suspeita de estarem em estado gestacional. Ora, essas limitações não alteram o fato de a mamografia ser de tamanha importância no diagnóstico das lesões na mama, devendo haver o devido rastreamento mamográficos no intuito de averiguar as situações em que não há sintomas tumorais, no entanto se fazem presentes outros fatores que podem ocasionar o neoplasma maligno (SILVA NETO, 2016).

A mamografia não detecta todos os cânceres de mama, de modo que conclui-se que apesar de sua assertividade ser alta, não é absoluta, sendo limitada devido às variações possíveis conforme as características de cada paciente. A sensibilidade do exame varia de 50% a 95%, sendo a densidade do parênquima mamário e a idade os fatos preditores mais importantes da assertividade. A maior sensibilidade ocorre com as mamas lipossustituídas¹ e em pacientes idosas, enquanto a menor sensibilidade

¹ Denominação das mamas com predominância de gordura.

ocorre nas pacientes e com mamas densas. Outra limitação é o risco falso-positivo, que pode resultar em biopsias desnecessárias, ansiedade e custo adicional. Essas limitações são maiores em pacientes na faixa de idade entre quarenta e quarenta e nove anos, quando comparadas às com idade superior a cinquenta anos (SILVA NETO, 2016).

Atualmente, há duas modalidades de acompanhamentos mamográficos específicos: para rastreamento e para diagnóstico. Eis que o mamograma para diagnóstico é aquele realizado em mulheres que já apresentam algum sintoma que pode indicar o tumor, enquanto rastreamento mamográfico é indicado nos casos de mulheres com idade superior aos 40 anos e não apresentam nenhum sintoma ou anomalia em suas mamas (SILVA NETO, 2016). Apesar da importância do mamograma de diagnóstico, trataremos no próximo capítulo tratando apenas sobre a modalidade de rastreamento na mamografia.

2.2.1 RASTREAMENTO MAMOGRÁFICO

A avaliação da eficácia de um teste proposto para a detecção precoce de um neoplasma requer o conhecimento dos princípios do rastreamento. As características da doença, do teste de diagnóstico proposto e do tratamento podem influenciar o impacto de um programa de rastreamento. A palavra rastreio, ou triagem, é derivada do inglês “*screening*”, e provem da ideia de peneira – do inglês “*sieve*” -, rica em furos, ou seja, todos os programas possuem resultados falso-positivos e falso-negativos e a ideia é selecionar, filtrar os resultados para melhores análises e certezas (GUIMARÃES; CHOJNIAK, 2019).

Para que o rastreamento tenha sucesso, se faz necessário o conhecimento da patologia específica que está sendo rastreada, conhecer sua incidência, os fatores de risco e seu comportamento biológico, incluindo a frequência de cânceres indolentes. Patologias muito raras, com baixa agressividade ou com índices elevados de variantes pouco agressivas podem comprometer a relação de custo-benefício do rastreamento. Os pesquisadores Marcos Duarte Guimarães e Rubens Chojniak (2019, p. 6) definem o acompanhamento de rastreio patológico como:

O rastreio pode ser definido como a realização sistemática de testes diagnósticos em busca de doença pré-clínica. Portanto, não existe rastreio em pacientes sintomáticos, pois estes já se encontram na fase clínica da doença.

O câncer de mama é um problema de saúde pública comum. Neste viés, estratégias de rastreamento mamográfico são aceitas como procedimento de escolha na maioria dos países para averiguação de pacientes assintomáticas com características que compõem uma classe de pessoas que o tumor mamário é bastante incidente. Desta forma, novas abordagens que adicionem maior acurácia ao diagnóstico dentro de um impacto econômico e social aceitável devem ser investigadas. Nas mamografias de rastreamento geralmente são obtidas imagens de cada mama em dois ângulos diferentes (RIBEIRO, 2017). O rastreamento mamográfico - também conhecido como mamografia de rotina - é realizado para verificar se existem sinais de doença na mama na ausência de sintomas ou alterações da mama, confirmando a importância da mamografia, Guimarães e Chojniak (2019, p. 9) são enfáticos que "a mamografia é a única modalidade de imagem que, quando implementada no rastreio do câncer de mama, comprovadamente possibilita a redução na mortalidade da doença". Mais especificamente,

O rastreamento mamográfico detecta 5 a 7 casos de câncer a cada 1.000 mulheres assintomáticas que se submetem ao primeiro exame e 2 a 3 casos a cada 1.000 mulheres que se submetem ao rastreamento sequencial. [...] Estima-se que desde a primeira divisão celular anômala até um nódulo mamário palpável de 1 cm exista um intervalo que pode chegar a até 10 anos. Nesse período, o melhor método diagnóstico com ação comprovadamente eficiente como *screening* é a mamografia pelo potencial de identificar lesões com 15 a 20 duplicações, medindo aproximadamente 0,5cm, ainda em fases pré-clínicas e, conseqüentemente, com evidentes melhoras e chances de cura e sobrevida da paciente (GUIMARÃES; CHOJNIAK (2019, p. 432).

Merece a devida atenção quanto à distinção entre as expressões "rastreamento" e "diagnóstico de doenças". Vez que quando um indivíduo exhibe sinais e sintomas de uma doença e um teste diagnóstico é realizado, este não representa um rastreamento. No rastreamento, exames ou testes são aplicados em pessoas saudáveis, o que implica, repetimos e enfatizamos, garantia de benefícios relevantes frente aos riscos e danos previsíveis e imprevisíveis da intervenção. Tanto o rastreamento como o diagnóstico podem usar vários métodos e exames (por exemplo, questionários, aparelhos portáteis para medir marcadores sanguíneos, como a glicose, colesterol, exames de laboratório etc.), assim como vários limiares ou pontos de corte para designar a condição (GUIMARÃES; CHOJNIAK, 2019).

Ainda no rastreamento, um exame positivo não implica fechar um diagnóstico, pois geralmente são exames que selecionam as pessoas com maior probabilidade de apresentar a doença em questão. Outro teste confirmatório (com maior especificidade

para a doença em questão) é necessário depois de um rastreamento positivo, para que se possa estabelecer um diagnóstico definitivo. Por exemplo, uma mamografia sugestiva de neoplasia deve ser seguida de uma biópsia e confirmação diagnóstica por anatomopatologia (GUIMARÃES; CHOINIAC, 2019).

Ainda, Guimarães e Chojniak (2019, p. 6):

A finalidade do rastreio é prevenir ou retardar o desenvolvimento de doença avançada mediante o diagnóstico pré-clínico e o tratamento precoce. Na prática clínica, este conceito é frequentemente ignorado ou mal compreendido. Muitos equivocadamente creem que a finalidade do rastreio é o diagnóstico e esquecem que o real objetivo de se rastrear é reduzir a mortalidade e aumentar a sobrevida.

No entanto, apesar de todo o avanço tecnológico e científico, era comum que houvesse confusão entre médicos, radiologistas e outros profissionais que tratam o câncer de mama perante a análise dos diagnósticos e dos laudos, o que causava problemas aos pacientes relativos às abordagens a serem tomadas e seguidas no rastreamento ou mesmo nos procedimentos incisivos no tratamento do neoplasma mamário. Estas diferenças ocorriam devido às tamanhas variedades das lesões mamárias existentes e a compreensão entre elas, pois quanto o avanço delas, alguns profissionais consideravam determinada lesão como em estado inicial e outros identificavam a mesma lesão como em situação intermediária, logo, as abordagens de procedimentos seriam divergentes, por fim, prejudicando a paciente (GUIMARÃES; CHOINIAC, 2019).

O BI-RADS foi concebido inicialmente como uma ferramenta exclusiva mamográfica, entretanto em 2003 sofreu uma atualização e foi disponibilizado também para a ultrassonografia e ressonância magnética. Sua versão mais atual (5ª edição), foi lançada em 2013, contudo trataremos apenas quanto à sua alçada no âmbito da mamografia a seguir.

2.3 CLASSIFICAÇÃO BI-RADS®

Esse termo BI-RADS vem da palavra em inglês da frase “*Breast Image Reporting and Data System*”, que em tradução livre significa “Sistema de Laudos e Registro de Dados de Imagem da Mama”, sendo este um sistema de padronização dos laudos em imagem da mama criado e desenvolvido pelo *American College of*

Radiology, em cooperação com outros institutos de saúde norte-americanos (ACR, 2016). Assim, sua principal função é reduzir a confusão na interpretação das imagens e facilitar o seguimento das pacientes além de proporcionar um sistema de auditoria interna de qualidade, devendo ser utilizado para diversos métodos diagnósticos. Conforme a descrição do *American College of Radiology* (2016, p.9) desenvolvedora do BI-RADS:

Um grande objetivo do BI-RADS é promover a consistência e a uniformidade dos laudos de exames de imagem da mama; assim, a aceitação dessa orientação quanto a um padrão de exibição uniforme das imagens destina-se a facilitar o cumprimento desse objetivo.

O BI-RADS foi desenvolvido devido às discrepâncias de entendimentos que ocorriam frequentemente entre os profissionais de radiologia e médicos quanto a um mesmo exame de mamografia, causando conflitos em entendimentos sobre determinada lesão mamária, ocasionando laudos diferentes de um mesmo prognóstico, vez que não havia uniformização na identificação e no modo de especificação das lesões (NASCIMENTO; SILVA; MACIEL, 2010).

Deste modo, o BI-RADS padronizou a linguagem médico-radiológica quanto às lesões na mama, melhorando a compreensão pelos profissionais da saúde dos laudos emitidos. Esse sistema, por sua vez, é baseado em uma classificação que permite ao profissional da saúde categorizar a lesão analisada de zero a seis. A partir dessa classificação, a lesão teria o seu estágio e o seu diagnóstico determinados; evitando diferenças nas interpretações, tornando assertiva a identificação do problema de saúde, e auxiliando nas recomendações de condutas. A classificação do BI-RADS segue uma estruturação como observam Nascimento, Silva e Maciel (2010, p. 92):

O BI-RADS é estruturado em quatro seções: seção I – léxico da imagem mamográfica; seção II – sistematização do laudo mamográfico; seção III – acompanhamento e monitoramento do resultado final; seção IV – criação de um banco de dados nacional.

A categoria 0 do BI-RADS deve ser reservada aos casos de rastreamento mamográficos incompletos, que uma avaliação adicional se faz necessária para a definição da *status* da lesão (ACR, 2016).

Os casos de BI-RADS 1 são denominados por “negativos” ou “não-achados”, que significa se tratar de um seio normal, sem problemas ou lesões, de modo que a mulher deve seguir realizando seus exames de rotina sem nenhuma preocupação com lesões enquanto estiver esta classificação (GUIMARÃES; CHOJNIK, 2014). Já o BI-RADS de categoria 2, são os “achados benignos” e tem valores preditivos de

malignidade próximos a zero por meio de exames anuais de rotina e comparações de exames anteriores, de modo que não haja risco de tumor, pois a lesão faz parte da anatomia mamária da paciente, quer seja de origem natural ou cirúrgica, um exemplo disso é a microcalcificação de uma paciente que realizou procedimentos cirúrgicos na mama. A recomendação nessa classificação é que a paciente continue realizando os exames de rotina (ACR, 2016).

Embora as lesões aparentem ser benignas, há até 2% de chances delas tornarem-se tumor, sendo descritas na categoria 3 do sistema BI-RADS. Em razão disso, é aconselhado às pacientes que efetuem o acompanhamento semestral até completar dois ou três anos (conforme a gravidade da lesão), sob a pretensão de acompanhar a estabilidade ou desenvolvimento da lesão (GUIMARÃES; CHOJNIAK, 2014).

Após o acompanhamento, se as lesões permanecerem estáveis, passam a ser caracterizadas como um “achado benigno” e a sua classificação é reduzida para a categoria 2, retornando ao grupo de rastreamento anual para os exames de rotina. Para exemplificar a redução de categoria 3 para categoria 2, são os fibroadenomas que apresentam macrocalcificações em seu interior em mulheres com idade superior aos 40 anos. Entretanto, caso não haja calcificação ou demonstrem evolução, a sua classificação persiste como BIRADS 3 ou pode sofrer reclassificação para a categoria 4 se a lesão estiver mais agressiva; neste caso, os procedimentos cabíveis são tomados (ACR, 2016).

O BI-RADS tipo 4 refere-se sobre uma suspeita de malignidade e é a única categoria que possui subdivisão: 4A, 4B ou 4C. Esta diferença é devido ao fato de na classificação tipo 4 há um espectro muito amplo de possibilidade de ser uma neoplasia. A subcategoria 4A diz respeito à lesão com risco baixo de malignidade, cuja chance de ser câncer seja inferior a 3%, e exige controle semestral após realizada a biópsia ou citologia negativa. A subcategoria 4B indica risco intermediário, com chances variando de 10 a 50% de não ser uma lesão benigna, sendo necessária uma boa correlação anatomopatológica ou anatomorradiológica, para se ter certeza se a lesão é benigna ou maligna. A subcategoria 4C inclui achados de suspeita moderada a elevada, ou seja, lesões com chance de 50 a 95% de serem tumores. Essas lesões apresentam todas as características de uma neoplasia, nas quais o resultado maligno é o mais esperado (ACR, 2016).

As lesões classificadas como alta suspeita para a malignidade estão

localizadas na categoria BIRADS 5. Neste tipo, associam-se um risco superior a 95% de que a lesão encontrada seja um câncer. Esta categorização pode ser reservada para o grupo de lesões clássicas de tumores representadas nos diversos métodos, como massas espiculadas, calcificações pleomórficas ou calcificações ductais; neste grupo, uma lesão maligna somente pode ser descartada após avaliação cirúrgica da região (GUIMARÃES; CHOJNIAK, 2014).

Finalizando as categorias de BI-RADS, no BIRADS 6 estão classificadas as lesões que já tem diagnóstico histológico de câncer quando retornam para controle de quimioterapia ou radioterapia neoadjuvantes ou mesmo em casos para segunda opinião. Ou seja, no tipo 6, não há dúvidas quanto à lesão ser um blastoma, e as pacientes já estão sendo submetidas a tratamentos médicos, bem como estão sendo recebendo acompanhamento para verificação das respostas às medidas submetidas, cirúrgicas ou não. Esta categoria não é apropriada nos casos de seguimento após cirurgia conservadora da mama (GUIMARÃES; CHOJNIAK, 2014).

Devido ao fato de os raios-X emitidos pelo mamógrafo não penetrarem de maneira absoluta nas mamas densas, acrescido das limitações inerentes à mamografia, outros procedimentos se tornam indispensáveis para que interpretação dos profissionais de saúde sejam mais assertivas. Logo, para resolver tais questões e para que não seja exigida uma segunda leitura que eleva os custos, surgiram então os sistemas CADe e CADx (GUIMARÃES; CHOJNIAK, 2014).

Além disso, discorrem Nascimento, Silva e Maciel (2010, p. 92) quanto à limitação da interpretação subjetiva e à necessidade de outras leituras e processos de análise:

Sabe-se que a acurácia dos exames de imagens mamárias pode ser afetada por inúmeros fatores, como aspectos técnicos, diferenças relacionadas às características da população em estudo, idade do paciente, experiência do médico radiologista, utilização de técnica de dupla leitura ou de programa de computador (*computer-aided detection systems – CADs*), bem como a variabilidade nas interpretações do médico radiologista no uso do BI-RADS.

Na literatura, existem sistemas CADs para todos os tipos de exames e para várias partes do corpo, como crânio, tórax, abdômen, ossos, sistema vascular, entre outros. Porém, os principais objetos de pesquisa para o desenvolvimento de sistemas CADe e CADx têm sido nos âmbitos da tomografia computadorizada, mamografia e angiografia. Por isso, esta pesquisa vai tratar no quesito na mamografia, verificada sua proposta inicial (GUIMARÃES; CHOJNIAK, 2014).

2.4 SISTEMAS CADe e CADx

As siglas CADe (ou, simplesmente, CAD) e CADx são termos provenientes da língua inglesa e significam, respectivamente, *Computer Aided Detection* (Detecção Auxiliada por Computador, em português) e *Computer Aided Diagnosis* (traduzido, Diagnóstico Auxiliado por Computador) (GUTFILEN; PEREIRA, 2002). Diante da possibilidade de falha e das limitações dos aparelhos nos diagnósticos da atualidade, outras técnicas de análise auxiliadas por computador têm sido propostas com o objetivo de oferecer melhores parâmetros para a elaboração de um diagnóstico mais preciso, indicando áreas suspeitas, bem como anomalias. Essas técnicas têm sido desenvolvidas por diversos grupos de pesquisas, sendo integradas em sistemas CADs visando auxiliar na detecção precoce do câncer da mama (GUTFILEN; PEREIRA, 2002).

Tais sistemas foram desenvolvidos devido à subjetividade no processo de avaliação de imagens pelo radiologista, estando sujeita a variações intra e interpessoais, bem como perda de informação devido à baixa qualidade da imagem, sobreposição de estruturas, fadiga visual ou distração. Calas, Gutfilen e Pereira (2002, p. 47) discorrem sobre às falhas humanas nos diagnósticos como:

Há evidências de que muitos dos cânceres detectados nas mamografias de rastreamento são visíveis retrospectivamente, entretanto não identificados pelo radiologista no momento da análise. Cerca de 10% a 30% das lesões mamárias são perdidas nos exames de rotina, devido a limitações próprias dos observadores humanos.

Com o intuito de minimizar os casos de falsos-negativos e de falsos-positivos foram desenvolvidos os sistemas CADs. Tais sistemas têm proporcionado um importante auxílio ao diagnóstico médico em diversas aplicações radiológicas, sobretudo na mamografia (GUTFILEN; PEREIRA, 2002).

O CADe identifica regiões suspeitas em uma imagem e ajuda os radiologistas a encontrar o câncer na mamografia de rastreio, fazendo o uso de técnicas de visão computacional. Nas palavras de Calas, Gutfilen e Pereira (2002, p. 47): “o objetivo do CADe não é de diagnóstico, todavia sim de alertar o radiologista para áreas específicas, em que a análise do radiologista da área demarcada irá decidir a necessidade de estudos adicionais”.

E, o Diagnóstico Auxiliado por Computador (*Computer Aided Diagnosis* -CADx) classifica as regiões suspeitas, permitindo que os radiologistas decidam se a lesão é

benigna ou maligna. O uso de técnicas de Inteligência Artificial tornam-se ferramentas importantes no auxílio do diagnóstico por especialistas.

Já foi demonstrado que uma dupla leitura, ou seja, a verificação de uma mesma imagem por dois profissionais diferentes, pode aumentar consideravelmente a sensibilidade do diagnóstico. Nesta perspectiva, os sistemas de auxílio podem atuar como um segundo especialista ou uma segunda opinião para a tomada de decisão, melhorando a consistência da interpretação da imagem radiológica (RIBEIRO, 2017). Ainda sobre a dupla leitura,

A dupla leitura da mamografia mostrou-se um benefício significativo, reduzindo o número de casos falso-negativos em 5% a 15%, melhorando as taxas de detecção do câncer de mama. Apesar dos seus benefícios de diagnóstico já comprovados, a utilização da dupla leitura nem sempre é possível, por causa de problemas logísticos e financeiros de cada instituição (CALAS; GUTFILEN; PEREIRA, 2012, p. 46-47)

De maneira geral, o CADe divide-se em quatro etapas: aquisição de imagens, melhoramento, segmentação e redução de falsos-positivos. Enquanto o CADx é voltado exclusivamente para o diagnóstico, logo, sua função está relacionada à diferenciação entre padrões de lesões, permitindo reconhecimento e identificação entre malignas e benignas (RIBEIRO, 2017).

Nesta alçada, os CADe e CADx se complementam, uma vez que o procedimento comum do especialista é detectar a região suspeita, validar se a região está clinicamente comprometida e se esta apresenta algum comportamento maligno ou benigno. Independente do emprego, é consenso na literatura o alerta para a não utilização de sistemas CAD como único meio de detecção e diagnóstico. Isto decorre porque o seu objetivo principal é contribuir para superar limitações humanas na realização de tarefas repetitivas, comuns na análise de grande quantidade de exames (NISHIKAWA, 2010).

Além do CADe e do CADx, o crescente avanço do processamento de imagem digital (através de novos hardwares ou softwares), o reconhecimento de padrões e uso de inteligência artificial proporcionam um cenário favorável para desenvolvimento de mecanismos computacionais que, no uso destas tecnologias, proporcionem aos especialistas uma forma de auxílio (NISHIKAWA, 2010).

Para apresentar bons resultados e boa performance, é necessário que os CADs mamográficos sejam avaliados com diferentes bancos de imagens. Essa necessidade é justificada, porque a aplicação computacional poder estar super ajustada para uma

determinada base de imagens. Existem algumas bases de mamogramas disponíveis na internet, entretanto nesta pesquisa utilizaremos uma base de domínio público o *Digital Database for Screening Mammogram* (DDSM).

2.5 PROCESSAMENTO DIGITAL DE IMAGEM

O presente trabalho utilizou como base imagens extraídas do *Digital Database for Screening Mammography* (DDSM), em tradução livre “Banco de Dados Digital para Rastreamento de Mamografia”, disponibilizado pela Universidade do Sul da Florida, nos Estados Unidos, sendo este um recurso para uso pela comunidade de pesquisa de análise de imagens mamográficas. O DDSM é organizado conforme "casos" e "volumes". Todos os volumes e casos estão disponíveis no próprio site da Universidade². De modo sucinto, pode-se aferir que o processamento digital de imagens é definido como sendo o conjunto de técnicas computacionais que manipulam uma imagem de entrada em uma saída, sendo que, na maioria dos casos, a saída é outra imagem digital. Dessa maneira, é possível, além de tratar aspectos visuais e estruturais, fornecer outros elementos que possibilitem a interpretação visual e computacional da imagem (PEDRINI et al, 2008).

2.5.1 AQUISIÇÃO DE IMAGEM

A aquisição de imagens é a etapa que caracteriza o início do processamento digital de imagens. Neste momento, a imagem pode ser obtida diretamente de um equipamento de raios-X digital, ou por meio de um digitalizador. O digitalizador é a aparelhagem que tem objetivo de receber os sinais analógicos e convertê-los em digitais (GONZALES et al, 2010). O DDSM disponibiliza para os usuários informações do exame e do próprio paciente e disponibiliza imagens obtidas no formato *Screen Film Mammography* (SFM). O processo de aquisição das imagens totalmente digitais, em formatos de qualidade, como *Full Field Digital Mammography* (FFDM), incide em mapear intensidades detectadas por células fotossensíveis em impulsos elétricos sem qualquer passo analógico, assim reduzindo drasticamente os ruídos gerados na

² Link de acesso: <http://www.eng.usf.edu/cvprg/Mammography/Database.html>

digitalização de dados originados nestes mamogramas, emitidos pelos raios-x (SILVA NETO, 2017).

Cada caso mamográfico está disposto pela DDSM em imagens obtidas nas duas mamas de cada paciente, nas posições Médio-Lateral Oblíqua e Crânio-Caudal. Tais imagens são disponibilizadas no padrão LJPEG e representadas em níveis de cinza com resolução de 12 ou 16 bits por pixels. Os laudos gerados por especialistas estão inseridos às imagens contendo informações como o código de cada paciente, a idade da mesma na época do estudo, a classificação na categoria BI-RADS (*Breast Imaging Report and Data System*) e na composição da densidade mamária segundo o mesmo padrão, além de estarem acompanhados de tipo de anormalidade e a localização da lesão (BOWYER, 1996).

2.5.2 PRÉ-PROCESSAMENTO

Em seguida, se faz necessária a remoção das estruturas não-desejadas, “limpando” a imagem dos ruídos, bordas, marcações, músculos desinteressantes, e outras informações que divergem do estudo pretendido, mas que estão visíveis na imagem e podem atrapalhar a clareza e assertividade do diagnóstico (PAVAN, 2017). Tal procedimento precede a segmentação das imagens. Após a remoção destas estruturas, é aplicada a técnica de realce baseada no histograma e o filtro da média, com finalidade de destacar as estruturas internas das mamografias, ou seja, das regiões que o estudo deve ser focado (SILVA NETO, 2017).

2.5.3 SEGMENTAÇÃO

Após o pré-processamento, ocorre a segmentação, que consiste na separação das regiões mais homogêneas e similares na imagem, concluindo nos objetos considerados suspeitos, tais como nódulos, massas estranhas, microcalcificações, assimetrias, ou qualquer outra anomalia que foge da anatomia da região das mamas em questão (SILVA NETO, 2016). Neste projeto, é proposta a abordagem de segmentação dividida em três etapas, na primeira utilizamos o método OTSU, a segunda o método PSO e a terceira é composta pelo *Graph Clustering*.

2.5.4 ALGORITMO OTSU

O algoritmo de OTSU reconhece que a imagem está dividida em duas classes, objeto e fundo, e busca localizar especificamente o limiar que melhor divide essas anomalias por meio de múltiplas iterações, calculando a diferenciação entre as classes de objeto e fundo para todos e quaisquer possíveis limiares (OTSU, 1979). Quando definido, todos os pixels da imagem que estiverem destacadas, ora, cujos valores de nível acinzentado são menores ao limiar definido, esses serão selecionados para o primeiro cluster e, assim, os valores cujo nível de cinza forem maior ou igual ao estipulado em limiar, são selecionados como pertencentes ao segundo cluster, permitindo melhor diferenciação e identificação da região (SILVA NETO, 2016).

2.5.5 PARTICLE SWARM OPTIMIZATION (PSO)

O PSO é uma técnica de computação evolucionária desenvolvida por James Kennedy, um psicólogo social, e por Russell Eberhart, um engenheiro elétrico, em 1995, inspirada na simulação de um sistema social simplificado. A intenção original era simular graficamente o comportamento de um bando de pássaros em voo com seu movimento localmente aleatório, mas globalmente determinado (VAN et all, 2003).

Silva Neto (2017), esclarece que dado uma divergência, o PSO mantém uma população de partículas em que cada partícula representa uma solução para o problema e está ligada a uma posição numa região de busca multidimensional. O PSO consiste em repetidas iterações das equações anteriores e toda vez que uma partícula for melhorada através do cálculo de fitness, a mesma é atualizada. Nesta apresentação, o PSO será um dos instrumentos utilizados para encontrar os melhores limiares, para a devida geração de locais mais homogêneos em grupos (SILVA NETO, 2017).

2.5.6 GRAPH CLUSTERING

O Graph Clustering é um método de agrupamento dos vértices do gráfico em clusters considerando a estrutura das arestas das imagens. Neste trabalho, definimos

alguns parâmetros: a região será definida em 3x3, o gráfico será formado a partir das regiões de interesse (ROI) de uma imagem, verificando sua vizinhança com todas as ROIs encontrados no cluster original, assim, será direcionado o grafo. Ao concluir o processo, os nós do grafo que possuírem mais de duas ligações serão excluídos e os nós que tenham nenhuma ou no máximo duas ligações permanecerão. Cada nó do gráfico resposta desse processo, simula uma ROI (SCHAEFFER, 2007).

Em seguida, (SCHAEFFER, 2007) esclarece os cálculos do *graph cluster* da seguinte forma: A partir de um nó qualquer do grafo, é calculado seu valor de Fator de Forma Circular (FFC) (Equação FFC), da seguinte forma: a) se FFC for menor que 10%: o nó será descartado e será escolhido outro nó, e o processo se repete; b) se FFC for maior que 10%: verifica-se suas adjacências, realizando as uniões do nós; c) após cada união, o FFC é calculado novamente, caso este, seja maior que 10% a união é válida e será verificado (caso exista) a próxima adjacência; d) se o FFC, após cada união, resultar em um valor menor que 10%, esta união não será válida, e o nó que foi unido será descartado.

Na equação FFC, que menciona Ribeiro (2017), **A** corresponde à área e **P** ao perímetro de cada ROI e o cálculo se apresenta da seguinte forma: $FFC = \frac{4 \cdot \pi \cdot A}{p^2}$ O percentual da FFC foi adotado em 10%, em decorrência dos testes, sendo este o valor que apresentou melhores resultados.

2.5.7 EXTRAÇÃO DE CARACTERISTAS

Dentre as tarefas relacionadas à análise de imagens, uma das mais complexas está a extração de características de cada região contida na imagem estudada, pelo meio a serem utilizados em processos de mais alto nível. Em outras palavras, a etapa de caracterização e representação da imagem consiste em uma etapa de fundamental importância no modelo clássico do processamento de imagem digital. Através de modelos/métodos matemáticos, as propriedades da imagem são representadas em formato de atributos, para então, servirem de entrada para uma posterior análise de reconhecimento e classificação de padrões (SILVA NETO, 2016).

Para analisar os padrões de textura das ROIs, a matriz de coocorrência foi computada verificando os pixels com distância 1 e foram utilizadas nove descritores: entropia, energia, contraste, homogeneidade, momento diferença inverso, momento diferença de ordem, probabilidade máxima, variância inversa e correlação (SILVA

NETO, 2016). Em resumo, a etapa de extração de características deve ser dividida em duas categorias para análises: textura e forma. Em uma análise por textura, a finalidade é descrever aspectos da imagem no que diz respeito a suavidade, rugosidade e regularidade. Já em uma análise baseada na forma da imagem, a ideia é extrair informações que calculam sobre propriedades morfológicas do gráfico (RIBEIRO, 2017).

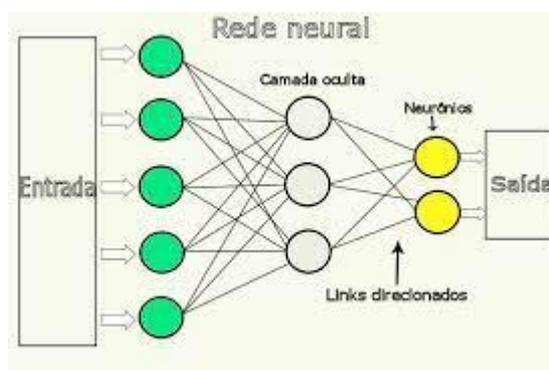
2.5.8 RECONHECIMENTOS DE PADRÕES

Conforme Silva Neto (2016), o uso do termo “reconhecimento” na computação significa que um computador reconhece que os padrões de objetos são aqueles que foram processados anteriormente. E em relação ao padrão de um objeto são dados espaciais e cronológicos recebidos do objeto por um dispositivo de entrada e levado ao sistema de reconhecimento. Nesta linha, reconhecimento de padrões é a disciplina científica que lida com métodos de descrição e classificação de objetos. Além disso, pode ser definido como a classificação de dados baseados em um conhecimento já obtido ou em informações estatísticas extraídas de padrões e/ou suas representações (SILVA NETO, 2016). Nesta seção discute-se reconhecimento de padrões, uma subárea da aprendizagem de máquina, que visa classificar informações, ou padrões, baseado em um conhecimento prévio ou em informações estatísticas extraídas dos padrões encontrados (RIBEIRO, 2017). Em seguida é apresentada as métricas utilizadas para validação da metodologia deste trabalho.

2.6 REDES NEURAIAS

As redes neurais artificiais, ou simplesmente redes neurais, são estruturas computacionais que tentam emular o comportamento do cérebro humano. Para tal, essas redes são formadas pela interligação de estruturas computacionais chamadas neurônios. Esses neurônios são processadores de dados que tentam se comportar como neurônios naturais, recebendo estímulos nas entradas, processando esses estímulos e através de uma função de ativação exibem uma saída, que para o caso mais simples (o *perceptron* de camada única) é a simples ativação, ou não, da saída, em valor alto ou baixo dependendo do valor das entradas e de um limiar (HUYNH; LI; GIGER, 2016). Como ilustrada na Figura 3.

Figura 3: Estrutura de uma rede neural



Fonte: (HUYNH 2016)

Existem basicamente dois tipos de aprendizado para redes neurais: supervisionado e não-supervisionado. Para o aprendizado supervisionado, temos um conjunto de treinamento com pares formados por informações de entrada e sua respectiva informação de saída (ou padrão desejado) que já é conhecida, com isso, depois de treinada a rede, outras entradas com saídas não conhecidas são apresentadas à mesma e ela é capaz de oferecer respostas a esses estímulos. Para as redes neurais de aprendizado não-supervisionado, não existe padrão desejado para o conjunto de treinamento (SILVA NETO, 2016). Isso quer dizer que a rede deve ser capaz de se auto-organizar com o propósito de agrupar os dados de entrada de forma totalmente inerente à própria arquitetura interna da rede. Para isso são usados métodos de treinamento competitivo entre os neurônios, onde essas redes ou mapas de neurônios vão recebendo os dados de entrada e o neurônio vencedor (e por vezes sua vizinhança) vão atualizando seus pesos (RIBEIRO, 2017).

Segundo Huynh, Li e Giger (2016), o modelo computacional de um neurônio consiste em um conjunto de conexões de entrada que alimentam a unidade com ativações vindas de outros neurônios. Essas entradas são ponderadas usando um conjunto de pesos denotados por $\{w\}$. A unidade de processamento soma todas as entradas e aplica uma função de limite não linear (também conhecida como função de ativação) para calcular a saída. A saída final é então transmitida a outras unidades neurais conectada (HUYNH; LI; GIGER, 2016).

2.6.1 TIPOS DE APRENDIZAGEM

Máquinas de aprendizado são sistemas artificiais com a capacidade de extrair informações e similaridades em conjuntos de dados, de forma a aprender padrões de informações relevantes dentro desses dados. Sistemas classificadores são tipos de máquinas de aprendizado que possuem a capacidade de inferir, a partir das características ou atributos de uma dada instância, a qual grupo ou classe a referida instância pertence, a partir de critérios internos à arquitetura de tal classificador (HUYNH; LI; GIGER, 2016). Segundo Kooi (et al, 2017) esses classificadores podem ser bio-inspirados, como no caso das redes neurais artificiais, onde normalmente são máquinas de aprendizado estatístico que realizam uma etapa de aprendizado (ensinamento) de maneira a alterarem suas características internas, moldando-se às propriedades relevantes de grupos de instâncias já classificadas. Depois de treinados, esses classificadores devem ser capazes de extrapolar o “conhecimento” adquirido e com isso, classificar com certa acurácia um indivíduo desconhecido, entretanto que pertença à uma das classes do conjunto de treinamento (HUYNH; LI; GIGER, 2016).

Soma-se a isso à necessidade de se estabelecer as redes neurais convolucionais (RNC).

2.6.2 REDES NEURAS CONVOLUCIONAL

Para Huynh, Li e Giger (2016), as redes neurais convolucionais (RNC) são análogas às redes neurais artificiais tradicionais, sua composição corresponde a um sistema de neurônios que se auto-otimizam por meio do aprendizado e do aprimoramento dos saberes. As redes neurais convolucionais são projetadas para processar dados que vêm na forma de várias matrizes, por exemplo, uma imagem colorida composta de três matrizes de duas dimensões contendo intensidades de pixel nos três canais de cores.

A arquitetura de uma RNC é composta por uma série de camadas que irão processar as imagens de entrada. Com relação às camadas, estão divididas em: convolução, *pooling* e camadas totalmente conectadas. A camada convolucional é o componente principal de uma RNC (KOOI, et al, 2017). Inclui um conjunto de filtros (também chamados de *kernels convolucionais*) que realizam a convolução com uma certa entrada e geram um conjunto de *feature maps* na saída. As unidades em uma camada convolucional são organizadas em *feature maps*, onde cada unidade é conectada a uma parte da camada anterior por meio de um conjunto de pesos

denominado de filtros. O resultado da soma dos pesos é então passado para uma *Leaky ReLU*, que é uma unidade linear retificadora que produz zero em metade de seu domínio (HUYNH; LI; GIGER, 2016).

Pooling é um conceito importante da RNC. Ele diminui a carga computacional, reduzindo o número de conexões entre as camadas convolucionais. A camada de *pooling* não afeta a dimensão da profundidade do volume. A camada de *pooling* opera sobre cada mapa de ativação na entrada e calcula sua dimensionalidade usando a função “MAX” Na maioria das RNCs, eles vêm como *max-pooling* com núcleos de uma dimensionalidade de 2x2 aplicados com um *stride* de 2 ao longo das dimensões espaciais da entrada. Isso reduz o mapa de ativação para 25% do tamanho original - enquanto mantém o volume de profundidade em seu tamanho padrão. Seguindo várias camadas convolucionais e de *pooling*, o raciocínio de alto nível na rede neural é realizado por meio de camadas totalmente conectadas (KOOI, et al, 2017).

2.7 MÉTRICAS DE AVALIAÇÃO

A heterogeneidade desta espécie de câncer pode ser observada pelas variadas manifestações histológicas e morfológicas, diferentes assinaturas genéticas e consequentes diferenças nas respostas terapêuticas.

Barboza (2014) destaca que uma gama de pesquisas relacionadas às neoplasias vêm sendo desenvolvidas pelos mais numerosos centros de pesquisas ao redor do globo, com a finalidade de melhorar e aumentar a expectativa de vida das mulheres que sofrem desta doença, o câncer de mama.

Através dos esforços da ciência e o apoio da tecnologia, novas ferramentas de auxílio ao diagnóstico estão sendo desenvolvidas para a prevenção e detecção precoce do câncer de mama. O aprimoramento das ferramentas de processamento digital em alta qualidade vem melhorando cada vez mais a acurácia diagnóstica de neoplasias mamárias, tornando-se ferramentas muito importantes na medicina moderna. As recorrentes e desenvolvimentos destas aplicações nestes métodos de diagnóstico proporcionam aos profissionais de mastologia informações importantíssimas, reduzindo exponencialmente a necessidade de intervenções cirúrgicas, aumentando os exames convencionais, como ultrassonografia e radiologia e proporcionando a segurança e aumento da sobrevida das pacientes acometidas por tal patologia (SILVA NETO, 2017).

O método padrão para identificação e rastreamento do câncer de mama é a

mamograma e, através dele, tem sido acompanhada a redução significativa na taxa de mortalidade ocasionada por essa neoplasia em todo o mundo. O rastreamento mamográfico permite a redução da mortalidade entre 20% a 30%, vez que sua intenção é antever qualquer sintoma ou anomalia mamária. Dessa maneira, permite que a paciente atue junto de seu profissional de saúde para tomar as devidas providências e trate a lesão antes de seu desenvolvimento.

No Brasil, o Ministério da Saúde aconselha que o exame ocorra com a frequência bianual nas mulheres com idade de 50 a 69 anos; para mulheres fora dessa faixa etária a mamografia não é recomendada, visto que, há mais riscos do que benefícios sobre a aplicação dessa técnica (INCA,2021). No entanto, mulheres com idade entre 35 e 49, que pertençam ao grupo de risco elevado para desenvolver câncer de mama, tem a recomendação de fazer o exame clínico em conjunto com a mamografia anualmente. O grupo de risco é definido como: mulheres com histórico familiar (mãe, irmã ou filha) de câncer de mama abaixo dos 50 anos, ou com parentes de primeiro grau (mãe, irmã ou filha) com câncer de mama bilateral ou câncer de ovário, em qualquer faixa etária, ou com histórico de câncer de mama masculino ou aquelas que apresentam diagnóstico histopatológico de lesão mamária proliferativa com atipia ou neoplasia lobular *in situ* (INCA, 2015).

Apesar da redução constante, a mamografia ainda possui limitações em seu procedimento, principalmente em mulheres com mama densa, de alto risco para o câncer de mama. Isto ocorre devido a sensibilidade do equipamento é de 85%, podendo chegar até a 15% de falso negativo, sua especificidade de 93% com falso positivo de 7%, e sua acurácia é de 89,6% para esta característica de mamas.

Portanto, sendo verificada tal possibilidade de falhar, a mamografia necessita de outros métodos complementares da imagem, com objetivo de melhorar a assertividade no que se referem às imagens. Geralmente, a leitura e interpretação de imagens da mamografia são procedimentos que exigem experiência e vivência do radiologista, e sua conclusão de diagnóstico depende, sobretudo, de treinamento, experiência e demais critérios subjetivos que são inerentes apenas ao próprio. Até experts apresentam uma taxa de acertos de classificação com variação de 65% a 75% entre observadores, tendo em vista que são limitações humanas (RIBEIRO, 2017).

A presença de falsos positivos e negativos na mamografia, apresentam repercussões físicas e emocionais para as pacientes. Tais variações tem ocasionado e intensificado a busca por recursos de imagens complementares à mamografia.

Nesta alçada, técnicas computacionais determinaram o surgimento de inúmeras pesquisas ao longo dos anos, com a intensão de desenvolver, aprimorar e criar sistemas computacionais. Esses sistemas estão presentes em diversos centros de diagnósticos e são conhecidos como Sistemas CAD. Eles estão ajudando os especialistas a cumprirem seu papel de interpretar e diagnosticar através de imagens radiológicas, conseguindo aumentar as taxas de acerto na identificação antecipada de doenças, como o carcinoma mamário (RIBEIRO, 2017).

Estes sistemas podem auxiliar os radiologistas na difícil tarefa de leitura e interpretação de imagens da mamografia, automatizando o processo de interpretação de imagens médicas. De 65% a 90% das biopsias de tecidos com suspeita de câncer mostram-se benignas, com traumas físicos e emocionais para as pacientes, o que mostra a importância do desenvolvimento de sistemas de detecção e diagnóstico mais precisos como os CADs, que permitam aprimorar a distinção entre lesões benignas e malignas (SILVA NETO, 2016).

Os sistemas CADs oferecem opiniões extraídas de imagens médicas, proveniente de várias modalidades e tipos, bem como: ressonânciamagnética, ultra sonografia, mamografia, entre outras. São utilizadas técnicas para o tratamento e processamento das imagens como, por exemplo: utilização de redes neurais artificiais, aprendizado de máquina, reconhecimento de padrões entre outros aspectos computacionais, que são aplicadas com o objetivo de melhorar tais imagens para extrair informações úteis ao diagnóstico (RIBEIRO, 2017). A propósito, o presente projeto de pesquisa desenvolveu um método de redes neurais para classificar nódulos mamários no padrão BI-RADS® nas categorias 4 e 5.

2.8 JUSTIFICATIVA

Levando em consideração o alto índice de morbidade e mortalidade do câncer de mama, observa-se a necessidade de melhorar a assertividade da leitura e interpretação dos exames mamográficos, em especial através dos sistemas (CADs) de diagnóstico de imagem auxiliada por computadores, que conforme o avanço tecnológico na área científica, vem recebendo cada vez mais importância quanto à acurácia e assertividade nos processos de diagnósticos das lesões na mama. Logo, nos dias de hoje, até onde conhece a área da biotecnologia, foram encontrados alguns métodos que utilizam as redes neurais, mas nenhuma focados somente nas lesões

malignas das categorias 4 e 5 do padrão BI-RADS®, o que nos levou à compreensão e entendimento da importância desta apresentação, tal como motivou a realização deste.

REFERÊNCIAS

AMERICAN COLLEGE OF RADIOLOGY (ACR) (2016). **Atlas BI-RADS do ACR:** sistema de laudos e registro de dados de imagem da mama / American College of Radiology; [tradução Angela Caracik] - 2. ed. - São Paulo: Colégio Brasileiro de Radiologia.

Barboza, S. H. I. Tomógrafo em nível de simulação utilizando micro-ondas em banda ultra larga (UWB) com transmissor em tecnologia CMOS para detecção precoce de câncer de mama (Doctoral dissertation, Universidade de São Paulo).

BRASIL. Ministério da Saúde (2010). Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Rastreamento. Ministério da Saúde, Secretaria de Atenção à Saúde, Departamento de Atenção Básica. – Brasília: Ministério da Saúde.

BOWYER, K; K., Kopans; Kegelmeyer, P. (1996) The digital database for screening mammography. In: Third international workshop on digital mammography. (pp. 457-460). Springer, Dordrecht.

Calas, M. J. G., Gutfilen, B., & Pereira, W. C. D. A. (2012). CAD e mamografia: por que usar esta ferramenta?. *Radiologia Brasileira*, 45, 46-52.

Carvalho Filho, A. O. D. (2016). Métodos para sistemas CAD e CADx de nódulo pulmonar baseada em tomografia computadorizada usando análise de forma e textura.

Duarte, R. D., Furtado, A. A., Lermen Jr, A., Borges, L., Carvalho, Ê. M., Neves, H. Z., ... & Duarte, D. L. (2005). Lesões mamárias incomuns: ensaio iconográfico. *Radiologia Brasileira*, 38, 371-376.

Gerótica, R. M. G., Azevedo, R. L., & Sanches, T. P. (2016). A importância da mamografia no diagnóstico precoce do câncer de mama. *UNILUS Ensino e Pesquisa*, 13(30), 251.

Global Cancer Observatory. *Cancer Today*. (2020). Disponível em: <<https://gco.iarc.fr/>>. Acesso em: 27 jul. 2021.

Gonzalez, R. C. (2010). *Digital image processing*. Pearson education india.

Guimaraes, M. D., Schuch, A., Hochegger, B., Gross, J. L., Chojniak, R., & Marchiori, E. (2014). Ressonância magnética funcional na oncologia: estado da arte. *Radiologia Brasileira*, 47, 101-111.

Huynh, B. Q., Li, H., & Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3), 034501.

Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) (2019). A situação do câncer de mama no Brasil: síntese de dados dos sistemas de informação. INCA.

_____. Controle do Câncer de Mama (2021): Conceito e Magnitude. Ministério da Saúde. Instituto Nacional de Câncer. – Rio de Janeiro: INCA. 11 jun.. Disponível em: <<https://www.inca.gov.br/controle-do-cancer-de-mama/conceito-e-magnitude>>. Acesso em: 31 jul. 2021.

_____. Mamografia (2007): da prática ao controle. Ministério da Saúde. Instituto Nacional de Câncer. – Rio de Janeiro: INCA.

Instituto Oncoguia (ONCOGUIA) (2020). Tipos de Câncer de Mama. Disponível em: <<http://www.oncoguia.org.br/conteudo/tipos-de-cancer-de-mama/1382/34/>>. Acesso em: 28 jul. 2021.

Kooi, T., van Ginneken, B., Karssemeijer, N., & den Heeten, A. (2017). Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Medical physics*, 44(3), 1017-1027.

MOORE, Keith L., Arthur F. Dalley, Anne M. R. Agur (2019) Anatomia orientada para a clínica, tradução Cláudia Lúcia Caetano de Araújo. - 8. ed. - Rio de Janeiro: Guanabara Koogan.

Nascimento, J. H. R. D., Silva, V. D. D., & Maciel, A. C. (2010). Acurácia dos achados mamográficos do câncer de mama: correlação da classificação BI-RADS e achados histológicos. *Radiologia Brasileira*, 43, 91-96.

Nishikawa, R. M. (2010). Computer-aided detection and diagnosis. In *Digital Mammography* (pp. 85-106). Springer, Berlin, Heidelberg.

Pavan, A. L. M., Vacavant, A., Trindade, A. P., & de Pina, D. R. (2017). Fibroglandular tissue quantification in mammography by optimized fuzzy C-means with variable compactness. *Irbm*, 38(4), 228-233.

Pedrini, H., & Schwartz, W. R. (2008). Análise de imagens digitais: princípios, algoritmos e aplicações. Cengage Learning.

Pisano, E. D., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J. K., Acharyya, S., ... & Rebner, M. (2005). Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353(17), 1773-1783.

Sampaio, W. B., Diniz, E. M., Silva, A. C., De Paiva, A. C., & Gattass, M. (2011). Detection of masses in mammogram images using CNN, geostatistic functions and SVM. *Computers in biology and medicine*, 41(8), 653-664.

Silva Neto, O. P. D. (2016). Detecção automática de massas em imagens mamográficas usando particle swarm optimization (PSO) e índice de diversidade funcional.

Neto, O. P. S., Silva, A. C., Paiva, A. C., & Gattass, M. (2017). Automatic mass detection in mammography images using particle swarm optimization and functional diversity indexes. *Multimedia Tools and Applications*, 76(18), 19263-19289.

Oliveira, M. E. C. D. (2019). *Análise do atraso da paciente no diagnóstico do câncer de mama–Paraíba*.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.

RIBEIRO, F. A. A. (2018). *OTIMIZAÇÃO DE IMAGENS MAMOGRÁFICAS PARA AUXÍLIO AO DIAGNÓSTICO MÉDICO* (Doctoral dissertation).

Santos, T. A. D., & Gonzaga, M. F. N. (2018). Fisiopatologia do câncer de mama e os fatores relacionados. *Rev Saúde Foco*, 10, 359-366.

Schaeffer, S. E. (2007). Graph clustering. *Computer science review*, 1(1), 27-64.

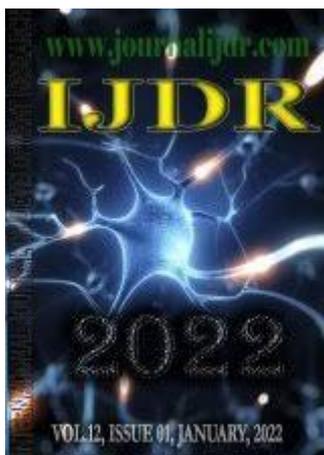
Tosteson, A. N., Stout, N. K., Fryback, D. G., Acharyya, S., Herman, B. A., Hannah, L. G., ... & DMIST Investigators*. (2008). Cost-effectiveness of digital mammography breast cancer screening. *Annals of internal medicine*, 148(1), 1-10.

Van der Merwe, D. W., & Engelbrecht, A. P. (2003, December). Data clustering using particle swarm optimization. In *The 2003 Congress on Evolutionary Computation, 2003. CEC'03*. (Vol. 1, pp. 215-220). IEEE.

3 CAPÍTULOS

3.1 Capítulo 1

Contribuição de Redes Neurais em Sistemas Computacionais para o Diagnóstico do Câncer de Mama



Artigo publicado,
Volume: 12 Article ID: 24331 4 pages
International Journal of Development Research

ISSN: 2230-9926

QUALIS: B1

Contribuição de Redes Neurais em Sistemas Computacionais para o Diagnóstico do Câncer de Mama

Elmo de Jesus Nery Júnior^{1,*}, Francisco Adelson Alves Ribeiro², Otilio Paulo da Silva Neto³, Larysse Maira Cardoso Campos Verdes⁴, Renato de Oliveira Pereira⁵, Danylo Rafael Costa Silva⁶, Maria da Conceição Barros Oliveira⁷, Josemeire da Costa Ximenes⁸, Victor Alves de Oliveira⁹ Isabel Cristina de Almeida Brito Nery¹⁰, Daiana Sousa da Silva Araújo¹¹, Alysson Ramirez de Freitas Santos¹², and Francisco das Chagas Alves Lima¹³

¹ Doutorando em Biotecnologia/Renorbio pela Universidade Federal do Piauí; ² Doutor em Biotecnologia/Renorbio, Docente do Instituto Federal do Maranhão, ³ Doutor em Engenharia Elétrica, Docente do Instituto Federal do Piauí, ⁴ Doutoranda em Biotecnologia/Renorbio pela Universidade Federal do Piauí, ⁵ Doutorando em Biotecnologia/Renorbio pela Universidade Federal do Piauí, ⁶ Doutor em Biotecnologia/Renorbio pela Universidade Federal do Piauí, ⁷ Doutor em Biotecnologia/Renorbio pela Universidade Federal do Piauí, ⁸ Mestre em Biodiversidade Ambiente e Saúde pela Universidade Estadual do Maranhão, ⁹ Doutorando em Alimentos e Nutrição pela Universidade Federal do Piauí, ¹⁰ Graduada em Pedagogia pelo Centro Universitário Unifacema, ¹¹ Mestre em Biodiversidade Ambiente e Saúde pela Universidade Estadual do Maranhão, ¹² Mestre em Engenharia de Produção pela Universidade Paulista, ¹³ Doutor em Química, Docente da Universidade Estadual do Piauí

Programa de Pós-Graduação em Biotecnologia, Universidade Federal do Piauí, Brasil;

Departamento de Química, Laboratório de Química Quântica Computacional, Universidade Estadual do Piauí, Teresina, Piauí, Brasil.

Resumo

Introdução: Os sistemas computacionais vêm se destacando progressivamente em diferentes áreas, especificamente na área da saúde, estes são utilizados na investigação de inúmeras doenças crônicas, em particular, o câncer de mama.

Método: Realizou-se busca de anterioridade dos registros de programa de computador que faz uso de redes neurais artificiais aplicadas ao diagnóstico de câncer de mama, no período de 2010 a 2020, realizada em três bases de depósitos de patentes: *European Patent Office (EPO)*, *World Intellectual Property Organization (WIPO)* e *United States Patent and Trademark Office's (USPTO)*. Utilizando as seguintes combinações de palavras-chave: software AND image processing, software AND breast câncer, software AND BI-RADS, software AND neural networks, software AND image processing AND breast câncer AND BI-RADS AND neural networks. **Resultado:** Encontrou-se um total de 24 patentes, sendo 22 na WIPO e 2 na USPTO, relacionadas a imagens de câncer mamário e sistema BI-RADS. Ressalta-se que a busca feita na base EPO não retornou registros. **Conclusão:** As análises demonstram que a área de redes neurais está em crescente expansão, abrindo oportunidades para avanço de projeto PD&I (pesquisa, desenvolvimento e inovação), na implantação e desenvolvimento de novas tecnologias/software voltadas ao melhoramento do diagnóstico das doenças em particular o câncer de mama.

Palavra Chave: Software, Câncer de mama, Redes Neurais, BI-RADS.

Introdução

O câncer é um problema de saúde pública mundial, estando entre as principais causas de morbidade e mortalidade do mundo, correspondendo a 9,9 milhões de mortes em 2020 (WHO, 2020). No Brasil, segundo dados do Instituto Nacional do Câncer (INCA), foi estimado para biênio de 2020-2021 o diagnóstico de 66.280 novos casos de câncer de mama que equivale a uma taxa 29,7% dos casos na população feminina e cerca de 15.000 mortes pela doença, com um risco estimado de 43,74 para cada 100 mil mulheres (INCA, 2019).

O Câncer de Mama é a neoplasia que mais acomete mulheres em todo o mundo, sua incidência é maior nos países desenvolvidos, entretanto a mortalidade é maior nos países menos desenvolvidos (RIBEIRO, et al 2017). Estimativa mundial mostra que em 2018 haveria aproximadamente 209 milhões de novos casos e cerca de 630 mil mortes por câncer de mama, o que representa de 24,2% e 15% de todos os novos casos de câncer e morte, respectivamente

(WANG, et al 2020). Há métodos utilizados para reduzir este quantitativo de mortes, como por exemplo, exames radiológicos que são utilizados para esta finalidade, destaca-se a mamografia que é o exame padrão para o rastreamento e diagnóstico do câncer de mama. Silva (2020) afirma que o exame monográfico consiste em um diagnóstico por imagem que tem por finalidade o estudo do tecido mamário por meio do uso de radiação ionizante.

O rastreamento por meio da mamografia visa além da detecção precoce, a redução da mortalidade de mulheres acometidas pelo câncer de mama. O controle da doença se dá por meio do diagnóstico prévio e os meios mais eficazes para a detecção desse tipo de câncer são os exames clínico de mama e a mamografia (PRADO, 2020) (GIESS, 2012). Procedimentos foram desenvolvidos e estão sendo aprimorados para a correta prevenção e detecção deste tipo de câncer, sendo a mamografia digital um dos métodos mais eficazes (TORTORA, 2016) (BALAS, et all 2020). Ressalta-se que a leitura e interpretação das imagens mamográficas são tarefas que exigem experiência do radiologista, mesmo experts bem treinados apresentam uma taxa de acertos de classificação com variação de 65 a 75% entre observadores (JUSTO, et al 2013).

No entanto, ferramentas computacionais podem auxiliar os profissionais médicos na interpretação e no diagnóstico dessas imagens, como por exemplo, os sistemas *Computer Aided Design* – CADs, os quais aumentam a sensibilidade, especificidade e acurácia do câncer da mama. Tais ferramentas aplicadas em conjunto com as técnicas de redes neurais para o processamento de imagens mamográficas, podem melhorar e extrair informações úteis ao diagnóstico preciso do câncer de mama melhorando o índice de acertos médicos (EL ATLAS et al., 2014) (GANESAN et al., 2012).

A prospecção tecnológica revelou uma crescente no desenvolvimento de novas ferramentas computacionais para auxílio ao diagnóstico do câncer de mama, por isso este artigo tem como objetivo realizar uma pesquisa de anterioridade no intuito de verificar a quantidade de patentes/registros de software que utilizam as redes neurais com essa finalidade.

Metodologia

Para o desenvolvimento da prospecção tecnológica foi utilizado a busca de publicações contidas no portal de periódicos CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, através do endereço eletrônico: <http://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br>, (CAPES 2020).

As buscas foram realizadas em três bases de depósitos de patentes: Instituto *European Patent Office* (EPO), *World Intellectual Property Organization* (WIPO) e *United States Patent and Trademark Office's* (USPTO).

Um estudo quantitativo foi realizado baseado na análise comparativa de resultados, buscados no período de 01/01/2010 à 15/12/2020 utilizando as seguintes descritores com critério de inclusão: Software, Image processing, breast câncer, BI-RADS, neural networks, os termos em inglês foram utilizados para as bases escolhidas. Aplicou-se como critério de exclusão a remoção da duplicidade e também os artigos que em sua pesquisa utilizavam como descritores: ultrassonografia e ressonância magnética.

Resultados

Utilizou-se os conectivos “Software AND Image processing” obteve-se o resultado somado a todas as bases de patentes pesquisadas, um total de > 200.000, “Software AND breast cancer” retornou > 80.000, “Software AND BI-RADS” retornou 310 patentes, “Software AND neural networks” retornou > 60.000, “Software AND Image processing AND breast cancer AND neural networks” retornou 585 patentes, “Software AND Image processing AND breast cancer AND BI-RADS AND neural networks” a base de dados do *European Patent Office* (EPO) retornou 0 publicações, o *World Intellectual Property Organization* (WIPO) retornou 22 publicações, já o *United States Patent and Trademark Office's* (USPTO), retornou 2 publicações, totalizando 24 publicações.

Após remoção das duplicidades encontradas nas bases de dados foram catalogadas 212 publicações. Como critérios de exclusão utilizou-se como descritores a ultrassonografia e a ressonância magnética. Após leitura dos títulos e resumos das publicações e aplicação dos critérios de exclusão, 63 publicações foram selecionadas para leitura completa. Aplicando os critérios de exclusão nas

publicações selecionadas para leitura completa, obteve-se como resultado 24 publicação.

Discussão

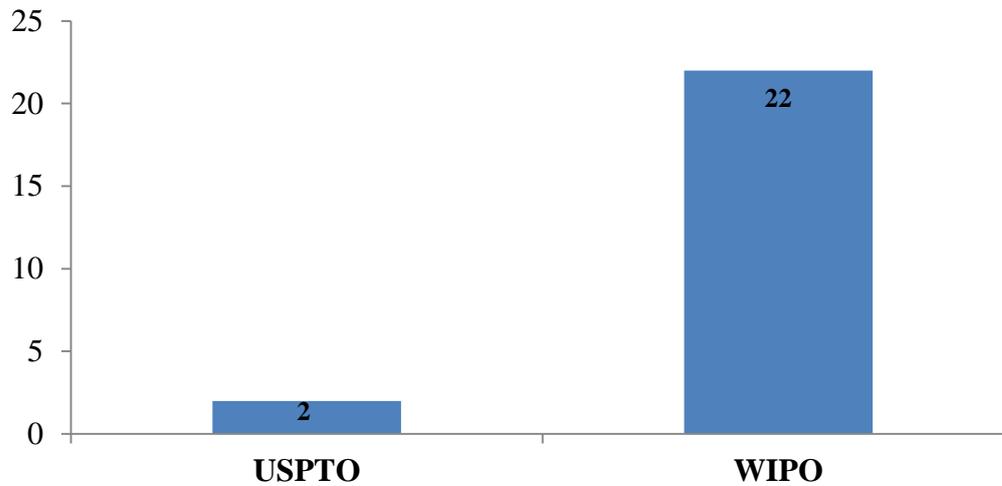
O Quadro 1 mostra o resultado das patentes encontradas por meio da busca utilizando as palavras-chave na combinação informada. Para isso observa-se que quando mais refinado a pesquisa com todas as palavras-chave o resultado teve uma proporção menor do que as demais com 24 patentes.

Quadro 1. Quantidade de registros de patentes obtidos por palavras-chave.

Palavras-Chave	Número de Patentes encontradas
Software AND Image processing	> 200.000
Software AND breast câncer	> 80.000
Software AND BI-RADS	310
Software AND neural networks	> 60.000
Software AND Image processing AND breast cancer AND neural networks	585
Software AND Image processing AND breast cancer AND BI-RADS AND neural networks	24

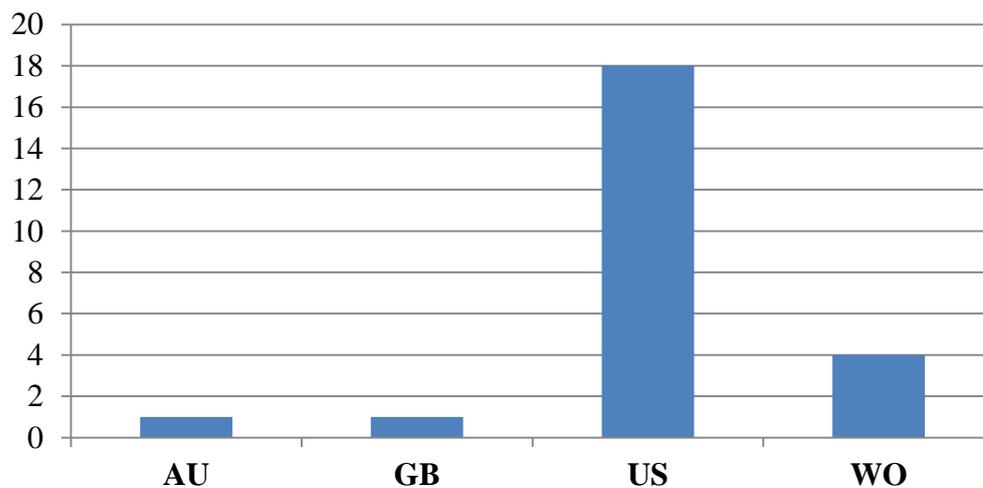
Para realizar a análise dos dados da pesquisa utilizamos o software Microsoft Excel 2013 para tabular o resultado e gerar os gráficos de forma dinâmica (LOLI, 2019); (OLIVEIRA, et al 2020); (SANTOS 2021).

A Figura 1 mostra o número de patentes por base de dados. Observou-se que o maior número de depósito de patente está nos bancos da WIPO com 22 patentes e USPTO com 2 depósitos. Não houve resultado no banco EPO.

Figura 1. Número de patentes por banco

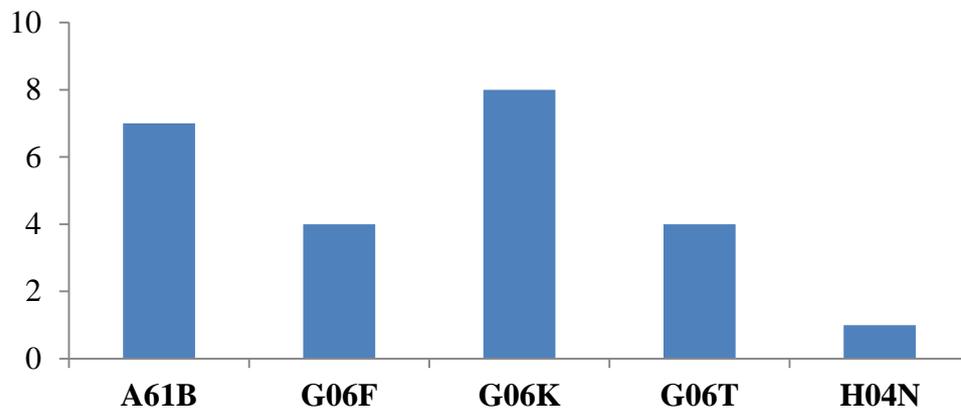
A Figura 2 mostra o número de patentes por países, com destaque para os Estados Unidos (US), com 18 registros, em seguida vem as patentes registradas via Organização Mundial de Propriedade Intelectual (WO), com 4 registros e o Reino Unido (GB) e Austrália (AU) com 1 registro cada.

Analisando a pesquisa de prospecção tecnológica de software para processamento de imagem da mama foi possível observar também, que os países desenvolvidos em destaque como os Estados Unidos (US), estão bem à frente em pesquisa e desenvolvimento de patentes de software para diagnóstico do câncer de mama em comparação com os países em desenvolvimento.

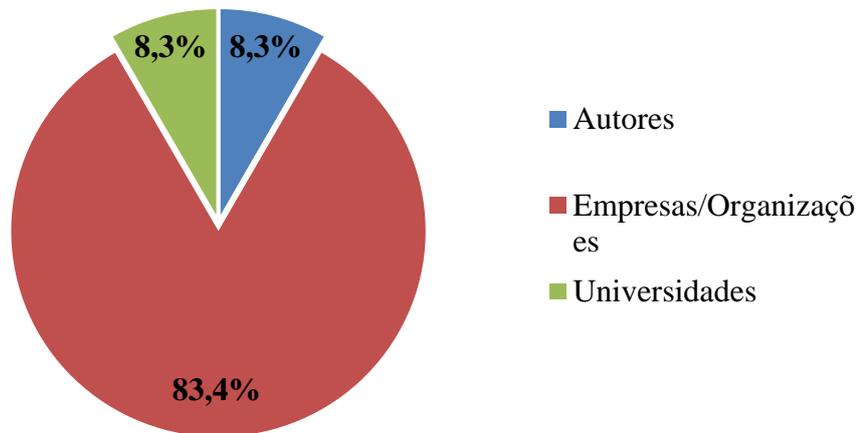
Figura 2. Número de patentes por país

Na Figura 3, observamos o número de patentes por área de classificação internacional de patentes – IPC (INPI, 2019). Observa-se que não há uma classificação específica para a área de computação ou engenharia (RIBEIRO et al., 2016), na prospecção teve o seguinte resultado: G06K (identificação/apresentação de dados) com oito patentes, A61B (diagnóstico, cirurgia) com sete registros de patentes, G06F (processamento elétrico de dados digitais) e G06T (processamento/geração de dados de imagem) com quatro patentes cada e por fim H04N (comunicação de imagem) com uma patente (WIPO, 2021).

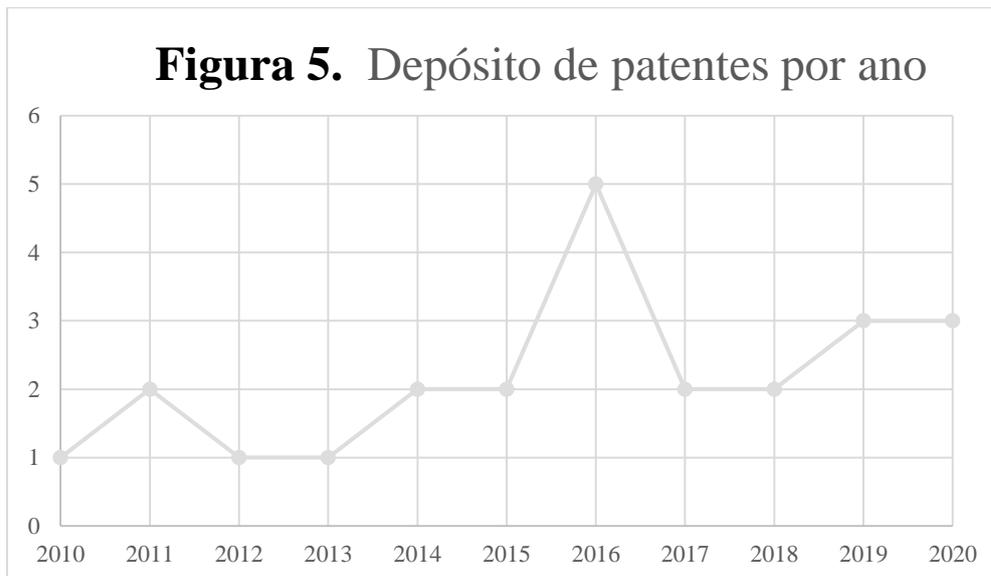
Figura 3. Número de patentes por área de classificação



Observa-se na Figura 4 a porcentagem de depositantes por autores, empresas/organizações e universidades nos últimos dez anos. Com o resultado de 83,4% as empresas mostraram maior índice de pedidos de depósito de patentes, isto demonstra um maior investimento em inovação tecnológica para se tornarem competitivas no mercado, desenvolvendo softwares inteligentes com uso de redes neurais, que auxiliem no diagnóstico do câncer de mama.

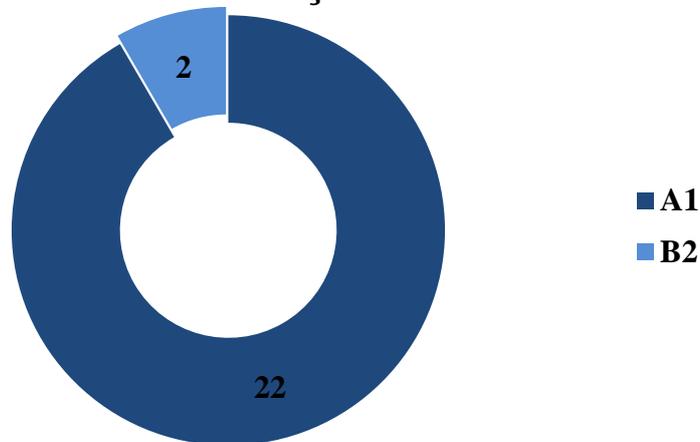
Figura 4. Porcentagem de depositantes

Observando a quantidade de patentes entre os anos de 2010 e 2020, percebe-se que houve uma pequena evolução no decorrer desses anos pesquisados com destaque para o ano de 2016 que teve um total de 6 patentes. Isso mostra que, a pesquisa sobre as redes neurais aplicadas ao diagnóstico do câncer de mama vem em uma crescente constante.

Figura 5. Depósito de patentes por ano

Na Figura 6, obteve-se o seguinte resultado quanto ao tipo de documento publicado de patentes: A1 (necessidade humana) com 22 patentes e B2 (operações de processamento) com 2 patentes, de acordo com (INPI, 2020). O pedido de patentes é publicado com seus devidos relatórios de pesquisas, apresentado em duas vias, protocolado e atribuído o número automaticamente pelo sistema PAG (Sistema de Protocolo Automatizado Geral).

Figura 6. Número de patentes por área de classificação



Conclusão

A prospecção realizada utilizando os depósitos de publicações de base de patentes EPO, WIPO e USPTO de registros de programas de computadores com uso de redes neurais para o diagnóstico do câncer de mama, levando em consideração a classificação BI-RADS, mostrou destaque somente para as bases da WIPO com 22 registros e USPTO com 2 registros.

O país que apresentou maior número de registro de patentes encontrado foram os Estados Unidos (US), onde obteve um destaque na pesquisa com 18 registros em um total de 24 depósitos de patente de software. Observa-se que os países desenvolvidos estão investindo em pesquisa sobre redes neurais e desenvolvendo software com técnicas de inteligência Artificial para melhorar o diagnóstico do câncer de mama.

Ressalta-se que a maioria das patentes de softwares encontrada na pesquisa estão presentes nas empresas/organizações com um total de 83,4% enquanto universidades e autores detêm cada um 8,3% das patentes. Este resultado demonstra a necessidade das universidades desenvolverem um projeto de extensão voltada para aumentar a quantidade de patentes de softwares.

Os resultados obtidos na pesquisa prospectiva demonstram que a área de redes neurais vem em uma crescente expansão, abrindo um leque de oportunidades para avanço de projeto PD&I (pesquisa, desenvolvimento e inovação), na implantação e desenvolvimento de novas tecnologias (software) voltadas ao melhoramento do diagnóstico de qualquer tipo de doença crônica, em particular o câncer de mama.

Referências

Who. Cancer. World Health Organization. 2020 [Online]. Available: <http://www.who.int/en/news-room/fact-sheets/detail/cancer> [Accessed: 04/08/2021].

INCA National Cancer Institute. Controle do Cancer de Mama, 2019 [Online]. Available: <https://www.inca.gov.br/controle-do-cancer-de-mama/dados-e-numeros/incidencia>. [Accessed: 12/09/2021].

Alves-Ribeiro, F. A., Costa-Silva, D. R., Escórcio-Dourado, C. S., da Silva-Neto, O. P., de Castro-Gonçalves, M. E., de Ribeiro, V. O., ... & da Silva, B. B. (2017). Masse detection in mammographic images using texture feature extraction and neural networks. *Journal of Computational and Theoretical Nanoscience*, 14(4), 2064-2068.

Wang, J. M., Zhao, H. G., Liu, T. T., & Wang, F. Y. (2020). Evaluation of the association between mammographic density and the risk of breast cancer using Quantr.

Silva, H. C. D. (2020). Desenvolvimento de vestimenta de proteção radiológica para tireoide para exames de mamografia.

Prado, N., Loiola, P., Guimarães, T., Ohara, E. C. C., & Oliveira, L. D. R. (2020). Gestante com diagnóstico de câncer de mama: prevenção, diagnóstico e assistência. *Brazilian Journal of Health Review*, 3(1), 1109-1131.

Giess, C. S., Frost, E. P., & Birdwell, R. L. (2012, August). Difficulties and errors in diagnosis of breast neoplasms. In *Seminars in Ultrasound, CT and MRI* (Vol. 33, No. 4, pp. 288-299). WB Saunders.

Tortora, G. J., & Derrickson, B. (2016). *Corpo Humano-: Fundamentos de Anatomia e Fisiologia*. Artmed Editora.

Balas, N., Yun, H., Jaeger, B. C., Aung, M., & Jolly, P. E. (2020). Factors associated with breast cancer screening behaviors in a sample of Jamaican women in 2013. *Women & health*, 60(9), 1032-1039.

Justo, N., Wilking, N., Jönsson, B., Luciani, S., & Cazap, E. (2013). A review of breast cancer care and outcomes in Latin America. *The oncologist*, 18(3), 248-256.

El Atlas, N., El Aroussi, M., & Wahbi, M. (2014, November). Computer-aided breast cancer detection using mammograms: A review. In *2014 Second World Conference on Complex Systems (WCCS)* (pp. 626-631). IEEE.

Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C., Abraham, K. T., & Ng, K. H. (2012). Computer-aided breast cancer detection using mammograms: a review. *IEEE Reviews in biomedical engineering*, 6, 77-98.

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. 2020 [Online] Available: <http://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br> [Accessed: 08/06/2021].

Loli, A. C. (2019). Utilização do excel para a construção e análise de gráficos no curso de administração em um centro universitário de guarapuava-PR. *Reso-Revista de Estudos Sociais*, 2(2), 03-19.

Oliveira, S. S., Camargo, M. E., Russo, S. L., Priesnitz, M. C., & da Silva, M. B. (2020). Technological Prospecting: A Mapping of the Patent Applications Related of Internet of Things. *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, 67(1), 155-170.

Santos, W. B., Galhardo, C. X., & Diniz, M. C. (2021). Prospecção Tecnológica de Softwares Utilizados por Universidades Públicas Brasileiras para Promover a Interação com o Setor Produtivo. *Cadernos de Prospecção*, 14(4), 1295-1309.

Alves-Ribeiro, F. A., de Sousa Freitas, M., da Silva, T. L. B., de Oliveira Júnior, W., de Castro-Gonçalves, M. E., Costa-Silva, D. R., ... & da Silva, B. B. (2016). Software Survey for Breast Image Processing. *International Journal of Electronics Communication and Computer Engineering*, 7(5), 266.

INPI. Guia básico de programa de computador. Serviços - Programa de Computador, 2019. Disponível em: <http://www.inpi.gov.br/menu-servicos/programa-de-computador/guia-basico-programa-de-computador>. Acesso em: 25 julho. 2020.

WIPO World Intellectual Property Organization IPC 2021. (Disponível: <https://www.wipo.int/classifications/ipc/ipcpub/?notion=scheme&version=20210101&symbol=none&menulang=en&lang=en&viewmode=f&fipcp=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>). Acessado: 01/02/2021.

INPI (Instituto Nacional de Propriedade Industrial) 2020 Manual de Marcas do INPI. Disponível: <http://manualdemarcas.inpi.gov.br>, Acessado: 08/05/2021

3.2 Capítulo 2

Classificação de lesões mamárias das categorias 4 e 5 do padrão BI-RADS® utilizando redes neurais



Artigo aceito para publicação,
Research, Society and Development

Fator de Impacto: 1,78

ISSN: 2525-3409

QUALIS: A3

Classificação de lesões mamárias das categorias 4 e 5 do padrão BI-RADS® utilizando redes neurais

Recebido: 00/01/2022 | Revisado: 00/01/2022 | Aceito: 00/01/2022 | Publicado: 00/01/2022

Elmo de Jesus Nery Júnior

ORCID: <https://orcid.org/0000-0003-0225-5384>

Universidade Federal do Piauí, Brasil

E-mail: elmojunior.cx@gmail.com

Otílio Paulo da Silva Neto

ORCID: <https://orcid.org/0000-0003-1940-9773>

Instituto Federal do Piauí, Brasil, Brasil

E-mail: otilio.paulo@ifpi.edu.br

Francisco Adelson Alves Ribeiro

ORCID: <https://orcid.org/0000-0003-2850-8028>

Instituto Federal do Maranhão, Brasil

E-mail: adelton@ifma.edu.br

Francisco das Chagas Alves Lima

ORCID: <https://orcid.org/0000-0002-0447-4911>

Universidade Estadual do Piauí, Brasil

E-mail: fdcalima@gmail.com

Larysse Maira Cardoso Campos Verdes

ORCID: <https://orcid.org/0000-0001-6064-149X>

Universidade Federal do Piauí, Brasil

E-mail: mayra9_3@hotmail.com

Danylo Rafael Costa Silva

ORCID: <https://orcid.org/0000-0001-6782-5312>

Universidade Federal do Piauí, Brasil

E-mail: danylorafael@hotmail.com

Maria da Conceição Barros Oliveira

ORCID: <https://orcid.org/0000-0002-8436-3604>

Universidade Federal do Piauí, Brasil

E-mail: mariah.da.concecao@hotmail.com

Pedro Henrique Bandeira Diniz

ORCID: <https://orcid.org/0000-0002-4178-6738>

Universidade Federal do Maranhão, Brasil

E-mail: pedro_hbd@hotmail.com

Anselmo Cardoso de Paiva

ORCID: <https://orcid.org/0000-0003-4921-0626>

Universidade Federal do Maranhão, Brasil

E-mail: paiva@nca.ufma.br

Aristófaes Corrêa Silva

ORCID: <https://orcid.org/0000-0003-0423-2514>

Universidade Federal do Maranhão, Brasil

E-mail: ari@nca.ufma.br

Resumo

O câncer de mama é a doença com mais incidência entre as mulheres em todo mundo, estimativa para Brasil no biênio de 2020-2021 cerca de 66.280 novos casos de câncer de mama que corresponde a uma taxa 29,7% dos casos na

população feminina e cerca de 15.000 mortes pela doença. A mamografia é um dos exames mais utilizado para detecção precoce desde tipo de neoplasia. No entanto, erros acontecem na leitura e interpretação dos laudos, mesmo um profissional bem treinado apresenta uma taxa de acertos entre 65% a 75% com uma quantidade de falso negativo variando entre 15% a 30% e um falso positivo de 7% a 10%, resultando em quantidade desnecessária de biópsia, de 65% a 90% das biópsias de tecido com suspeita de câncer apresentam-se benigna, causando repercussão emocional e física para as pacientes. Sistemas computacionais podem ser desenvolvidos para auxiliar no diagnóstico médico. Este artigo aplicou as técnicas de redes neurais para desenvolver uma ferramenta computacional capaz de classificar lesões das categorias 4 e 5 do padrão BI-RADS®. Os resultados adquiridos pelo software, observaram que o melhor classificador no que diz respeito à taxa de acerto acurácia, foi o Deep Learning, atingindo um percentual de 82,60%, o Support Vectors Machine - SVM teve um percentual de 73,97%. Isto demonstra que as técnicas de redes neurais utilizadas no projeto do software mostram uma eficácia na tarefa de classificação das lesões.o resumo.

Palavras-chave: Câncer de Mama; Classificação BI-RADS®; Processamento de imagem; Redes neurais.

1. Introdução

O câncer de mama é a neoplasia maligna mais incidente em mulheres na maior parte do mundo, caracterizado especificamente pela divisão incorreta e desordenada das células na região dos seios, motivo pelo qual leva este nome, podendo ocorrer tanto no corpo masculino quanto no feminino, no entanto, são as mulheres que têm a massacrante maioria dos casos de câncer mamário (BATISTA et al., 2020; SANTOS, 2018).

Segundo as estatísticas extraídas pelo Observatório Global do Câncer (GLOBOCAN, 2020), em 2020, foram estimados 2,2 milhões de casos novos de câncer e 690 mil óbitos pela doença em 185 países do globo, comprovando que este tipo de câncer é o que mais afeta a população no mundo atual.

A estimativa para o Brasil, de acordo com Instituto Nacional do Câncer (INCA), para biênio de 2020-2021 é o diagnóstico de 66.280 novos casos de

câncer de mama que corresponde a uma taxa 29,7% dos casos na população feminina e cerca de 15.000 mortes pela doença, com um risco de aproximadamente 43,74 para cada 100 mil mulheres (Inca, 2019). Todavia há estratégias para se reduzir a mortalidade por câncer de mama. A mamografia é o exame mais utilizado para a detecção precoce deste tipo de neoplasia, o que facilita a identificação de lesões mamárias em sua fase inicial, fator primordial para o sucesso do tratamento (ALVES, RIBEIRO et al., 2017).

A mamografia, também chamada mamograma ou mamografia digital, é um exame realizado que submete as mamas a raios-X para produzir imagens de alta resolução e permitir que o especialista analise e identifique de forma visual onde se encontra a lesão (MOORE et al., 2018). De acordo com estudos observacionais, quando realizado o rastreamento por meio da mamografia visando a detecção precoce, reduz-se o índice de mortalidade em mulheres acometidas pelo câncer de mama em cerca de 15% a 25% (SALA, 2021).

O diagnóstico prévio é a melhor estratégia para prevenção desse tipo de câncer, os exames clínico de imagem da mama são os mais recomendados, com destaque a mamografia se torna um dos métodos mais eficazes (BALAS et al., 2020; GIESS, 2012; PRADO, 2020; TORTORA, 2016). Ressalta-se que a leitura e interpretação das imagens são tarefas difíceis que exigem experiência do radiologista visto que mesmo um profissional bem treinado apresenta uma taxa de acertos entre 65% a 75% (JUSTO et al., 2013).

No entanto, apesar de todo o avanço tecnológico e científico, era comum que houvesse confusão entre médicos que tratam o câncer de mama perante a análise dos diagnósticos e dos laudos, o que causava problemas aos pacientes relativos às abordagens terapêuticas (GUIMARÃES & CHOINIAC, 2019).

Com base nestas divergências, no ano de 1993, o Colégio Americano de Radiologia, padronizou os achados mamográficos para que os especialistas “falassem a mesma língua” reduzindo a confusão na interpretação dos laudos, foi criado o sistema de classificação BI-RADS® (Breast Image Reporting and Data System - Sistemas de Dados e Relatório de Imagem da Mama), tem o papel importante na indicação de biópsia em lesões suspeitas (ACR, 2016). A sua classificação é dividida por categoria variando de 0 a 6, em cuja classificação a categoria 0: incompleta; 1 e 2 achado benigno; 3 achado provavelmente benigno;

4 achado suspeito variando de 2% a 95%; 5 achado maligno e 6 malignidade comprovada por biópsia.

A mamografia apresenta falha na sua interpretação com uma quantidade de falso negativo de 15% a 30% e um falso positivo variando de 7% a 10% em mulheres com idade reprodutiva, resultando em quantidade desnecessária de biópsia, de 65% a 90% das biópsias de tecido com suspeita de câncer apresentam-se benigna, causando repercussão emocional e física para as pacientes (ALVES, RIBEIRO et al., 2017; FERNANDES, 2014).

No entanto, ferramentas computacionais podem auxiliar os profissionais médicos na interpretação e no diagnóstico dessas imagens, como por exemplo, os sistemas Computer Aided Detection – CADe (Detecção auxiliado por computador) e os sistemas Computer Aided Diagnosis – CADx (Diagnóstico orientado por computador) os quais aumentam a sensibilidade, especificidade e acurácia do câncer da mama. Tais ferramentas aplicadas em conjunto com as técnicas de redes neurais para o processamento de imagens mamográficas, podem melhorar e extrair informações úteis ao diagnóstico preciso do câncer de mama melhorando o índice de acertos médicos (EL ATLAS et al., 2014; GANESAN et al., 2012).

As redes neurais artificiais, ou simplesmente redes neurais, são estruturas computacionais que tentam emular o comportamento do cérebro humano. Para tal, essas redes são formadas pela interligação de estruturas computacionais chamadas neurônios. Esses neurônios são processadores de dados que tentam se comportar como neurônios naturais, recebendo estímulos nas entradas, processando esses estímulos e através de uma função de ativação exibem uma saída, que para o caso mais simples (o perceptron de camada única) é a simples ativação, ou não, da saída, em valor alto ou baixo dependendo do valor das entradas e de um limiar (HUYNH et al., 2016).

O presente artigo visa aplicar as técnicas de redes neurais para desenvolver uma ferramenta computacional capaz de classificar lesões das categorias 4 e 5 do padrão BI-RADS®.

2. Metodologia

O presente trabalho é um estudo experimental sobre classificação BI-RADS® de lesões em imagens mamográficas auxiliadas por computador, a partir do qual foi desenvolvido o software SPIBIRADS – Sistemas de Processamento de Imagem para classificar lesões mamárias baseado no padrão BI-RADS®, certificado pelo Instituto Nacional de Propriedade Industrial - INPI registrado com nº BR512019001754-4. O Estudo incluiu 350 imagens divididas em duas categorias: BI-RADS® 4 e BI-RADS® 5. As imagens utilizadas foram adquiridas na base de dados de domínio público DDSM (Digital Database for Screening Mammography – Banco de dados Digital para Mamografia de Rastreamento), da Universidade do Sul da Flórida, EUA. E em seguida, realizou-se o pré-processamento da imagem para remoção de estrutura indesejada (PAVAN, 2017), seguido pela segmentação esse processo é realizado para buscar a melhor limiar da imagem. Utilizamos três procedimentos, o primeiro foi método de OTSU e o segundo foi o método PSO (Particle Swarm Optimization - Otimização por exame de partículas) e o terceiro aplicando o algoritmo Graph Clustering - Agrupamento de grafo, aplicados nas pesquisas de (SILVA, NETO et al., 2016) (ALVES, RIBEIRO et al., 2017). Próximo passo fez-se a separação da base, extração das características e classificação das lesões com SVM (Support Vector Machines - Máquina de Vetores de Suporte) e Deep Learning - Aprendizado Profundo.

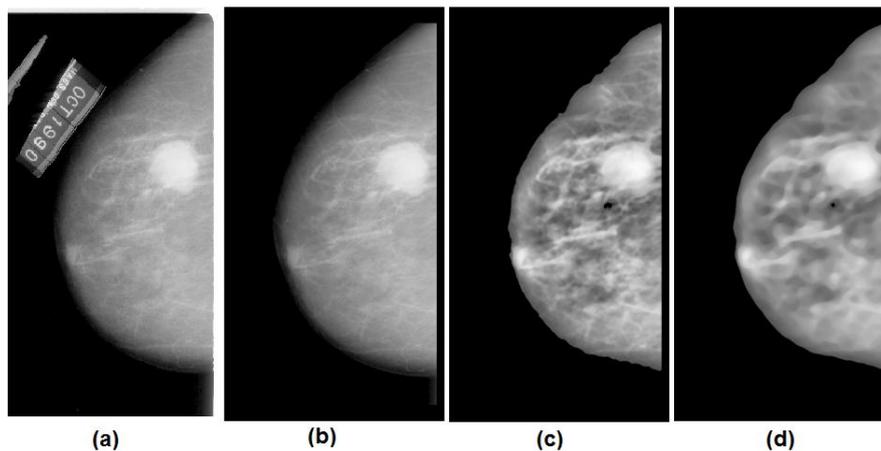
2.1. Base de Imagens

As imagens mamográficas foram adquiridas na base de dados de domínio público DDSM com uma quantidade aproximadamente de 2.600 imagens. Cada caso mamográfico está disposto pela DDSM em imagens obtidas nas duas mamas de cada paciente, nas posições Médio-Lateral Oblíqua e Crânio-Caudal. Tais imagens são disponibilizadas no padrão LJPEG e representadas em níveis de cinza com resolução de 12 ou 16 bits por pixels. Os laudos gerados por especialistas estão inseridos às imagens contendo informações como o código de cada paciente, a idade da mesma na época do estudo, a classificação na categoria BI-RADS® (Breast Imaging Report and Data System) e na composição da densidade mamária segundo o mesmo padrão, além de estarem acompanhados de tipo de anormalidade e a localização da lesão (HEATH & BOWYER, 2000).

2.2. Pré-processamento

Em seguida, se faz necessária a remoção das estruturas não-desejadas, “limpando” a imagem dos ruídos, bordas, marcações, músculos desinteressantes, e outras informações que divergem do estudo pretendido, mas que estão visíveis na imagem que podem atrapalhar a clareza e assertividade do diagnóstico (PAVAN, 2017). Tal procedimento precede a segmentação das imagens. Após a remoção destas estruturas, é aplicada a técnica de realce baseada no histograma e o filtro da média, com finalidade de destacar as estruturas internas das mamografias, ou seja, das regiões que o estudo deve ser focado como mostra a Figura 1 (ALVES, RIBEIRO et al., 2017) (SILVA, NETO et al., 2017).

Figura1. Resultado das etapas do melhoramento: (a) Imagem original; (b) Imagem sem as bordas e marcações; (c) Realce local; (d) Filtro da média.



Fonte: Autoria Própria.

2.3. Segmentação

Após o pré-processamento, ocorre a segmentação, que consiste na separação das regiões mais homogêneas e similares na imagem, concluindo nos objetos considerados suspeitos, tais como nódulos, massas estranhas, microcalcificações, assimetrias, ou qualquer outra anomalia que foge da anatomia da região das mamas em questão (ALVES, RIBEIRO et al., 2017; SILVA, NETO, 2016). Neste projeto, é proposta a abordagem de segmentação dividida em três etapas, na primeira utilizamos o método (OTSU, 1979), a segunda o método PSO - Otimização por exame de partículas (VAN DEN

BERGH, 2007) e a terceira é composta pelo Graph Clustering - Agrupamento de grafo.

2.4. Otsu

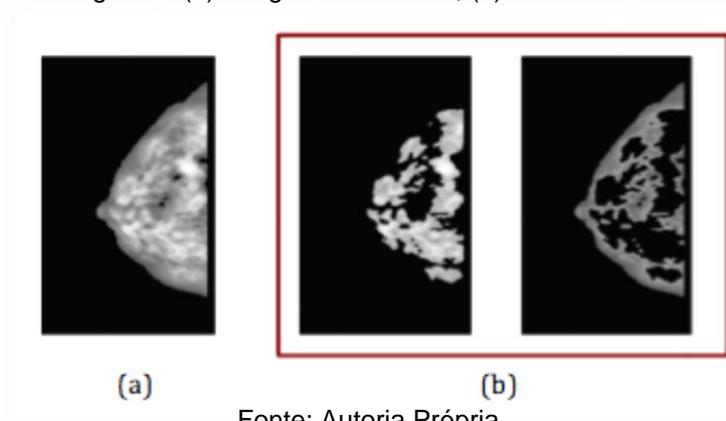
O algoritmo de (OTSU, 1979) reconhece que a imagem está dividida em duas classes, objeto e fundo, e busca localizar especificamente encontrar a limiar que melhor divide essas anomalias por meio de múltiplas iterações, calculando a diferenciação entre as classes de objeto e fundo para todos e quaisquer possíveis limiares, aplica-se a Equação 1 para encontrar o desvio padrão gerado.

$$\delta^2 = \frac{1}{N} \sum_{j=1}^N (y_i - m_{ij})^2, \quad (1)$$

Então o $\delta^2 > dp_{mi}$, aplica o centróide g de cada classes/cluster com o novo limiar para se dividir este cluster em dois novos classes/clusters. Aplica-se mais um vez o cálculo do desvio padrão deste cluster, se $(\delta^2 > dp_{mi})$, conseqüentemente, este passo é executado de forma recursiva até que δ^2 não seja maior que dp_{mi} . N é o número de pixels das imagens.

Quando definido, todos os pixels da imagem que estiverem destacadas, ora, cujos valores de nível acinzentado são menores ao limiar definido, esses serão selecionados para o primeiro cluster e, assim, os valores cujo nível de cinza forem maior ou igual ao estipulado em limiar, são selecionados como pertencentes ao segundo cluster, permitindo melhor diferenciação e identificação da região como mostra a Figura 2. (SILVA, NETO et al., 2017).

Figura 2. (a) Imagem de entrada; (b) Processamento Otsu.



Fonte: Autoria Própria.

2.5. Otimização por exame de partículas - PSO

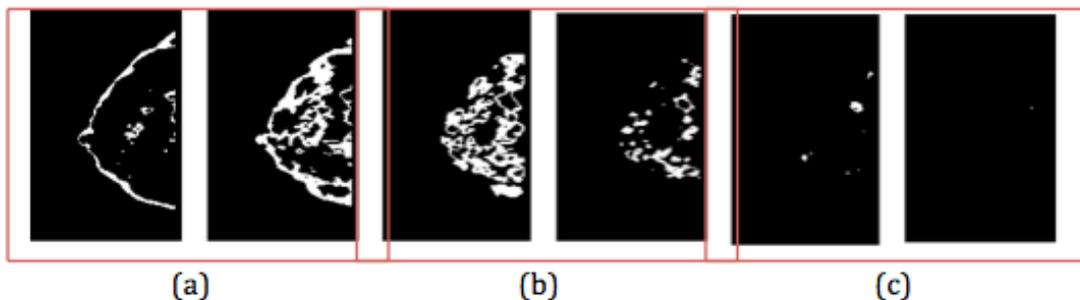
Depois de receber a imagem processada aplica-se o algoritmo PSO (Particle Swarm Optimization) para otimizar processo e encontrar a melhor limiar da imagem e conseqüentemente os melhores clusters/agrupamento, de acordo (MERWE et al., 2003), cada partícula do PSO é formada por um vetor de centróides de tamanho N_c é a quantidade de grupos encontrados, e caracteriza um resultado para o problema. Quando acontece uma iteração, atribui-se os dados ao centróide ao qual estão mais próximos, avalia os resultados atualizados de acordo com a melhor posição de entrada por alguma partículas.

Para cada partícula X_i é tido como um conjunto de clusters, sendo assim, cada partícula representaria uma resposta completa de clusterização, isto é, no enxame do PSO, cada partícula possui um vetor de k centróides que são construídos de acordo com Equação 2 (RIBEIRO, ALVES et al., 2017).

$$x_i = (mi_1 mi_2, \dots, mi_j mi_{N_c}) \quad (2)$$

Onde N_c é número de clusters, criados, mi_j corresponde ao j -ésimo centroide da i -ésima partícula, desse modo, uma única partícula caracteriza uma solução do problema de clusterização como mostra na Figura 3.

Figura 3. Resultado dos clusters gerados: (a) 10 cluster gerado; (b) 20 cluster gerado a partir do anterior; (c) 30 cluster gerado a partir do segundo.



Fonte: Autoria Própria.

Depois das imagens mamográficas segmentada através do método OTSU e PSO verifica as regiões suspeitas de lesões, o próximo passo será aplicado o graph clustering para agrupar as regiões de interesse (ROI).

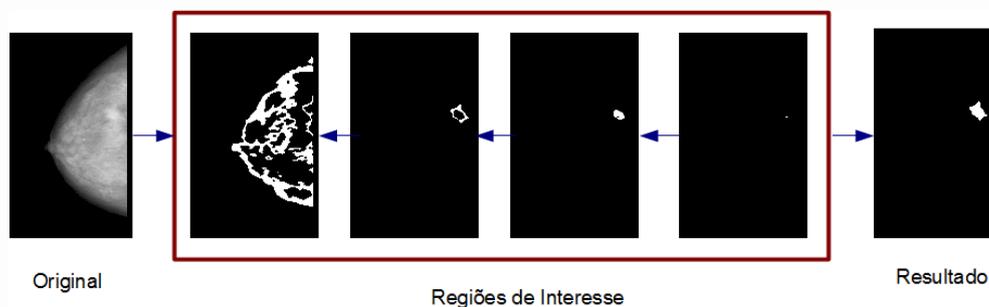
2.6. Graph Clustering - Agrupamento de Grafo

O Graph Clustering é um método de agrupamento dos vértices do gráfico em clusters considerando a estrutura das arestas das imagens. Neste trabalho, definimos alguns parâmetros: a região será definida em 3x3, o gráfico será formado a partir das regiões de interesse (ROI) de uma imagem, verificando sua vizinhança com todas as ROIs encontrados no cluster original, assim, será direcionado o grafo. Ao concluir o processo, os nós do grafo que possuírem mais de duas ligações serão excluídos e os nós que tenham nenhuma ou no máximo duas ligações permanecerão. Cada nó do gráfico resposta desse processo, simula uma ROI (SCHAEFFER, 2007; SILVA, NETO et al., 2017).

Em seguida, (SCHAEFFER, 2007) esclarece os cálculos do graph cluster da seguinte forma: A partir de um nó qualquer do grafo, é calculado seu valor de Fator de Forma Circular (FFC) (Equação FFC), da seguinte forma: a) se FFC for menor que 10%: o nó será descartado e será escolhido outro nó, e o processo se repete; b) se FFC for maior que 10%: verifica-se suas adjacências, realizando as uniões do nós; c) após cada união, o FFC é calculado novamente, caso este, seja maior que 10% a união é válida e será verificado (caso exista) a próxima adjacência; d) se o FFC, após cada união, resultar em um valor menor que 10%, esta união não será válida, e o nó que foi unido será descartado.

Na Equação 3 FFC, que menciona (ALVES, RIBEIRO et al., 2017), A corresponde à área e P ao perímetro de cada ROI e o cálculo se apresenta da seguinte forma: $FFC = (4 \cdot \pi \cdot A) / p^2$ O percentual da FFC foi adotado em 10%, em decorrência dos testes, sendo este o valor que apresentou melhores resultados como mostra na Figura 4.

Figura 4: Resultado do algoritmo Graph Clustering.



Fonte: Autoria Própia.

2.7. Separação da base

Para realizar o treinamento é muito importante organizar corretamente a base de dados de entrada de forma que as redes neurais façam o aprendizado correto, ou seja, o treinamento bem realizado depende da preparação dos dados, isso determina a taxa de aprendizado e predição na execução dos testes. A primeira etapa é organizar o conjunto de imagens que foram adquiridas conforme o trabalho (ALVES, RIBEIRO et al., 2017; SILVA, NETO et al., 2017).

Em seguida, as lesões segmentadas passam por um pré-processamento para identificar o maior X (largura) e maior Y (altura) dentre todas as lesões, para se construir uma boundingbox - Caixa delimitadora padrão com fundo preto, inserindo as lesões no centro dessa boundingbox e deixando todas as imagens com as mesmas dimensões de largura e altura. Conforme ilustrado na Figura 5 (a) e (b).

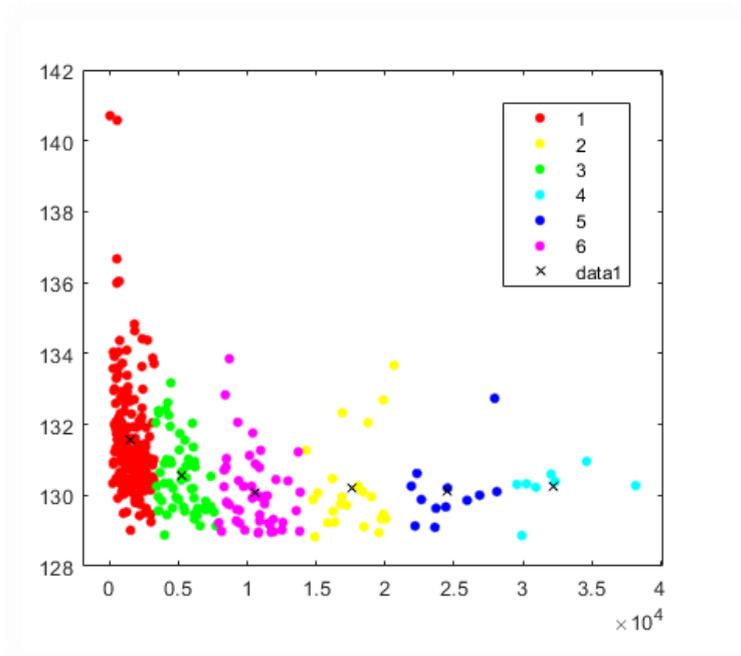
Figura 5: Nódulo em fundo preto



Fonte: Autoria Própria.

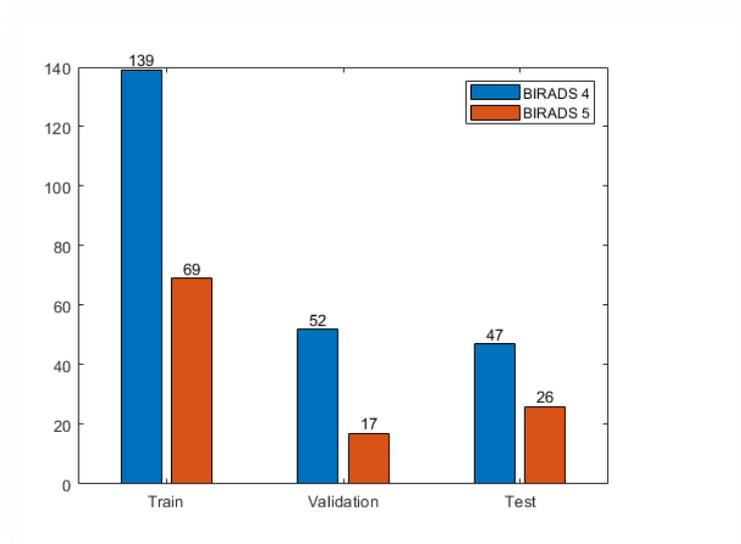
Na Figura 6, ilustra-se o algoritmo de agrupamento k-means (PACIFICO, 2020) utilizado para criar os grupos de lesões baseado no desvio padrão, média e tamanho das lesões, gerando seis grupos para subdividir as lesões grandes e pequenas igualmente nas bases de treinamento, validação e teste como mostra na Figura 7. As imagens são rotuladas em duas categorias BI-RADS® 4 e BI-RADS® 5.

Figura 6: Visualização das dimensões das lesões



Fonte: Autoria Própria.

Figura 7: base treinamento, validação, teste



Fonte: Autoria Própria.

2.8. Extração de características

A extração das características pode ser realizada através da forma e textura dos nódulos. A extração utilizada nesta pesquisa será a partir da textura

dos nódulos mamários, aplicando-se os descritores de textura que expõe essas características através do objeto da imagem bem como, suavidade, rugosidade, uniformidade e granulação (GONZALEZ; WOODS & EDDINS, 2010). Para caracterizar a textura ROI – Região de Interesse, utiliza-se as diversidades funcionais de acordo com (TILMAN, 2001) que são as variações e valores de uma espécie e suas características influenciadoras no funcionamento de uma determinada comunidade. Foram utilizados os índices de diversidades funcionais para complementar a extração conforme a pesquisa (SILVA, NETO et. al., 2017), as ROIs - regiões de interesse dispõem da seguinte características: FADa - índice de diversidade funcional abundante, FADe - índice de diversidade funcional abundante da espécie, FADp - índice de diversidade funcional atributo pixel. Os índices extraídos permitem características mais detalhadas das regiões.

Outra extração de característica utilizada foi baseada nos descritores de Haralick para verificar os padrões de textura das ROIs. Foram aplicados nove descritores de acordo com (HARALICK, 1973) contraste, homogeneidade, energia, entropia, momento diferença inverso, momento diferença de ordem, probabilidade máxima, variância inversa e correlação, conforme a pesquisa (ALVES, Ribeiro et al., 2017).

3. Resultados e Discussão

3.1 Experimentos

Uma vez que realizado toda metodologia aplicada na pesquisa para as etapas de preparação das imagens mamográficas, aquisição da base de dados DDSM, pré-processamento, segmentação com os métodos OTSU e PSO, graph clustering para encontrar as regiões de interesse (ROI), separação das bases para treinamento, validação, teste e extração das características, os mesmos foram utilizados como entrada para classificação de duas classes BI-RADS® categoria 4 e 5.

Utilizou-se técnicas de redes neurais artificiais para classificação das lesões mamárias baseado no padrão BI-RADS®. A extração e separação dos dados foram realizados em 350 nódulos mamários variando entre as duas categorias BI-RADS® 4 e BI-RADS® 5. Com este grupo de dados, o software foi

utilizado para realizar as classificações dos nódulos mamários através dos algoritmos Support Vectors Machine - Máquina de Vetores de Suporte (SVM) e Deep Learning - Aprendizado Profundo, a métrica utilizada para avaliar os classificadores, foi a acurácia (GONÇALVES, 2017).

3.2 Achados

Os teste foram realizados em 350 nódulos escolhidos na base de dados de domínio público Digital Database for Screening Mammography - DDSM da Universidade do Sul da Flórida. Em seguida, se fez necessário realizar o pré-processamento das imagens para remoção das estruturas não-desejadas, após o pré-processamento, ocorre a segmentação da imagem utilizando o algoritmo de OTSU e PSO, o próximo passo será aplicar o graph clustering para agrupar as regiões de interesse (ROI). Depois de encontrar as ROIs é de suma importância realizar a separação da base para preparar as imagens de treinamento, validação e teste. As imagens foram rotuladas em duas categorias BI-RADS® 4 e BI-RADS® 5. Para a categoria 4 foram utilizadas 139 lesões, enquanto que para a categoria 5 utilizou-se 69 lesões para treino. Ao passo que, na realização da validação foram usadas 52 lesões de categoria 4 e 17 lesões de categoria 5. Para aplicação do teste foram utilizadas 47 lesões de categoria 4 e 26 lesões de categoria 5. O passo seguinte foi aplicar a extração das características utilizando o índice de diversidade e os descritores Haralick. Após realizar todas as etapas descritas acima efetuou-se a classificação dos nódulos mamários através dos algoritmos Support Vectors Machine - Máquina de Vetores de Suporte (SVM) e Deep Learning - Aprendizado Profundo.

O Quadro 1 apresenta os resultados obtidos por cada algoritmo que foi utilizado na etapa de classificação das lesões, observou que o melhor classificador que teve a melhor acurácia foi o Deep Learning, atingindo um percentual de 82,60%, o Support Vectors Machine - SVM obteve um percentual de 73,97%.

Quadro: 1 Resultado dos Classificadores

Técnica	Treino		Validação		Teste		Acurácia média(%)
	Birads4	Birads5	Birads4	Birads5	Birads4	Birads5	
SVM	139	69	52	17	47	26	73,97
Deep Learning	139	69	52	17	47	26	82,60

Fonte: Autoria Própria.

O resultado encontrado demonstra que o Deep Learning é um algoritmo complexo devido ao empilhamento de diversas camadas ocultas que são similares aos neurônios, alimentam-se de uma grande quantidade de dados e são capazes de reconhecer imagens e tem a possibilidade de aprender tarefas extremamente avançadas com poder de tomar decisões intuitivas e inteligentes. O algoritmo *Support Vector Machine (SVM)* não demonstrou um bom desempenho na extração e classificação das imagens mamográficas por utilizar uma menor quantidade de camadas. O bom desempenho do algoritmo depende da precisão com que os recursos serão identificados e extraídos. Qualquer alteração desses parâmetros de entrada, seu desempenho pode melhorar ou piorar os seus resultados (REESE, 2017).

4. Conclusão

Após a análise e discussão com os resultados da pesquisa verificou-se que a metodologia utilizada para detecção e classificação das lesões mamárias das categorias 4 e 5 do padrão BI-RADS® somado com pré-processamento, segmentação OTSU e PSO, Graph clustering, separação da base, extração de característica, e classificação com Deep learning resultou em uma taxa de classificação de 82,60% de acurácia. Este estudo permitiu verificar o potencial do algoritmo Deep learning para classificação e detecção das lesões mamárias e as vantagens de aplicar este método independentemente do tipo de mama.

Esta pesquisa limitou-se às imagens mamográficas utilizando a base de dados de domínio público DDSM da Universidade do Sul da Flórida ao qual está disponível desde 1996 e não houve atualizações recentes. Foram também utilizados apenas uma quantidade de 350 imagens limitando assim o resultado e só foram classificados em duas categorias suspeitas de malignidade que são os BI-RADS® 4 e 5. Desta forma recomenda-se otimizar este método utilizando novas bases de dados e aumentando tanto a quantidade de imagens como

também a quantidade de classificação por categoria de acordo com o padrão BI-RADS® variando categoria de 1 a 3.

Referências

Alves-Ribeiro, F. A., Costa-Silva, D. R., Escórcio-Dourado, C. S., da Silva-Neto, O. P., de Castro-Gonçalves, M. E., de Ribeiro, V. O., ... & da Silva, B. B. (2017). Masse detection in mammographic images using texture feature extraction and neural networks. *Journal of Computational and Theoretical Nanoscience*, 14(4), 2064-2068.

AMERICAN COLLEGE OF RADIOLOGY (ACR) (2016). Atlas BI-RADS do ACR: sistema de laudos e registro de dados de imagem da mama / American College of Radiology; [tradução Angela Caracik] Colégio Brasileiro de Radiologia. 2. ed. - São Paulo

Balas, N., Yun, H., Jaeger, B. C., Aung, M., & Jolly, P. E. (2020). Factors associated with breast cancer screening behaviors in a sample of Jamaican women in 2013. *Women & health*, 60(9), 1032-1039.

Batista, G. V., Moreira, J. A., Leite, A. L., & Moreira, C. I. H. (2020). Câncer de mama: fatores de risco e métodos de prevenção. *Research, Society and Development*, 9(12), e15191211077-e15191211077.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).

Costa, F., & Ferreira, D. D. (2020, December). Classificação de nódulos mamários com máquina de vetores de suporte. In *Congresso Brasileiro de Automática-CBA* (Vol. 2, No. 1).

El Atlas, N., El Aroussi, M., & Wahbi, M. (2014, November). Computer-aided breast cancer detection using mammograms: A review. In *2014 Second World Conference on Complex Systems (WCCS)* (pp. 626-631). IEEE.

Fernandes, I., & dos Santos, W. (2014). Classificação de mamografias utilizando extração de atributos de textura e redes neurais artificiais. In *Congresso Brasileiro de Engenharia Biomédica* (Vol. 8).

Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C., Abraham, K. T., & Ng, K. H. (2012). Computer-aided breast cancer detection using mammograms: a review. *IEEE Reviews in biomedical engineering*, 6, 77-98.

Giess, C. S., Frost, E. P., & Birdwell, R. L. (2012, August). Difficulties and errors in diagnosis of breast neoplasms. In *Seminars in Ultrasound, CT and MRI* (Vol. 33, No. 4, pp. 288-299). WB Saunders.

Giger, M. L. (2000). Computer-aided diagnosis of breast lesions in medical images. *Computing in Science & Engineering*, 2(5), 39-45.

International Agency for Research on Cancer. (2020). *Global cancer observatory: cancer today*. 2020.

Gonçalves, C. B. (2017). *Detecção de câncer de mama utilizando imagens termográficas*.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2010). *Morphological reconstruction. Digital image processing using MATLAB, MathWorks*.

Guimaraes, M. D., Bitencourt, A. G., Marchiori, E., Chojniak, R., Gross, J. L., & Kundra, V. (2014). Imaging acute complications in cancer patients: what should be evaluated in the emergency setting?. *Cancer Imaging*, 14(1), 1-12.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.

Heath, M. D., & Bowyer, K. W. (2000, June). Mass detection by relative image intensity. In *Proceedings of the 5th International Workshop on Digital Mammography (IWDM-2000)* (pp. 219-225).

Huynh, B. Q., Li, H., & Giger, M. L. (2016). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3), 034501.

INCA National Cancer Institute. *Controle do Cancer de Mama*, (2019) [Online]. Available: <https://www.inca.gov.br/controle-do-cancer-de-mama/dados-e-numeros/incidencia>. [Accessed: 12/09/2021].

Justo, N., Wilking, N., Jönsson, B., Luciani, S., & Cazap, E. (2013). A review of breast cancer care and outcomes in Latin America. *The oncologist*, 18(3), 248-256.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.

Moore, K. L., Dalley, A. F., & ARTHUER, F. (2018). *AgurAMR. Anatomia orientada para a clínica*. Guanabara 2014.

Neto, O. P. S., Silva, A. C., Paiva, A. C., & Gattass, M. (2017). Automatic mass detection in mammography images using particle swarm optimization and functional diversity indexes. *Multimedia Tools and Applications*, 76(18), 19263-19289.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.

Pacífico, L. (2020, October). Agrupamento de imagens baseado em uma abordagem híbrida entre a otimização por busca em grupo e k-means para a segmentação automática de doenças em plantas. In *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional* (pp. 152-163). SBC.

Pavan, A. L. M., Vacavant, A., Trindade, A. P., & de Pina, D. R. (2017). Fibroglandular tissue quantification in mammography by optimized fuzzy C-means with variable compactness. *Irbm*, 38(4), 228-233.

Prado, N., Loiola, P., Guimarães, T., Ohara, E. C. C., & Oliveira, L. D. R. (2020). Gestante com diagnóstico de câncer de mama: prevenção, diagnóstico e assistência. *Brazilian Journal of Health Review*, 3(1), 1109-1131.

Reese, H. (2017). Understanding the differences between AI, machine learning, and deep learning. URL: <https://www.techrepublic.com/article/understandingthedifferencesbetweenaimachinelearninganddeeplearning>.

Sala, D. C. P. (2021). Rastreamento mamográfico no Brasil: determinantes à implementação no Sistema Único de Saúde e contribuições da Atenção Primária à Saúde (Doctoral dissertation, Universidade de São Paulo).

Santos, T. A., & Gonzaga, M. F. N. (2018). Fisiopatologia do câncer de mama e os fatores relacionados. *Revista Saúde em Foco*, 10, 359-366.

Schaeffer, S. E. (2007). Graph clustering. *Computer science review*, 1(1), 27-64.

Tilman, D. (2001). Functional diversity. *Encyclopedia of biodiversity*, 3(1), 109-120.

Tortora, G. J., & Derrickson, B. (2016). *Corpo Humano-: Fundamentos de Anatomia e Fisiologia*. Artmed Editora.

Van Den Bergh, F. (2007). An analysis of particle swarm optimizers (Doctoral dissertation, University of Pretoria).

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5), 988-999.

4 CONSIDERAÇÕES FINAIS

Concluimos esta pesquisa, expondo que a metodologia aplicada atingiu seu objetivo de desenvolver um método de redes neurais para classificar nódulos mamários no padrão BI-RADS® nas categorias 4 e 5. O primeiro capítulo realizou busca de anterioridade dos registros de programa de computador que faz uso de redes neurais artificiais aplicadas ao diagnóstico de câncer de mama, por meio de uma prospecção tecnológica, a qual foi realizada em três bases de depósitos de patentes: European Patent Office (EPO), World Intellectual Property Organization (WIPO) e United States Patent and Trademark Office's (USPTO). As bases de depósitos de patentes que mais se destacaram foram WIPO e USPTO. Encontrou-se um total de 24 patentes, sendo 22 na WIPO e 2 na USPTO, relacionadas a imagens de câncer mamário e sistema BI-RADS. Ressalta-se que busca feita na base de dados EPO não retornou registros relacionados ao tema pesquisado. O segundo capítulo exibe os resultados adquiridos pelo software, mostraram eficácia na utilização das técnicas de redes neurais. Nos algoritmos que foram utilizados na etapa de classificação das lesões de categorias 4 e 5 do padrão BI-RADS, observou que o melhor classificador no que diz respeito a taxa de acerto a acurácia, foi o Deep Learning, atingindo um percentual de 82,60%, o *Support Vectors Machine* - SVM obteve um percentual de 73,97%. Este estudo permitiu verificar o potencial do algoritmo Deep Learning para classificação e detecção das lesões mamárias e as vantagens de aplicar este método independentemente do tipo de mama. Desta forma recomenda-se otimizar este método utilizando novas bases de dados e aumentando a quantidade de classificação por categoria de acordo com o padrão BI-RADS variando categoria de 1 a 3.



REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA ECONOMIA
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512019001754-4**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 28/01/2018, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: SPIBIRADS

Data de criação: 28/01/2018

Titular(es): UNIVERSIDADE FEDERAL DO PIAUÍ

Autor(es): BENEDITO BORGES DA SILVA; ELMO DE JESUS NERY JÚNIOR

Linguagem: C++

Campo de aplicação: SD-05; SD-06; SD-08; SD-09

Tipo de programa: GI-08; IA-01; TC-04

Algoritmo hash: SHA-512

Resumo digital hash:

63FF143DE40DE92EC0170E0AB9F13A04F217470E61C97E6B6DD0A2855F896E253B06CCCC918B71B3EC60A018
91209D037FB129AC41F76DC80E43885FB344B056

Expedido em: 20/08/2019



Aprovado por:
Helmar Alvares

Chefe da DIPTO - Portaria/INPI/DIRPA Nº 09, de 01 de julho de 2019

APÊNDICES



Apêndice – Dados dos Inventores

DADOS DO(S) INVENTOR(ES)

Nome civil completo:

Benedito Borges da Silva

Unidade: Centro de Ciências da Saúde

Departamento: Materno Infantil

Fone institucional: 86 3215 0470

Fax: 86 3215 0470

E-mail: beneditoborges@globo.com

Identidade Nº: 93756

Órgão expedidor: ssp-pi

Data de emissão: 17/03/1980

CPF: 027.232,633-04

Data nascimento: 20/03/1950

Estado Civil: casado

Nacionalidade: Brasileiro

Naturalidade: Teresina

Endereço Residencial Completo: Avenida Elias João Tajra, 1260

Bairro: Jockey

CEP: 64049-300

Telefone Residencial: 86 3232 5063

Celular: 86 99981 5388

Vínculo com a UFPI

Professor

Aluno Especialização

Técnico-administrativo

Aluno Mestrado

Aluno Graduação

Aluno Doutorado

Participante Externo UFPI

Professor

Informe Instituição (Empresa):

Técnico-administrativo

Aluno Graduação

Aluno Pós-Graduação

Outro:

50% Contribuição no presente invento:

Nome civil completo:

Elmo de Jesus Nery Júnior

Unidade: RENORBIO-PI

Departamento: Biotecnologia

Fone institucional: 99 98168-2920

Fax:

E-mail: elmojunior.cx@gmail.com

Identidade Nº: 1195835

Órgão expedidor: SSP-PI

Data de emissão: 30/06/2004

CPF: 968.562.113-68

Data nascimento: 18/03/1980

Estado Civil: casado

Nacionalidade: Brasileiro

Naturalidade: Caxias – MA

Endereço Residencial Completo: Rua Benedito Leite, 1046

Bairro: Centro

CEP: 65604-020

Telefone Residencial: (99) 98807-2845

Celular: (99) 98168-2920

Vínculo com a UFPI

Professor

Aluno Especialização

Técnico-administrativo

Aluno Mestrado

Aluno Graduação

Aluno Doutorado

Participante Externo UFPI

Professor

Informe Instituição (Empresa):

Técnico-administrativo

Aluno Graduação

Aluno Pós-Graduação

Outro:

50% Contribuição no presente invento: