

UNIVERSIDADE CANDIDO MENDES – UCAM
PROGRAMA DE PÓS-GRADUAÇÃO EM PESQUISA OPERACIONAL E
INTELIGÊNCIA COMPUTACIONAL
CURSO DE MESTRADO EM PESQUISA OPERACIONAL E INTELIGÊNCIA
COMPUTACIONAL

SUMARA VARGAS HÜBNER VALINHO

**USO DE TECNOLOGIAS DIGITAIS NO APOIO AO DIAGNÓSTICO
PRECOCE DO CÂNCER DE MAMA**

CAMPOS DOS GOYTACAZES, RJ
Outubro de 2019

UNIVERSIDADE CANDIDO MENDES – UCAM
PROGRAMA DE PÓS-GRADUAÇÃO EM PESQUISA OPERACIONAL E
INTELIGÊNCIA COMPUTACIONAL
CURSO DE MESTRADO EM PESQUISA OPERACIONAL E INTELIGÊNCIA
COMPUTACIONAL

Sumara Vargas Hübner Valinho

**USO DE TECNOLOGIAS DIGITAIS NO APOIO AO DIAGNÓSTICO
PRECOCE DO CÂNCER DE MAMA**

Dissertação apresentada ao Programa de
Mestrado em Pesquisa Operacional e
Inteligência Computacional da
Universidade Candido Mendes –
Campos/RJ, para obtenção do grau de
MESTRE EM PESQUISA OPERACIONAL
E INTELIGÊNCIA COMPUTACIONAL

Orientador: Prof. Ítalo de Oliveira Matias, D.Sc.
Coorientador: Prof. Tiago Samaha Cordeiro, M. Sc

CAMPOS DOS GOYTACAZES, RJ
Outubro de 2019

Catálogo na Fonte

Preparada pela Biblioteca da **UCAM – CAMPOS**

021/2020

Valinho, Sumara Vargas Hubner.

Uso de tecnologias digitais no apoio ao diagnóstico precoce do cancer de mama. / Sumara Vargas Hubner Valinho. – 2019.
73 f. : il.

Orientador: Ítalo de Oliveira Matias.
Coorientador: Tiago Samaha Cordeiro.

Dissertação de Mestrado em Pesquisa Operacional e Inteligência Computacional – Universidade Candido Mendes – Campos. Campos dos Goytacazes, RJ, 2019.

1. Câncer de mama. 2. Câncer de mama - Diagnóstico. 3. Tecnologias digitais. I. Universidade Candido Mendes – Campos. II. Título.

CDU – 618.19-006:004.8

Bibliotecária Responsável: Flávia Mastrogirólamo CRB 7^a-6723

SUMARA VARGAS HÜBNER VALINHO

**USO DE TECNOLOGIAS DIGITAIS NO APOIO AO DIAGNÓSTICO
PRECOCE DO CÂNCER DE MAMA**

Dissertação apresentada ao Programa de
Mestrado em Pesquisa Operacional e
Inteligência Computacional da
Universidade Candido Mendes –
Campos/RJ, para obtenção do grau de
MESTRE EM PESQUISA OPERACIONAL
E INTELIGÊNCIA COMPUTACIONAL

Aprovada em 05 de outubro de 2019.

BANCA EXAMINADORA

Prof. Ítalo de Oliveira Matias, D.Sc. – orientador
UNIVERSIDADE CANDIDO MENDES - CAMPOS

Prof. Tiago Samaha Cordeiro, M.Sc. - coorientador
UNIVERSIDADE CANDIDO MENDES - CAMPOS

Prof. Eduardo Shimoda, D.Sc.
UNIVERSIDADE CANDIDO MENDES - CAMPOS

Prof^a. Claudia Boechat Seufitelli, D.Sc.
INSTITUTO FEDERAL FLUMINENSE

CAMPOS DOS GOYTACAZES, RJ
2019

Dedico este trabalho a Deus e a todas as
pessoas que acreditaram em mim e em
especial aos meus pais, esposo e filhos.

AGRADECIMENTOS

Primeiramente a Deus pela concessão do indispensável para realizar este Mestrado.

Ao meu orientador, Prof. Dr. Ítalo de Oliveira Matias, pela paciência diante de minha ansiedade e pela direção adequada em face de minhas hesitações e medo. Pela prontidão e ajuda em todo o tempo. Mais que um professor, foi um amigo nesse período de incertezas e angústias. Faltarão palavras para descrever toda minha gratidão.

Ao meu coorientador, Prof. Tiago Samaha Cordeiro, por toda ajuda e empenho durante o desenvolvimento desse trabalho.

Ao Prof. Dr. Eduardo Shimoda, pelo apoio nos momentos certos, pela paciência e disponibilidade, sempre pronto a ajudar. O senhor é um dos homens mais nobres que conheci. Sua sabedoria, didática, acessibilidade e bondade fazem de você um exemplo a ser seguido por todos. Eu o admiro muito.

À professora Claudia Boechat Seufitelli, por integrar a banca. Muito obrigada por tudo.

Aos meus pais, Gilson e Ester, por todo amor, ensino e apoio. Jamais poderei retribuir tudo que fizeram por mim. Vocês são meu exemplo, meu porto seguro aqui na Terra. Minha eterna gratidão.

Ao meu esposo e também colega de mestrado, Hugo Valinho, pela parceria, por ser incansável e sempre pronto a ajudar, por ter estado ao meu lado em todos os momentos, tornando essa caminhada um pouco mais leve. Agradeço-lhe por ser meu amigo, companheiro e ajudador. Amo você.

Aos meus filhos, Gabriel e Elisa, por entenderem que por muitas noites tive que estar ausente, os privando da convivência, das brincadeiras e dos filmes de que tanto gostamos de assistir juntos, porque precisava estudar. Vocês são minha inspiração para prosseguir. Amor sem fim.

Ao meu sogro e a minha sogra, João Batista e Vânia, por todo carinho, pela presença constante, por fazerem parte da minha rede de apoio. Obrigada pela disposição constante. Vocês são maravilhosos.

Aos meus avós e a todos os meus familiares por serem tão presentes, amigos e parceiros. Família, o bem preciso que temos.

À professora e amiga Dulce pelo apoio, encorajamento e por acreditar no meu sonho. Você também faz parte dessa vitória.

A todos os amigos do mestrado que tornaram nossas terças e quartas mais leves, pelos laços que construímos e que levaremos por toda a vida. Essa turma ficará para sempre guardada em meu coração.

A todos os professores do mestrado MPOIC da Universidade Candido Mendes (UCAM) por nos passarem seus ricos conhecimentos.

A todos os que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Meu agradecimento e profundo respeito a todos.

*“O homem pode tanto quanto sabe”
(Francis Bacon).*

RESUMO

USO DE TECNOLOGIAS DIGITAIS NO APOIO AO DIAGNÓSTICO PRECOCE DO CÂNCER DE MAMA

Tendo como foco o diagnóstico precoce do câncer de mama, o que mais mata mulheres no planeta, esta dissertação objetivou demonstrar que a utilização das tecnologias digitais (processamento de imagens e inteligência artificial) funciona como suporte no diagnóstico precoce do câncer de mama, auxilia o médico na análise das imagens a fim de elaborar o diagnóstico. Assim, detectada precocemente a doença, aumentam as chances de um tratamento menos traumático e muitas vezes a cura da paciente. A estrutura do desenvolvimento da pesquisa se compõe de três artigos: a) Análise bibliométrica de produções nacional e internacional sobre os temas “microcalcificação, mamografia, câncer de mama, inteligência artificial e Pacs” na base Scopus; b) Radiologia vinculada às novas tecnologias: suporte diagnóstico e terapêutico com técnica de Thresholding; c) Segmentação automática de áreas de calcificações em imagens digitais de mama – a partir dos quais se chegou à conclusão da eficácia do processamento de imagens na identificação de região suspeita no exame mamográfico quando se integram a este as técnicas de segmentação inteligente a fim de diagnosticar o câncer de mama. Por certo, há carência de estudos neste âmbito para que o aperfeiçoamento da técnica possa mitigar ainda mais os danos da doença.

PALAVRAS-CHAVE: Tecnologias digitais. Processamento de imagens. Câncer de mama. Diagnóstico. Prevenção.

ABSTRACT

USE OF DIGITAL TECHNOLOGIES IN SUPPORT OF EARLY BREAST CANCER DIAGNOSIS

Focusing on the early diagnosis of breast cancer, which kills most women on the planet, this dissertation aimed to demonstrate that the use of digital technologies (image processing and artificial intelligence) works as a support in the early diagnosis of breast cancer. doctor in the analysis of the images in order to make the diagnosis. Thus, early detection of the disease increases the chances of less traumatic treatment and often cure of the patient. The research development structure consists of three articles: a) Bibliometric analysis of national and international productions on the themes “microcalcification, mammography, breast cancer, artificial intelligence and Pacs” in the Scopus database; b) Radiology linked to new technologies: diagnostic and therapeutic support with Thresholding technique; c) Automatic segmentation of areas of calcifications in digital breast images - from which the effectiveness of image processing in identifying suspicious regions on mammographic examination has been concluded when intelligent segmentation techniques are integrated to the diagnosis. Breast cancer. Certainly, studies in this area are lacking so that the improvement of the technique can further mitigate the damage of the disease.

KEYWORDS: Digital technologies. Image processing. Breast cancer. Diagnosis. Prevention.

LISTA DE FIGURAS

Figura 1– Número de publicações a respeito dos temas “microcalcificação, mamografia, câncer de mama, inteligência artificial e Pacs” por ano.....	30
Figura 2– Países que mais publicaram sobre o tema entre 1991 e 2017.....	31
Figura 3– Universidades em todo o mundo que mais publicaram sobre o tema entre 1991 e 2017.	32
Figura 4– Principais autores que publicaram sobre o tema no período avaliado.	33
Figura 5– Principais áreas que publicaram sobre o tema no período avaliado.	34
Figura 6– Principais periódicos que publicaram sobre o tema no período investigado.	35
Figura 7– Descritores de maior evidência nos artigos encontrados.	36
Figura 8– Histogramas: 8a) Imagem clara; 8b) imagem escura.	48
Figura 9– 9a) Mama seio esquerdo; 9b) Recorte da área de interesse.	49
Figura 10– 10a) Histograma da Figura 10b; 10b) Limiar com valor $L = 200$	49
Figura 11– 11a) Mama seio direito; 11b) Recorte da área de interesse.....	50
Figura 12– 12a) Histograma da Figura 12b; 12b) Limiar com valor $L = 61$	50
Figura 13– 13a) Imagem Original; 13b) Recorte da área de interesse.....	51
Figura 14– 14a) Histograma da Figura 13b; 14b) Limiar com valor $L = 171$	51
Figura 15– Mamografia	61
Figura 16– Escala de cinza	61
Figura 17 – Imagem Binária	62
Figura 18– Modelo de um neurônio artificial.....	63
Figura 19– Modelo de uma Floresta Aleatória.....	63
Figura 20– Vetor de atributos de uma imagem aplicado a um algoritmo	64
Figura 21– Tela de parâmetros por Segmentação com Random Forest.....	65
Figura 22- à esquerda: Imagem Original; à direita: Imagem com as classes	66
Figura 23- à esquerda: Predição utilizando Random Forest; à direita: Mapa de probabilidade Random Forest	66
Figura 24– à esquerda: Predição utilizando Rede Neural; à direita: Mapa de probabilidade Rede Neural.....	67

LISTA DE ABREVIATURA E SIGLAS

BI-RADS Breast Imaging – Reporting and Data System
CIPHAR-10 Canadian Institute for Advanced Research
DDSM Dados Digital para Mamografia de Triagem
DICOM Digital Imaging and Communications in Medicine
EUA Estados Unidos da América
IA Imagem artificial
IRM Imagens de ressonância magnética
INCA Instituto Nacional do Câncer
OMS Organização Mundial de Saúde
PACS Picture Archiving and Communication System
Pixel Picture Element
TCIA The Cancer Imaging Archive

SUMÁRIO

1	INTRODUÇÃO	13
1.1	CONTEXTUALIZAÇÃO	13
1.2	OBJETIVOS	15
1.2.1	Objetivos gerais	15
1.2.2	Objetivos específicos	15
1.3	JUSTIFICATIVA	16
1.4	DELIMITAÇÃO DO PROBLEMA	17
1.5	ESTRUTURAÇÃO DO TRABALHO	17
2	METODOLOGIA	19
2.1	CLASSIFICAÇÃO DA PESQUISA	19
2.2	PROCEDIMENTOS TÉCNICOS E TÉCNICAS DE ANÁLISE	21
2.3	TRABALHOS CORRELATOS	21
3	ARTIGO A – ANÁLISE BIBLIOMÉTRICA DE PRODUÇÕES NACIONAL E INTERNACIONAL SOBRE OS TEMAS “MICROCALCIFICAÇÃO, MAMOGRAFIA, CÂNCER DE MAMA, INTELIGÊNCIA ARTIFICIAL E PACS” NA BASE SCOPUS	23
3.1	RESUMOS	23
3.1.1	Resumo em português	23
3.1.2	Abstract	24
3.2	INTRODUÇÃO	24
3.3	REVISÃO DE LITERATURA	25
3.3.1	Câncer de mama	25
3.3.2	Bibliometria	28
3.4	METODOLOGIA	29
3.5	RESULTADOS E DISCUSSÃO	30
3.6	CONCLUSÕES	36
3.7	REFERÊNCIAS	37
4	ARTIGO B – RADIOLOGIA VINCULADA ÀS NOVAS TECNOLOGIAS: SUPORTE DIAGNÓSTICO E TERAPÊUTICO COM TÉCNICA DE THRESHOLDING	39
4.1	RESUMOS	39
4.1.1	Resumo em português	39
4.1.2	Abstract	40

4.2	INTRODUÇÃO	40
4.3	REVISÃO DE LITERATURA	41
4.3.1	Novas tecnologias de imagem	41
4.3.2	Prevenção do câncer de mama favorecida pelas novas tecnologias	44
4.4	METODOLOGIA	47
4.4.1	Imagem digital	47
4.4.2	Histograma.....	47
4.4.3	Limiarização (<i>Thresholding</i>).....	48
4.5	RESULTADOS E DISCUSSÃO	48
4.6	CONCLUSÕES	51
4.7	REFERÊNCIAS.....	52
5	ARTIGO C – SEGMENTAÇÃO AUTOMÁTICA DE ÁREAS DE CALCIFICAÇÕES EM IMAGENS DIGITAIS DE MAMA	54
5.1	RESUMOS.....	54
5.1.1	Resumo em português.....	54
5.1.2	Abstract.....	55
5.2	INTRODUÇÃO	55
5.3	REVISÃO DE LITERATURA - CÂNCER DE MAMA	57
5.4	METODOLOGIA	60
5.4.1	Imagem digital	60
5.4.2	Imagem monocromática	61
5.4.3	Limiarização (<i>Thresholding</i>).....	62
5.4.4	Redes neurais	62
5.4.5	Random Forest (Floresta Aleatória).....	63
5.5	RESULTADOS E DISCUSSÃO	64
5.6	CONCLUSÕES	68
5.7	REFERÊNCIAS.....	68
6	CONSIDERAÇÕES FINAIS	70
	REFERÊNCIAS.....	72

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

O câncer é uma doença crônica não transmissível, que afeta indistintamente pessoas de qualquer nacionalidade, raça, etnia, idade, classe social e pelo fato da sua imprevisão a recepção do diagnóstico desse mal é quase sempre alarmante para o paciente. Embora possa ser do desconhecimento do senso comum, o câncer é uma doença milenar, conhecida como tumor maligno, cujo termo – do grego karkíno (=caranguejo) – fora utilizado pela primeira vez, ainda na antiguidade, pelo grego Hipócrates (460-377 a.C.), considerado o pai da medicina, que definiu o mal como tumor duro, que, mesmo após extração, reaparecia (INCA, 2011).

Hoje, os casos de câncer vêm se tornando cada vez maior. O de mama, principalmente entre as mulheres, ocupa posição de destaque nesse ranking. Entretanto, se diagnosticado precocemente, tem grandes chances de cura. Nesse sentido, as tecnologias digitais vêm avançando em ritmo cada vez mais acelerado para apoiar o diagnóstico precoce da doença (PULTZ, 2015).

De acordo com dados do Instituto Nacional do Câncer (INCA, 2018), este é o tipo de câncer mais comum entre as mulheres no Brasil e no mundo, depois do câncer de pele não melanoma, respondendo por cerca de 28% dos casos novos a cada ano. Geograficamente, sem considerar os tumores de pele não melanoma, esse tipo de câncer é o mais frequente nas mulheres das Regiões Sul, Sudeste, Centro-Oeste e Nordeste. Pode acometer homens também, porém estes

representam apenas cerca de 1% dos casos. Existem vários tipos de câncer de mama, sendo muitas vezes impossível prever sua evolução, já que há casos em que a evolução do quadro ocorre lentamente, em outros não. A estimativa para 2018, ainda de acordo com o INCA, era de 59.700 novos casos. É preciso considerar que o diagnóstico tardio da doença aumenta ainda mais o índice de mortes entre os pacientes deste tipo de câncer.

Tal aumento da incidência de câncer de mama na atualidade é certamente cada vez mais alarmante. Segundo dados da Organização Mundial de Saúde (OMS), em 2018 foi estimado que 627.000 mulheres tenham morrido em decorrência deste tipo de câncer, o que corresponde aproximadamente a 15% de todas as mortes ocasionadas por câncer entre as mulheres (ACR, 2016).

A análise clínica do câncer de mama comumente é realizada a partir do uso de diagnóstico por imagem e um dos exames mais utilizados, considerado na verdade o principal para a detecção precoce deste tipo de câncer, é a mamografia – um exame que permite visualizar o tecido mamário a partir do rastreamento por imagem com a utilização de raios-X. A recomendação oficial da OMS é que mulheres acima dos 40 anos, ou consideradas dentro do grupo de alto risco de câncer mamário, façam o exame anualmente. Portanto, percebe-se que a mamografia é um importante instrumento de controle do câncer de mama (ACR, 2016).

Ainda para auxiliar o diagnóstico precoce do câncer de mama, em 1993, foi desenvolvido o sistema de Laudos e Registros de Dados de Imagens da Mama – BI-RADS (em inglês, Breast Imaging Reporting and Data System), com o objetivo de padronizar os relatórios mamográficos e diminuir as divergências na interpretação do exame, que muitas vezes se apresentam como subjetivas. O sistema também visa facilitar o controle de resultados (ACR, 2016).

Como já foi exposto, quanto mais precoce e correto o diagnóstico do câncer mamário, melhor tratadas serão as mulheres portadoras dessa doença, porém é um processo complexo, que depende, dentre outros fatores, do raciocínio e experiência de profissionais da área médica. Quando a mamografia é interpretada erroneamente pode ocorrer superdiagnósticos, gerando procedimentos desnecessários, além de efeitos psicológicos danosos às pacientes, tais como estresse e preocupações com a possível evolução da doença; por sua vez, resultados falso-negativos trazem uma segurança inexistente, levando à ausência de tratamento adequado. Nesse sentido, o uso de técnicas computacionais é essencial para auxiliar os profissionais da área

médica no diagnóstico dessa doença, tornando assim a prevenção e o diagnóstico precoce ainda mais eficazes no contexto atual (CALAS; GUTFILEN; PEREIRA, 2012).

1.2 OBJETIVOS

1.2.1 Objetivos gerais

O presente trabalho teve como objetivo geral utilizar tecnologias digitais (processamento de imagens e inteligência artificial) para auxiliar no diagnóstico precoce do câncer de mama, ajudando o profissional de medicina na análise das imagens, facilitando o diagnóstico precoce e aumentando a chances de cura dessa doença.

1.2.2 Objetivos específicos

Os objetivos específicos da dissertação são:

- Utilizar imagens de mamografia com calcificações (macro e micro) de um repositório de imagens médicas gratuito, o The Cancer Imaging Archive (TCIA) – serviço que hospeda um grande arquivo de imagens médicas de câncer acessíveis para download público, no formato de Digital Imaging and Communications in Medicine (DICOM) –, fazendo uso apenas de imagens, sem idade, nome ou qualquer outro tipo de identificação da paciente, portanto, sem necessidade de consentimento legal para uso das imagens. Os dados de suporte relacionados às imagens, como resultados dos pacientes, detalhes do tratamento, genômica, patologia e análises especializadas também são fornecidos quando disponíveis. Pode ser acessado pelo link: <https://www.cancerimagingarchive.net/primary-data/>
- Segmentar as imagens;
- Classificar manualmente as calcificações e utilizar as técnicas de machine learning: Redes Neurais e Floresta Aleatória;

- Gerar resultados compatíveis com o profissional da área para o apoio ao diagnóstico do câncer de mama.

1.3 JUSTIFICATIVA

De acordo com as estatísticas do câncer publicadas em 2018 (SIEGEL; MILLER; JEMAL, 2018), estima-se que uma a cada cinco mamografias resultam em falsos negativos, ou seja, os efeitos psicológicos para as pacientes podem ser incalculáveis. Segundo os estudos de Hubbard *et al...* (2011), 7% a 9% de mulheres que realizam mamografias anualmente receberam um resultado falso-positivo recomendando a etapa de biopsia.

Quando há um exame de imagem da mama em curso pode haver diversos componentes passivos de análise pelo profissional, como a presença de nódulos, distorção arquitetural, calcificações, assimetria, linfonodo intramamário, lesão de pele e ducto único dilatado (D'ORSI *et al.*, 2013). A análise desses componentes ocorre a partir de uma orientação complexa e resulta em um agrupamento baseado em probabilidade com diferentes formas de ação que podem causar efeitos psicológicos negativos para a paciente, como o estresse, ou mesmo o adiamento de um tratamento que já deveria ter sido realizado. Portanto, a preocupação com essas questões no ambiente acadêmico-científico é uma forma de evolução para o conhecimento e, por isso, justifica esta pesquisa. Ademais, é interessante perceber a relação cada vez mais crescente entre biologia e medicina com inteligência artificial. O aprimoramento de novas técnicas é crucial para o avanço de ambos os campos científicos. Ao receber o auxílio de ferramentas de software direcionados especificamente para a área da saúde, o profissional de medicina tem em suas mãos meios mais eficazes de análise, evitando tratamentos desnecessários e possibilitando um tratamento mais assertivo ao paciente.

Estudos têm cada vez mais se preocupado em diminuir os possíveis erros nas análises dos exames. Como exemplo, tem-se a categorização de imagens por meio de redes convolucionais, um processo que possui diversas ferramentas para facilitar o desenvolvimento de soluções eficazes. O estudo que obteve a menor taxa de erro no desafio de classificação que utiliza o Canadian Institute for Advanced Research

(CIPHAR-10) foi o AutoAugment, que usou de técnicas de transformações das imagens da base de dados para aumentar a precisão de seu algoritmo (CUBUK *et al...*, 2018). Portanto, trazer à tona tal discussão é um modo de contribuir com o embate para o tratamento precoce do câncer mediado pelas novas tecnologias digitais da inteligência artificial.

1.4 DELIMITAÇÃO DO PROBLEMA

Este trabalho se circunscreve à utilização de tecnologias digitais (processamento de imagens e inteligência artificial) que cooperam com o diagnóstico precoce do câncer de mama, o que dá suporte ao profissional da medicina para analisar imagens e conseguir diagnosticar, desde a fase inicial, o surgimento da doença, aumentando, sobremaneira, as chances de curá-la.

1.5 ESTRUTURAÇÃO DO TRABALHO

Além das partes pré-textuais e pós-textuais desta Introdução geral (Capítulo 1 – parte que contém contextualização, objetivos, justificativa, delimitação do problema e a estrutura da pesquisa) e Conclusão (Capítulo 6) desta dissertação, o seu desenvolvimento se estrutura em mais sete capítulos, sobre os quais se passa a descrevê-los sumariamente.

O Capítulo 2 é acerca da metodologia da pesquisa: classificação, etapas de seu desenvolvimento e apresentação de trabalhos relacionados.

O Capítulo 3 é pertinente ao Artigo A (“Análise bibliométrica de produções nacional e internacional sobre os temas “microcalcificação, mamografia, câncer de mama, inteligência artificial e pacs” na base scopus”).

O Capítulo 4 apresenta o Artigo B (“Radiologia vinculada às novas tecnologias: suporte diagnóstico e terapêutico com técnica de thresholding”).

O Capítulo 5 é referente ao Artigo C (“Segmentação automática de áreas de calcificações em imagens digitais de mama”).

O Capítulo 6 expõe as conclusões finais do pesquisador com o fito de averiguar até que ponto o objetivo proposto foi alcançado.

O trabalho se encerra com as Referências gerais (Capítulo 7) utilizadas nos Capítulos 1, o da introdução geral e o 2, referente à metodologia.

2 METODOLOGIA

Além da classificação da pesquisa conforme as escolhas metodológicas, este capítulo descreve os procedimentos técnicos e técnicas de análise e, por fim, apresenta trabalhos correlatos.

2.1 CLASSIFICAÇÃO DA PESQUISA

Para a classificação da pesquisa, o teórico-metodológico consultado foi Maxwell Ferreira de Oliveira, professor da Universidade Federal de Goiás com dedicação exclusiva, e autor da obra “Metodologia científica: um manual para a realização de pesquisas em Administração”.

Os critérios de classificação de uma pesquisa são vários. Podem se dar quanto aos objetivos (descritiva, exploratória, explicativa, exploratório-descritiva), à natureza da pesquisa (qualitativa, quantitativa, quali-quantitativa), à escolha do objeto de estudo (estudos de caso único, estudo de casos múltiplos, amostragens não probabilísticas, estudo censitário), à técnica de coleta de dados (entrevista, questionário, observação, pesquisa documental, pesquisa bibliográfica, triangulação, pesquisa-ação, experimento) e à técnica de análise de dados (análise de conteúdo, estatística descritiva, estatística multivariada, triangulação na análise) (OLIVEIRA, 2011, p. 19).

Na dissertação em apreço, em relação aos objetivos, esta pesquisa se classifica como exploratória. Isso significa uma busca de proximidade com o fenômeno em foco que nem sempre se preocupa com formulação de hipóteses, mas sim com o conhecimento acerca dos fatos pela análise de diversos aspectos relativos ao fenômeno, para desenvolver, desvendar e recompor conceitos e princípios, e então se chegar à reformulação mais definida de problemas e hipóteses com vista a futuras pesquisas. E assim se obtém um panorama aproximativo do que é pesquisado (OLIVEIRA, 2011, p. 20).

Quanto à natureza da pesquisa, trata-se de um misto de duas espécies: qualitativa e quantitativa (quali-quantitativa). É qualitativa porque vai além das aparências procurando significado nos dados e, a partir daí explicar motivações, fontes, relações e mudanças, além de quaisquer percepções de individualidade e de significados diversificados. É quantitativa por se caracterizar pela quantificação (como no Artigo A, em que se aplicou a bibliometria) e por generalizar resultados. Assim, a pesquisa qualitativa se presta para explicar o que se coletou pela quantitativa (OLIVEIRA, 2011, p. 24-26).

No que se refere à escolha do objeto de estudo, trata-se de um estudo de caso único por se caracterizar como uma investigação profunda e exaustiva do objeto (diagnóstico de câncer de mama através das tecnologias da imagem), ampla e pormenorizada no conhecimento da realidade do objeto pesquisado em seu contexto real, cujos dados vão se convergindo para uma forma de triângulo, que facilitará “o desenvolvimento prévio de proposições teóricas para conduzir a coleta e a análise de dados” (YIN apud OLIVEIRA, 2011, p. 28).

A forma de triângulo, isto é, a triangulação é a técnica de coleta de dados utilizada. Para se coletar os dados um só artigo não seria o bastante como fonte de informações; daí a opção por três que trabalham com o mesmo objeto de estudo, cercando-o com maior abrangência sua descrição explicação e compreensão (OLIVEIRA, 2011, p. 42).

Também com relação à técnica de análise de dados, o critério utilizado foi o da triangulação na análise, até porque coleta e análise formam um todo e se retroalimentam durante o processo da pesquisa. Interpretando Creswell, Oliveira (2011, p. 53-54) explica ser essa técnica usada para “validar os dados por meio da comparação entre fontes de dados distintas, examinando-se a evidência dos dados e usando-os para construir uma justificativa para os temas”.

2.2 PROCEDIMENTOS TÉCNICOS E TÉCNICAS DE ANÁLISE

É de praxe nas pesquisas acadêmico-científicas iniciar-se o desenvolvimento com uma revisão bibliográfica tomando como referência o tema em foco. Realizou-se, desse modo, nos três artigos que compõem o desenvolvimento desta dissertação, um estudo teórico da área de câncer de mama e de uso do machine learning com imagens digitais. Numa primeira etapa (Artigo B), realizou-se um pré-processamento nas imagens com o uso de algumas operações entre elas: limiarização e binarização; após isso, procedeu-se à segmentação de imagens através de um processamento digital; Numa etapa em seguida (Artigo C), foram classificadas manualmente as calcificações, e não calcificações, descrevendo e diferenciando o que é do que não é (aprendizado supervisionado); para isso foram utilizadas tanto as Redes Neurais Artificiais quanto o algoritmo de Floresta Aleatória (Random Forest). Obteve como saída um mapa de probabilidade em tons de cinza no qual as maiores probabilidades na imagem emitiam a maior taxa de acerto dos métodos de machine learning acima descritos no encontro das calcificações. Para o desenvolvimento do protótipo do sistema, utilizou-se a linguagem Python (Artigo B) juntamente com o software FIJI (uma distribuição de software livre do projeto ImageJ), utilizando o kit de ferramentas de mineração de dados e aprendizado de máquina que contém o Waikato Environment for Knowledge Analysis (WEKA) (HALL *et al.*, 2009).

2.3 TRABALHOS CORRELATOS

Realizada a revisão bibliográfica sobre vários artigos que abordam o tema dos três artigos (A, B e C) do desenvolvimento desta dissertação, foi selecionado um trabalho correspondente a cada um desses artigos, dos quais se passa a descrevê-los sumariamente.

No Brasil, Karla Rangel Ribeiro, Rosalee Santos Crespo e Fernanda Castro Manhães (2016) utilizaram indicadores bibliométricos na base Scopus em pesquisa sobre evasão escolar relativa a julho de 2016, com o objetivo de identificar o maior

número e publicações nacionais e internacionais considerando autores, instituições, periódicos e áreas. Essa técnica metodológica de computação de produção científica configura-se como uma ciencitometria e se revela um modelo eficiente de mensuração e diagnóstico dos interesses de pesquisa e publicação referentes a certo tema, contribuindo, portanto, para o fornecimento de dados estatísticos referentes aos aspectos de interesse: evolução temporal, confluência geográfica de publicação, áreas de maior investimento e assim por diante.

Priscila do Carmo Santana, em 2010, apresentou ao Centro de Desenvolvimento de Tecnologia Nuclear (CDTN), sua pesquisa de mestrado, em que desenvolveu uma metodologia computacional a fim de comprovar que os mecanismos de análise computacional, pela sua precisão, praticidade e rapidez, podem melhorar a avaliação da qualidade de imagens, inclusive das mamográficas. Utilizou técnicas de processamento digital de imagens (PDI) para determinar características geométricas e radiométricas das imagens avaliadas. “Os parâmetros avaliados incluem resolução espacial, detalhes de alto contraste, limiar [threshold] de baixo contraste, detalhes lineares de baixo contraste, massas tumorais, índice de contraste e densidade óptica de fundo”. Comparou os resultados com os da Vigilância Sanitária de Minas Gerais apresentados nas avaliações visuais, daí depreendendo que a “metodologia automatizada se apresenta como uma promissora alternativa para redução ou supressão da subjetividade existente na metodologia de avaliação visual, atualmente em uso”.

Neste ano de 2009, Silva, Leal e Lima, publicaram sua pesquisa, que faz uma comparação entre 50 simulações realizadas para classificar nódulos de mama entre “modelos de rede neural de perceptrons de múltiplas camadas e máquina de vetores de suporte não linear”. Os resultados médios obtidos evidenciam que o algoritmo de máquina de vetor de suporte não linear se sobressai no desempenho ultrapassando 90,0%, relativo à acurácia no conjunto de teste, destacando-se, portanto, “quando comparado ao algoritmo de rede neural de perceptrons de múltiplas camadas proposto, com acurácia de $\approx 99,0\%$ e taxa de falso negativo de $\approx 2,0\%$ ”. Os autores concluem que “o modelo de rede neural apresentou desempenho inferior ao classificador de máquina de vetor de suporte não linear. Os resultados médios obtidos, com a aplicação dos modelos propostos, mostram-se promissores, na classificação do câncer de mama”.

3 ARTIGO A – ANÁLISE BIBLIOMÉTRICA DE PRODUÇÕES NACIONAL E INTERNACIONAL SOBRE OS TEMAS “MICROCALCIFICAÇÃO, MAMOGRAFIA, CÂNCER DE MAMA, INTELIGÊNCIA ARTIFICIAL E PACS” NA BASE SCOPUS

3.1 RESUMOS

3.1.1 Resumo em português

No mundo e no Brasil, dentre os vários tipos de câncer em mulheres, o câncer de mama é o segundo mais recorrente. O objetivo deste artigo é apresentar indicadores bibliométricos referentes aos temas “microcalcificação, mamografia, câncer de mama, inteligência artificial e PACS” na base Scopus. A busca foi realizada no dia 08 de agosto de 2018, sendo identificadas 28 publicações, e destas foram obtidas informações relacionadas aos principais países, autores, instituições, periódicos, áreas e ano de publicação. Como resultados, verificou-se que: a) o ano de 2006 foi o de maior publicação; b) os Estados Unidos se destacam com 11 artigos, sendo a Universidade de Chicago a que teve mais publicações (5); c) e a área em que mais se publicou foi a de Medicina (60,7%). Além disso, os periódicos com mais publicações foram aqueles relacionados à radiologia e de informática aplicada à área médica. De forma geral, conclui-se que ainda são poucos os trabalhos que abordam os temas estudados e que esses artigos são relativamente concentrados em instituições dos Estados Unidos.

Palavras-chave: Câncer de mama. Publicações científicas. Bibliometria. Inteligência artificial.

3.1.2 Abstract

In the world and in Brazil, among the various types of cancer in women, breast cancer is the second most recurrent. The objective of this article is to present bibliometric indicators referring to the topics "microcalcification, mammography, breast cancer, artificial intelligence and PACS" in the Scopus database. The search was carried out on August 08, 2018, with 28 publications identified, from which information related to the main countries, authors, institutions, periodicals, areas and year of publication were obtained. As a result, it was verified that: a) 2006 was the year of greatest publication; b) the United States stands out with 11 articles, with the University of Chicago having the most publications (5); c) and the area in which it was most published was Medicine (60.7%). In addition, the journals with more publications were those related to radiology and computer science applied to the medical field. In general, it is concluded that there are still few papers that address the subjects studied and that these articles are relatively concentrated in institutions of the United States.

Keywords: Breast câncer. Scientific publications. Bibliometrics. Artificial intelligence.

3.2 INTRODUÇÃO

Doença é um estado de anormalidade, de desequilíbrio do corpo e/ou da mente. Uma das doenças que mais atemorizam e carregam em si perplexidades sobre o amanhã é o câncer. Este, durante muito tempo, envolvido num reducionismo negativista e enredado na cultura das civilizações com grande intensidade, amedrontou muitos que sequer pronunciavam a palavra câncer. Hoje, com os avanços tecnológicos na medicina, essa mentalidade vem-se transformando, já que,

se o tratamento for iniciado precocemente, há chances de cura. O câncer de mama cujos sinais/sintomas são detectados logo no início propicia resultados efetivos com o tratamento precoce. Este é o tipo de câncer que ocupa o segundo lugar no índice de ocorrências desta doença no mundo, atrás apenas do de pele não melanoma (INCA, 2018).

Sendo assim, há por toda parte estudos empenhados em discutir a questão, esclarecendo pontos obscuros e apresentando novas formas de tratamento e enfrentamento da doença. É exatamente sobre essas publicações que este estudo incide, pois objetiva demonstrar indicadores bibliométricos (mecanismos de avaliação de produção científica) referentes ao câncer de mama na base Scopus.

Após esta introdução, faz-se uma revisão bibliográfica do tema câncer de mama em razão da relevância de conhecimentos a respeito de uma doença alarmante e sempre carecendo de investimento em pesquisas, em face da mortalidade e dos prejuízos físicos e psicológicos dela decorrentes sobre os pacientes. Em seguida, vem a descrição da metodologia, o tipo de análise e a apresentação das variáveis. Na sequência, apresentam-se os resultados e a discussão, para, finalmente, apontar as conclusões.

3.3 REVISÃO DE LITERATURA

3.3.1 Câncer de mama

Definido como tumor maligno, o câncer, diferente de uma doença individualizada, caracteriza-se como um conjunto de doenças que abrange mais de 200 patologias, cujas células anormais malignas, geralmente muito agressivas, crescem descontroladamente invadindo órgãos e tecidos adjacentes, podendo disseminar pelo corpo originando tumores em outras regiões. Neste caso de disseminação, tem-se a denominada metástase. Tais células antes de se tornarem cancerígenas eram normais; entretanto, deformaram-se ao sofrerem danos em seu funcionamento, em algum momento anterior, que as levaram a se reproduzirem

rápida e desordenadamente com um aumento de consumo de glicose (LESSA; MARENGONI, 2016).

A despeito de o câncer ser “[...] uma das doenças que mais causam temor na sociedade, por ter se tornado um estigma de mortalidade e dor” (ALMEIDA *et al.*, 2005, p. 118), ele não é uma doença que surgiu na atualidade.

Desde a Antiguidade, médicos extraíam mamas doentes, acentuando sofrimento e morte. Com o surgimento de anestésias mais eficazes e da assepsia, foi possível, no final do século XIX, executar a chamada mastectomia radical, que retirava toda a mama, musculatura peitoral e os linfonodos axilares. Essa intervenção foi amplamente aceita até a década de 1950, quando técnicas cirúrgicas conservadoras, que evitavam mutilação das pacientes, passaram a ser utilizadas (INCA, 2014, p. 21).

A partir dos anos 1980, o tratamento vem evoluindo consideravelmente gerando altos índices de cura. Há vários tipos dessa doença. O de mama – cujas publicações sobre ele é o foco deste estudo – é resultado do crescimento indiscriminado de células da mama com características anormais, resultantes de uma ou mais mutações ocorridas no material genético de uma das células, formando um tumor que pode acometer outros órgãos (INCA, 2014).

Em razão de ser uma doença com altas taxas de incidência e mortalidade entre as mulheres, as tecnologias vêm avançando substancialmente na resolução do enfrentamento da doença e possível cura. A mamografia, por exemplo, é um método de obtenção de imagens indicado para detectar tecidos mamários densos, localizar as calcificações suspeitas e orientar biópsias. Entretanto, “Alguns autores sugerem que a ultrassonografia pode ser uma ferramenta útil, principalmente para as mulheres com mamas densas ou com alto risco de câncer de mama, por ser capaz de diagnosticar alguns tumores não identificados de outra forma” (NASTRI; MARTINS; LENHARTE, 2011, p. 97).

O câncer de mama tem aumentado significativamente entre as mulheres jovens. Um elemento fundamental no êxito do tratamento dessa doença e na redução da taxa de mortalidade decorrente dela é a precocidade do diagnóstico, caso contrário as chances de cura esvaem-se drasticamente (LELES; MARENGONI, 2016).

A propósito, várias áreas da medicina têm utilizado a Inteligência Artificial (IA; “campo da ciência da computação que imita os processos de pensamento humano, a capacidade de aprendizagem e o armazenamento de conhecimento”) no

aprimoramento de diagnóstico, prognóstico e tratamento de doenças, inclusive o de câncer de mama, revelando resultados animadores para a saúde humana fornecendo diretrizes mais precisas e confiáveis (BRAGA *et al.*, 2018).

Alana de Moraes *et al.* (2010) esclarecem ser a IA uma diretriz oriunda da ciência da computação que indica várias técnicas e dispositivos no desdobramento de programas inteligentes, que podem tomar decisões semelhantes às humanas, ou seja, é a competência de programação para que o computador desempenhe tarefas tais quais o homem seja capaz de naturalmente efetuar-las. Recentemente, porém, com o advento do computador moderno, a IA ganha protagonismo como ciência integral, é utilizada na solução de problemas especiais em áreas específicas e apresenta metodologias próprias. Os programadores de jogos são impelidos a desenvolverem cada vez mais a IA, conforme as exigências do mercado, tais como manipulação de conceitos, uso de heurísticas (com vista à resolução de problemas em que os algoritmos conhecidos não são capazes de resolvê-los), representação do conhecimento, permissão de dados imprecisos (suporta dados imprecisos ou incompletos), permissão de múltiplas soluções, capacidade de aprender (DE MORAIS *et al.* 2010, p. 1-2).

Com esse avanço, a área da Saúde vem utilizando cada vez mais de novos mecanismos tecnológicos da computação: IA, Realidade Virtual, Multimídia e Internet

Afirmam Azevedo-Marques e Salomão:

A revolução tecnológica ocorrida nas últimas décadas tem provocado mudanças importantes em várias áreas do conhecimento. Na Radiologia, a utilização em grande escala de sistemas digitais tem gerado um volume de dados cada vez maior. A melhor solução para gerenciar essas imagens digitais está na adoção de um Sistema de Arquivamento e Distribuição de Imagens (PACS, do inglês Picture Archiving and Communication System) (AZEVEDO-MARQUES; SALOMÃO, 2009, p. 131).

Como se vê a potencialidade das ferramentas de IA, com suas imagens termográficas, na detecção do câncer (incluindo-se aí o de mama – parte do objeto deste estudo) e a baixo custo tem evidenciado um expediente promissor e que vêm revolucionando a medicina a tal ponto que se percebe vivenciar o futuro em vez do presente em face de tamanhos avanços que demandam do médico radiologista uma atualização constante, caso não queira fazer parte do passado.

3.3.2 Bibliometria

A publicação por Cole e Eales de uma análise estatística da disciplina anatomia comparada, no decorrer da história, marca, em 2017, o surgimento da bibliografia estatística, hoje denominada bibliometria. De início, para utilidade específica nas bibliotecas, fazia-se a medição estatística de monografias, em seguida abarcou também periódicos e artigos científicos. Hoje, avalia a produtividade de autores, realização de estudos de citações, índices de produção e propagação do conhecimento, acompanhamento da trajetória evolutiva de inúmeros setores científicos, referências de autoria, editoração e utilização dos resultados de pesquisas. Trata-se, portanto, de técnica quantitativa e estatística pela medição de índices – evidenciando-se daí a sua relevância e benefícios práticos tanto nas bibliotecas quanto na comunidade acadêmica e científica (COSTA *et al.*, 2012).

A bibliometria é a metodologia de análise desta pesquisa – método que se caracteriza pela pesquisa quantitativa como meio de mensuração do conhecimento científico e de suas publicações de determinado tema e em certas áreas com o fim de detectar “o estado da ciência e da tecnologia” e averiguar que país/instituição/autor produz mais em certa área. “Portanto, a bibliometria representa todos os estudos que tentam quantificar os processos de comunicação escrita fornecendo subsídios na formulação da política científica e tecnológica nas diferentes áreas do conhecimento” (SILVA, 2013, p. 4).

A bibliometria – domínio da ciencitometria – muito contribui no fornecimento de dados estatísticos sobre a evolução temporal, os *loci* geográficos de concentração de esforços em certas áreas e tantas outras informações. Para demonstrar os indicadores bibliométricos identificados nas publicações que tratam do tema ora proposto, tomou-se a base Scopus (referencial da Editora Elsevier). Segundo a Capes/MEC,

[...] a base Scopus é disponibilizada pelo Portal de Periódicos por meio do contrato Elsevier B. V/Scopus. Essa base indexa títulos acadêmicos revisados por pares, títulos de acesso livre, anais de conferências, publicações comerciais, séries de livros, páginas web de conteúdo científico (reunidos no Scirus) e patentes de escritórios. Dispõe de funcionalidades de apoio à análise de resultados (bibliometria) como identificação de autores e filiações, análise de citações, análise de publicações e índice H. Cobre as áreas de Ciências Biológicas, Ciências da Saúde, Ciências Físicas e

Ciências Sociais. Período de acesso desde 1823 até o presente. O acesso a esse conteúdo pode ser feito pela opção buscar base (BRASIL, 2019).

Os indicadores bibliométricos (de qualidade científica, de atividade científica, de impacto científico e de associações temáticas) são mecanismos de avaliação, os quais devem ser aplicados cuidadosamente em razão de características e limitações que podem apresentar. É essencial lembrar que os resultados desse tipo de investigação são transmitidos por comunicação formal e que as publicações retratam o volume da produção investigada. Tais trabalhos são fontes primárias que, indexadas nas bases de dados, tornam-se fontes secundárias – mecanismos que possibilitam, em qualquer área científica, a recuperação da informação publicada (COSTA *et al.*, 2012).

3.4 METODOLOGIA

O primeiro passo desta pesquisa foi estabelecer os descritores (microcalcificação, mamografia, câncer de Mama, Inteligência Artificial e Pacs), uma interface entre as áreas da medicina e da inteligência artificial para, em seguida, proceder à busca, no dia 8 de agosto de 2018, na base Scopus, sendo encontrados 28 artigos publicados no período de 1991 a 2017.

Foram pesquisados artigos que tivessem os descritores no título, no resumo ou nas palavras-chaves, sendo utilizada a seguinte expressão booleana:

```
(((((title-abs-key("radiology information systems" or "picture archiving" or "radiologic picture" or "pacs (radiology)") or ((title-abs-key(mammograph*)) and (title-abs-key(microcalcification or calcinosis or calcinosis or microcalcinoses)))) and (title-abs-key("artificial intelligence" or "computational intelligence" or "computer vision systems" or "machine intelligence")) and (title-abs-key("breast neoplasms" or "breast cancer" or "mammary cancer" or "breast tumor")) and (limit-to (doctype,"ar")) and (limit-to (srctype,"j"))))
```

Os dados obtidos, sempre quantitativos, são referentes a número anual de artigos no recorte temporal 1991 a 2017, país, instituição, autor, área e periódico.

Em seguida, partiu-se para a análise quantitativa (resultados e discussões) desses dados, conforme orienta o método bibliométrico.

3.5 RESULTADOS E DISCUSSÃO

Na Figura 1, é possível observar o número de publicações por ano.

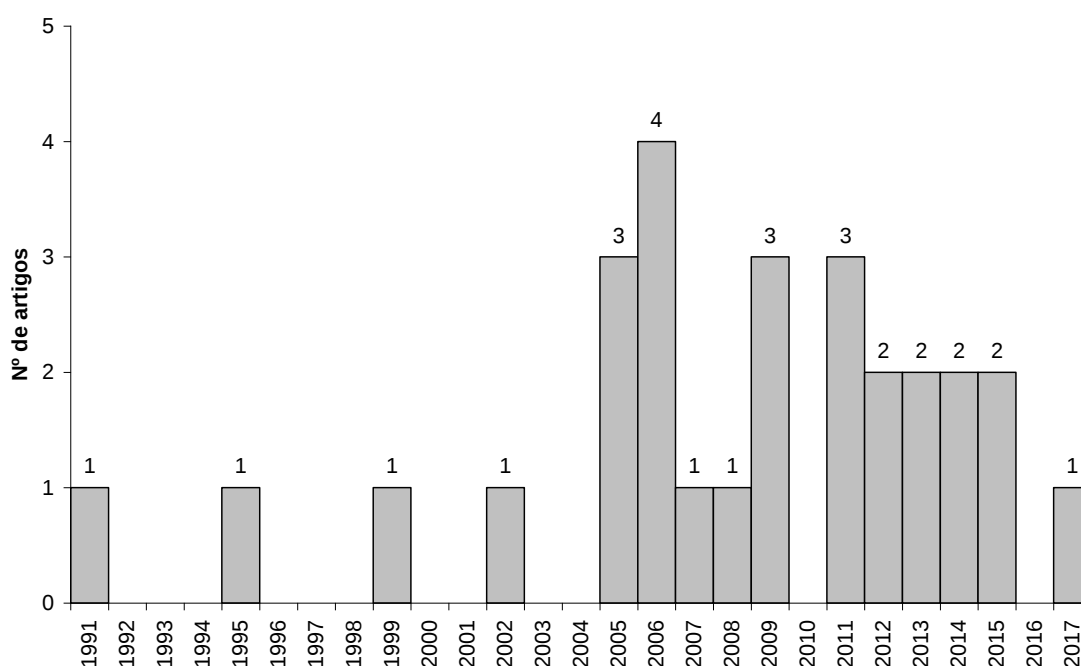


Figura 1– Número de publicações a respeito dos temas “microcalcificação, mamografia, câncer de mama, inteligência artificial e Pacs” por ano.

Fonte: Próprio autor

É possível verificar na Figura 1 que houve anos (1992, 1993, 1994, 1996, 1997, 1998, 2000, 2001, 2003, 2004, 2010 e 2016) em que nada se publicou com os referidos descritores e houve anos (1991, 1995, 1999, 2002, 2007, 2008 e 2017) em que se detectou apenas uma publicação. De 2012 a 2015, foram duas publicações em cada um desses anos. Em 2005, 2009 e 2011, foram três publicações de cada ano. O ano de 2006 foi o que apresentou maior publicação: quatro. Pode-se também afirmar que as publicações foram mais frequentes de 2005 em diante, ainda que em 2010 e 2016 não houvesse publicação.

Na Figura 2 é possível identificar os países que mais publicaram sobre o assunto.

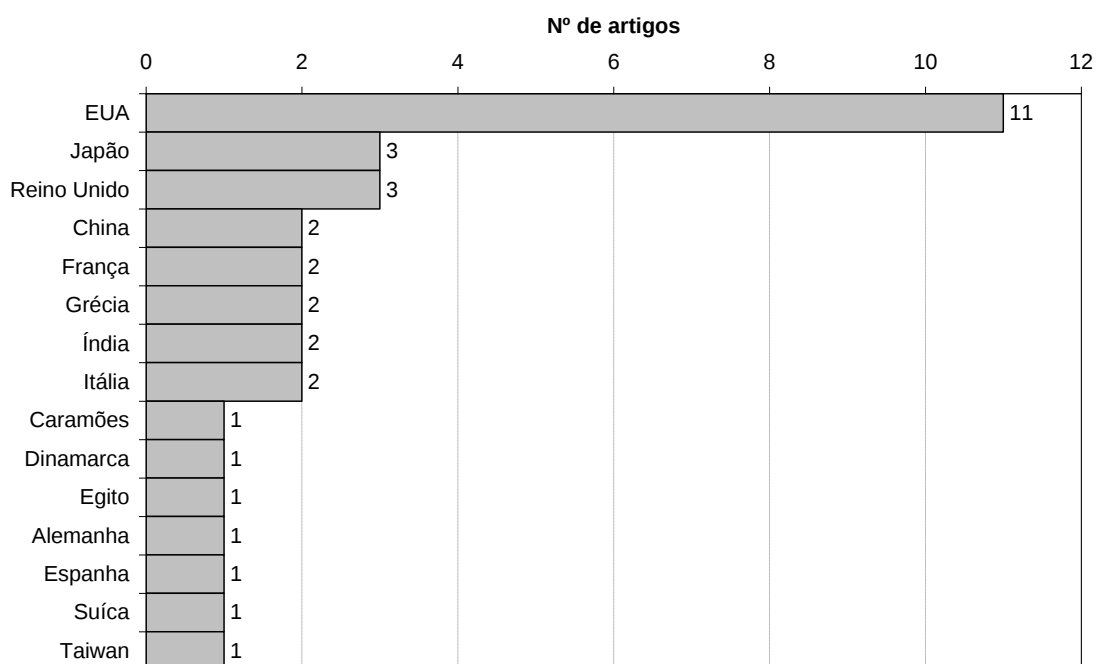


Figura 2– Países que mais publicaram sobre o tema entre 1991 e 2017.

Fonte: Próprio autor

Observa-se pela Figura 2 que, dentre essas publicações, não se encontrou nenhuma brasileira ou mesmo latino-americana. O país que mais produziu foram os Estados Unidos da América (EUA), com 11, seguido de Japão e Reino Unido com três cada um.

Já China, França, Grécia, Índia e Itália, com duas cada um. E, com apenas uma publicação cada um: Camarões, Dinamarca, Egito, Alemanha, Espanha, Suíça, Taiwan. Infere-se desses dados que os EUA dominam a quantidade de publicação científica em periódicos, em nível mundial.

Na Figura 3 são demonstradas as universidades que mais publicaram sobre o tema.

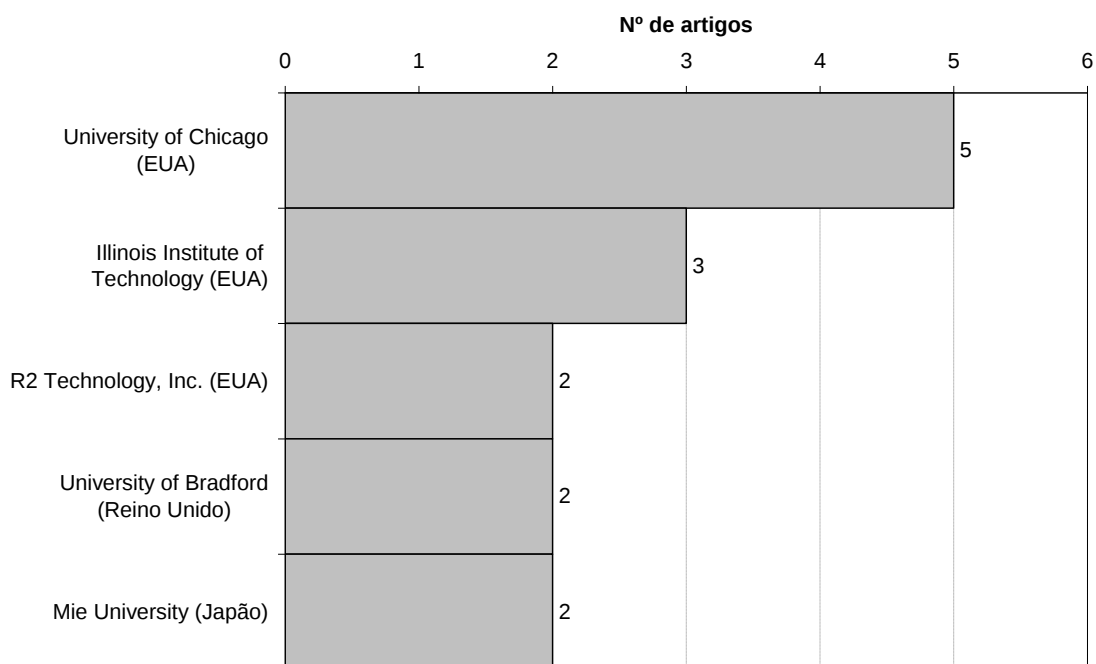


Figura 3– Universidades em todo o mundo que mais publicaram sobre o tema entre 1991 e 2017.

Fonte: Próprio autor

Verificou-se grande destaque de publicações nas instituições estadunidenses: University of Chicago (EUA), com quatro (4) publicações; Illinois Institute of Technology (EUA), com três (3); R2 Technology Inc., com duas (2). Também com duas (2) publicações estão a University of Bradford (Reino Unido) e a Mie University (Japão).

Esses dados corroboram o que já se comentou sobre o destaque dos EUA em produção científica internacional na temática em foco e em periódicos no recorte temporal estabelecido para esta pesquisa.

Na Figura 4, é possível avaliar os autores que mais escreveram sobre o tema.

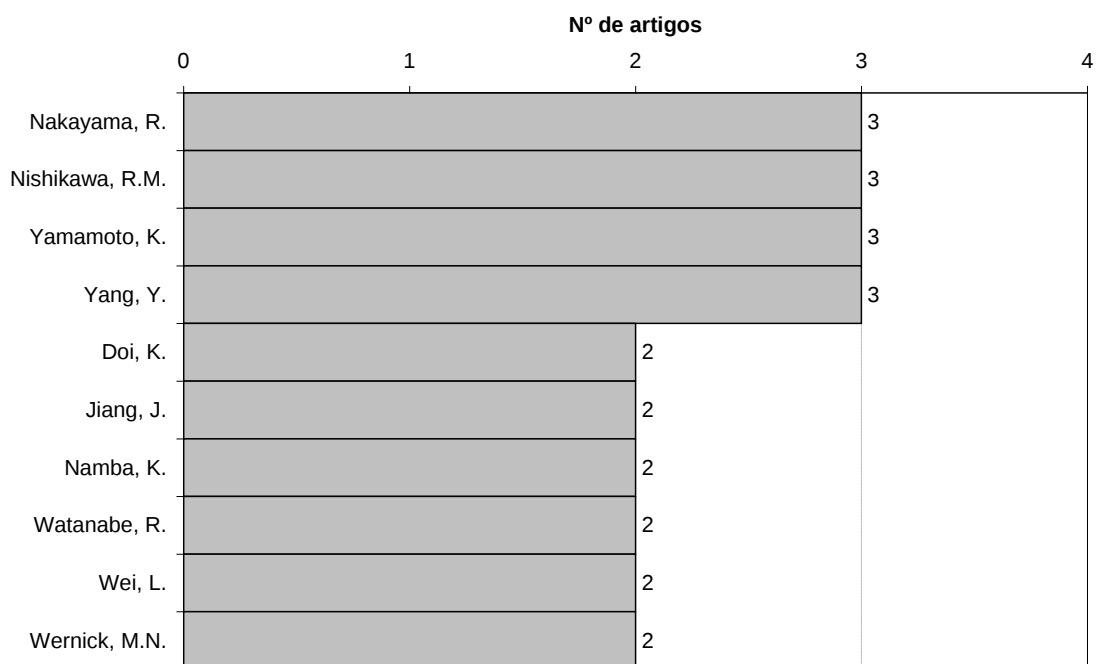


Figura 4– Principais autores que publicaram sobre o tema no período avaliado.

Fonte: Próprio autor

Os autores que mais produziram – três (3) pesquisas no período referido – foram Nakayama, R.; Nishikawa, R. M.; Yamamoto, K.; e Yang, Y.

Os demais autores (DOI, K.; JIANG, J.; NAMBA, K.; WATANABE, R.; e WEI, L.) tiveram duas (2) produções cada um.

Percebe-se claramente que as produções se concentram em certos autores e que até mesmo estes que sobressaíram também publicaram pouco nesse período estabelecido desta pesquisa, a despeito de a IA vir caminhando a passos largos e atravessando fronteiras entre conhecimentos que, num passado não muito longe, se fechavam em seus delimitados espaços.

Na Figura 5 são descritas as áreas que mais publicaram sobre o assunto no mundo.

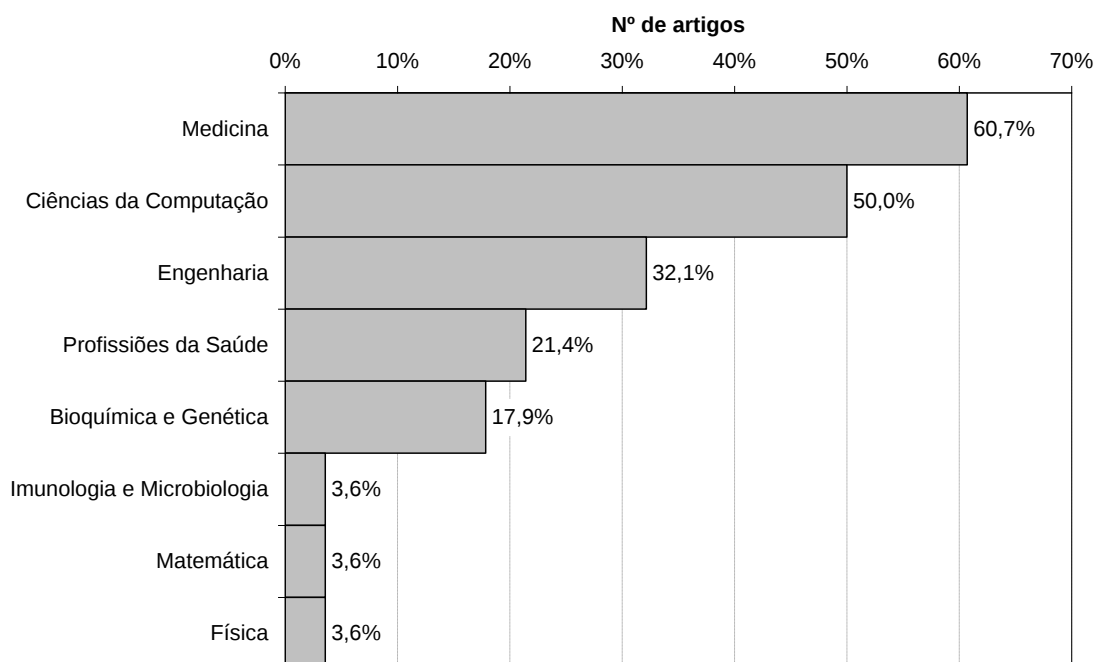


Figura 5– Principais áreas que publicaram sobre o tema no período avaliado.

Fonte: Próprio autor

Verifica-se que são poucos os autores que se atêm a produzir cientificamente estabelecendo a interface entre câncer de mama e IA.

A área de maior produção foi a Medicina (60,7%), seguida em ordem decrescente por Ciência da Computação (50%), Engenharia (32,1%), Profissões da Saúde (21,4%), Bioquímica e Genética (17,9%), Imunologia e Microbiologia (3,6%), Matemática (3,6%) e Física (3,6%).

O destaque da Medicina seguido das Ciências da Comunicação evidencia que a IA tem se empenhado no campo da saúde: “seja pela adoção de sistemas de apoio à decisão clínica, seja pelo uso integrado de novas tecnologias, incluindo as tecnologias vestíveis/corporais (wearable devices), seja pelo armazenamento de grandes volumes de dados de saúde de pacientes e da população” (LOBO, 2017, p. 185).

Na Figura 6 pode-se observar os periódicos que mais publicam sobre o tema.

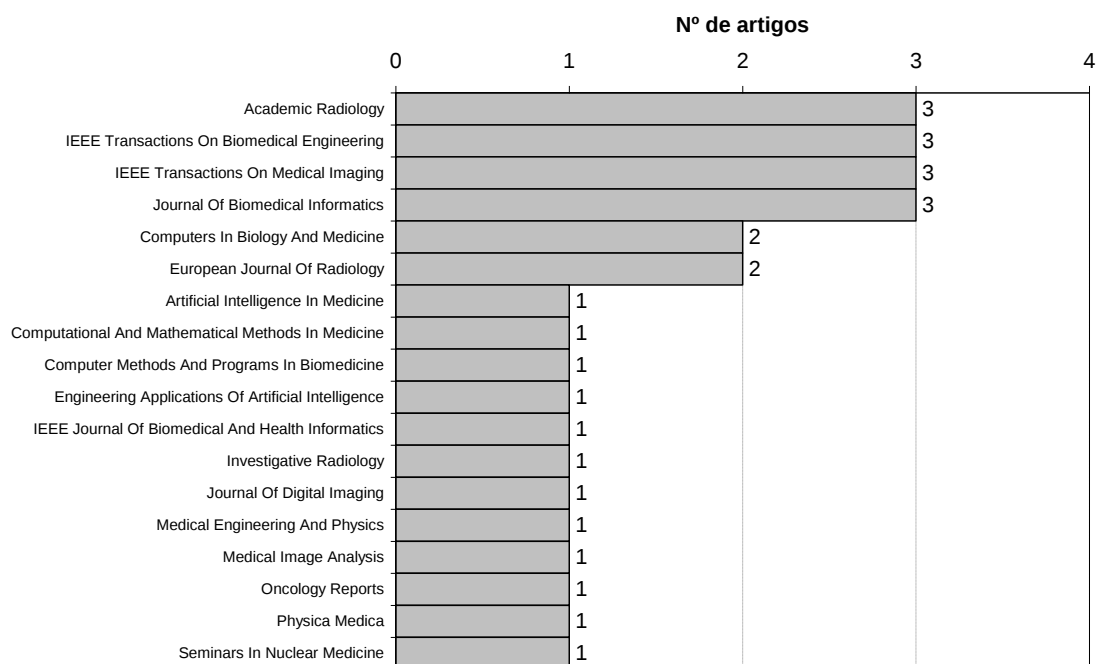


Figura 6– Principais periódicos que publicaram sobre o tema no período investigado.

Fonte: Próprio autor

Os periódicos Academic Radiology, IEEE Transactions On Biomedical Engineering, IEEE Transactions On Medical Imaging e Journal of Biomedical Informatics trazem três (3) dessas publicações.

Já o Computers in Biology and Medicine e European Journal of Radiology comportam duas (2) cada um.

E com apenas uma (1) publicação por periódico, citam-se: Artificial Intelligence in Medicine, Computational and Mathematical Methods in Medicine, Computer Methods and Programs in Biom Medicine, Engineering Applications of Artificial Intelligence, IEE Journal of Biomedical and Heath Informatics, Investigative Radiology, Journal of Digital Imaging, Medical Image Analysis, Oncology Reports, Physica Medica e Seminars in Nuclear Medicine.

Observa-se que todas as publicações foram feitas em periódicos da área de radiologia ou associam às áreas engenharia/informática aplicadas à medicina; são periódicos de temáticas sempre envolvendo a IA em sua transversalidade como, por exemplo, as imagens de ressonância magnética (IRM) da radiologia. A “de corpo inteiro está sendo cada vez mais utilizada no diagnóstico e acompanhamento de tumores como dos múltiplos mielomas e linfomas malignos” (MADUREIRA *et al.*, 2010, p. 18). Enfim, os periódicos voltados à medicina divulgam cada vez mais o que

“Essas tecnologias, consideravelmente menos invasivas, passaram a mostrar os tecidos biológicos de modo direto e com excelente resolução espacial da anatomia” (MADUREIRA *et al.*, 2010, p. 13).

Na Figura 7 é possível perceber os descritores de maior relevância entre as publicações.

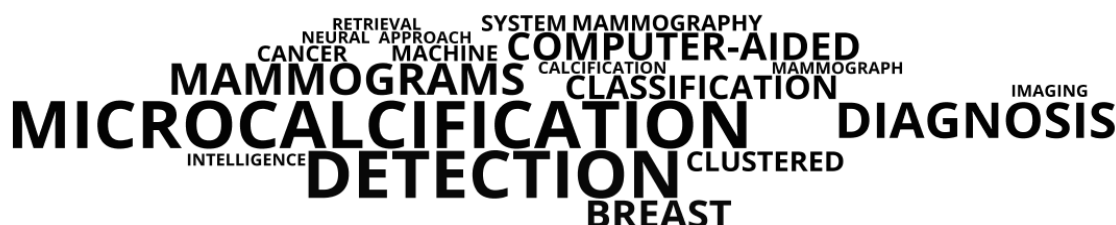


Figura 7– Descritores de maior evidência nos artigos encontrados.

Fonte: Próprio autor

Por fim, são apresentados os descritores de maior importância, a partir do maior destaque gráfico. Na ordem estão: microcalcification, detection, diagnosos, mammograms, computer-aided, Breast, classofication, clustered, system mammography, cancer, machine, mammograph, imaging, calcification, intelligence, retrieval e neural approach.

3.6 CONCLUSÕES

Como resultados, verificou-se o seguinte: o ano de 2006 foi o de maior publicação; os Estados Unidos se destacam com 11 artigos, sendo a Universidade de Chicago a que teve mais publicações (5), e a área em que mais se publicou foi a de Medicina (60,7%). Outrossim, os periódicos com o quantitativo maior de publicações se relacionam à radiologia e à informática aplicada à área médica, por certo, em decorrência da notabilidade do progresso das imagens de ressonância magnética com o seu uso intenso nos diagnósticos realizados por tecnologias cada vez menos invasivas e de resultados mais precisos.

Infere-se desta pesquisa, por fim, o reduzido número de trabalhos relacionados ao tema ora tratado. Além dessa restrição, as publicações se concentram relativamente em instituições estadunidenses.

3.7 REFERÊNCIAS

ALMEIDA, V. L. de *et al.* Câncer e agentes antineoplásicos ciclo-celular específicos e ciclo-celular não específicos que interagem com o DNA: uma introdução. **Química Nova**, São Paulo, v. 28, n. 1, p. 118-29, 2005.

AZEVEDO-MARQUES, P. M. de; SALOMÃO, S. C. PACS: Sistemas de Arquivamento e Distribuição de Imagens. **Revista Brasileira de Física Médica**, São Paulo, v. 3, n. 1, p. 131-139, 2009.

BORCHARTT, T. **Análise de imagens termográficas para a classificação de alterações na mama**. Niterói: UFF, 2013.

BRASIL. CAPES/MEC. Acervo. **Portal de periódicos**. Disponível em: <https://bit.ly/2DAvS8f>. Acesso em: 13 fev. 2019.

BRAGA, A. V. *et al.* Inteligência Artificial na medicina. **Cipeex**, Anápolis, v. 2, p. 937-941, 2018.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. **A mulher e o câncer de mama no Brasil**. Rio de Janeiro, INCA, 2014.

INCA. Instituto Nacional de Câncer José Alencar Gomes da Silva. **Conceito e Magnitude do câncer de mama**. 19 nov. 2018. Disponível em: <https://bit.ly/2VEsRit>. Acesso em: 28 abr. 2019.

LESSA, V.; MARENGONI, M. Applying artificial neural network for the classification of breast cancer using infrared thermographic images. *In: INTERNATIONAL CONFERENCE ON COMPUTER VISION AND GRAPHICS*, 2016, Warsaw, Poland. **Proceedings** [...]. [s.l.]: [s.n.], 2016, p. 429-438.

LOBO, L. C. Inteligência Artificial e Medicina. **Revista Brasileira de Educação Médica**, Brasília, ano 41, n. 2, p. 185-193, 2017.

MADUREIRA, L. C. A. *et al.* Importância da imagem por ressonância magnética nos estudos dos processos interativos dos órgãos e sistemas. **Revista de Ciências Médicas e Biológicas**, Salvador, n. 9, p. 13-19, 2010. Supl.1.

MORAIS, A. M. de *et al.* **Tomada de Decisão aplicada à Inteligência Artificial em Serious Games voltados para Saúde**. 2010. Disponível em: <https://docplayer.com.br/1808217-Tomada-de-decisao-aplicada-a-inteligencia-artificial-em-serious-games-voltados-para-saude.html>. Acesso em: 02 set. 2019.

NASTRI, C. O.; MARTINS, W.de P.; LENHARTE, R. de J. Ultrassonografia no rastreamento do câncer de mama. **Femina**, São Paulo, v. 39, n. 2, p. 97-102, fev. 2011.

SILVA, R. C. da. Avaliação da informação científica em Bibliometria aplicada às Ciências da Saúde. *IN*: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 25., Florianópolis, 2013, SC. **Anais [...]**. Florianópolis: CBBB, 2013, p. 1-16.

4 ARTIGO B – RADIOLOGIA VINCULADA ÀS NOVAS TECNOLOGIAS: SUPORTE DIAGNÓSTICO E TERAPÊUTICO COM TÉCNICA DE THRESHOLDING

4.1 RESUMOS

4.1.1 Resumo em português

O papel da radiologia tem sido preponderante nas transformações e evolução das trajetórias da medicina no mundo. As imagens radiológicas viabilizam dados relevantes para a tomada de decisões de um diagnóstico e subsequente tratamento/acompanhamento de um paciente. No caso de câncer de mama, elas podem representar um passo importante, mas não definitivo. Nesse sentido, este artigo tem por objetivo demonstrar a importância da radiologia com técnicas de processamento digital de imagem em especial a limiarização no diagnóstico do câncer de mama. Para tanto, primeiramente se discorre sobre os avanços das tecnologias da imagem, que têm seu primeiro passo na primeira radiografia, para só então discutir sobre a prevenção do câncer mamário possibilitada por tais avanços. Infere-se desta pesquisa o quão a descoberta dos raios X e o progresso nesse domínio aliados às tecnologias de ponta vêm beneficiando no diagnóstico precoce da doença e mesmo propiciando a cura de muitos casos.

Palavras-chave: Radiologia. Câncer de mama. Imagem. Limiarização. Prevenção.

4.1.2 Abstract

The role of radiology has been paramount in the transformations and evolution of medicine's trajectories in the world. Radiological images provide relevant data for decision making of a diagnosis and subsequent treatment / follow-up of a patient. In the case of breast cancer, they may represent an important but not definitive step. In this sense, this article aims to demonstrate the importance of radiology with digital image processing techniques, especially thresholding in the diagnosis of breast cancer. To do so, first discusses the advances in imaging technologies, which has its first step in the first radiography, only then discuss about the prevention of breast cancer made possible by such advances. It is inferred from this research how the discovery of X-rays and the progress in this field combined with the latest technologies have been benefiting in the early diagnosis of the disease and even in the cure of many cases.

Keywords: Radiology. Breast câncer. Image. Threshold. Prevention.

4.2 INTRODUÇÃO

Tida como técnica mais clara e acessível de prática para se processar imagens e de velocidade computacional, a limiarização/thresholding tem sido utilizada frequentemente no âmbito de segmentação de imagens. Tal técnica é conhecida também como binarização, que “consiste na bipartição do histograma, convertendo os pixels cujo tom de cinza é maior ou igual a um certo valor de limiar (T) em brancos e os demais em pretos” (SANTA CATARINA, 2019, p. 59).

Na verdade, todo esse avanço, que tem contribuído sobremaneira com a medicina, decorre da descoberta dos raios X, há pouco mais de 126 anos. No caso do câncer de mama, os avanços da radiologia têm elevado a sobrevivência de muitas mulheres e evitado que outras tantas desenvolvam a doença quando fazem o diagnóstico precoce. Nesse sentido, o sistema de binarização cria técnicas de processamento de imagens digitais, que possibilitam a geração de parâmetros

substanciais para análise das imagens mamográficas, também refina a precisão na revelação de tumores, alterações anatômicas e lesões estruturais, tornando o diagnóstico elucidativo para as tomadas de decisão sobre prevenção/tratamento da paciente (SANTA CATARINA, 2019, p. 59).

Isto posto, apresenta-se o objetivo deste artigo de revisão de literatura: demonstrar a relevância da técnica limiarização/thresholding no diagnóstico do câncer de mama. Para alcançar tal objetivo, discorre-se primeiro sobre imagem digital, histograma, imagem binária e limiarização/thresholding e, por fim, a importância na prevenção do câncer de mama a partir dessa evolução tecnológica.

4.3 REVISÃO DE LITERATURA

4.3.1 Novas tecnologias de imagem

Uma imagem é a representação visual de um objeto, realizado através de várias técnicas/procedimentos: pintura, fotografia, radiografia, filme e outras. Imagem digital literalmente remete a dados, mas em tecnologia remete à informação binária, em dois estados, bidimensional, que se constrói tendo por base uma matriz binária, constituída de vários uns e vários zeros, de diversas maneiras. Tais imagens – atualmente em evidência, como se vê na área da saúde – resultam do desenvolvimento das ciências da informação, primeiramente no âmbito da astronomia e depois progressivamente expandiram-se em outros domínios, como na medicina e até mesmo atingiram de modo geral o público. Uma imagem digital é formada por milhares de pixels (pixel é o menor elemento de uma imagem digital), cada qual tem uma cor. Na imagem digital, cada fragmento é definido pelos vários modelos de cor, os quais são sistemas utilizados a fim de “organizar e definir cores através de um conjunto de propriedades básicas que são reproduzíveis. Estes foram desenvolvidos de maneira a uniformizar a forma de como as cores são especificadas em formato digital, de modo a produzi-las com rigor em qualquer tipo de hardware” (SOARES, 2010, p. 10-11).

Há uma distribuição de frequências/diagrama das frequências representada graficamente em colunas/retângulos a partir de um conjunto de dados que, de antemão, é tabulado e segmentado em classes uniformes, conhecida como histograma. Numa imagem, este “indica o número ou o percentual de pixels que a imagem tem em determinado nível de cinza ou cor”. E assim, gera um indicador de qualidade da imagem referente a contraste e intensidade luminosa (CONCI, 2015, p. 46). Com as imagens binárias, isto é, segmentadas em objeto e fundo, obtêm-se dados significativos das regiões/objetos destacados. Através da digitalização, a imagem se relaciona com a verdade com uma potência virtual. Há uma simulação com propriedade infinita que a torna uma imagem-imaginação. Com um clique no mouse, o real se configura. Conforme Balan (2009, p. 1),

Processar e analisar imagens com auxílio de máquinas é uma ciência que permite modificar, analisar e manipular imagens digitais a partir de um computador. Os algoritmos de processamento de imagens são procedimentos realizados passo-a-passo, com alta velocidade de cálculos, para otimizar as operações de tratamento de imagens. A complexidade se torna maior quando o objetivo é tornar a imagem digital um elemento representativo de informações que possam ser transmitidas a distância com o máximo de fidelidade da imagem original.

No sistema de visão binária, destaca-se a limiarização/Thresholding como estratégia eficiente e de simples implementação que utiliza como separador a intensidade do pixel; é um tipo de segmentação (limiar local, limiar global, múltiplos limiares), que é o processo que fragmenta a imagem em distintas regiões, cada qual com pixels com atributos similares. “Um sistema criado para realizar o processamento de imagens digitais geralmente é composto por cinco elementos: a aquisição da imagem, o armazenamento, o processamento em si, a comunicação e a exibição do resultado final do processo”. Particularmente, na área da saúde, o sistema de diagnóstico em computadores detecta, automática ou semiautomaticamente, anomalias em exames de imagem. Segmenta-se a imagem, subdividindo-a, para a distinção do “objeto de interesse do plano de fundo da imagem” que se dá pela segmentação que subdivide a imagem. Pela utilização da limiarização (ou threshold), define-se “um valor de limiar capaz de separar o objeto de interesse do plano de fundo e então uma imagem em níveis de cinza se torna uma imagem binária”. A informação que se tem a partir dessas técnicas auxilia o trabalho do médico. Estes são apenas alguns conceitos e técnicas para tratar imagens médicas digitais (COLELLA, 2018, p. 1).

A imagem digital, de tecnologia avançada, na medicina, enfaticamente na radiologia, tem sido utilizada no diagnóstico por imagens, investigando e monitorando a saúde de órgãos, músculos e tecidos, de modo que sensores enviam imagens diretamente ao computador, o qual as processa e as direciona à análise e interpretação do médico radiologista. Nesse decurso, surgem os mais sofisticados equipamentos para capturar imagens direta ou indiretamente. Com a incessante evolução tecnológica, principalmente a partir do advento do computador, os resultados das imagens se tornam mais nítidos e acurados. Há de se registrar o progresso relevante a partir da computação em nuvem, extinguindo a inevitabilidade que havia de tornar as imagens compactas visando a economizar espaços físicos na máquina (COLELLA, 2018, p. 1).

Furquim e Costa (2009, p. 91) afirmam que o desenvolvimento de técnicas de controle de qualidade de equipamentos alcançou “serviços de radiologia convencional, odontológico, mamografia convencional, processadoras que deveriam ser controladas diariamente, equipamentos de fluoroscopia com intensificador de imagem e alguma iniciativa em tomografia computadorizada”. Foi muito importante instituir o domínio sobre a irradiação de pacientes. Os equipamentos passam a ser calibrados e a funcionar em ambientes mais controlados, o que desencadeou um maior número de serviços de cálculo de blindagens e a realização de levantamentos radiométricos. Com o tempo e o avançar da tecnologia digital, os detectores capturam imagem radiológica com propriedades capazes de levar “à avaliação de grandezas diferentes daquelas utilizadas em sistemas de écran-filme”.

Com efeito, o formato digital da informação circula dentro e fora do ambiente hospitalar de forma consistente, fidedigna e automática, obviamente na trilha dos protocolos de segurança da informação. O Picture Archiving and Communication Systems (PACS) – que transmite, armazena e visualiza através de imagens os dados dos diagnósticos – é a técnica preferida hoje. Tal sistema propicia retomar imagens embasadas em conteúdos pelo Content Based Image Retrieval (CBIR), arquitetado com o fim de reproduzir as características das imagens armazenadas no PACS e utilizá-las em subsequente recuperação por similitude. Também o Digital Imaging and Communications in Medicine (DICOM) foi adotado como padrão de comunicação e armazenamento tanto de imagens médicas quanto de informações do paciente. Todos esses avanços exigem do médico radiologista competências e habilidades outrora inimagináveis, focadas em tecnologia da informação, engenharia

biomédica e inteligência artificial compartilhadas com as instituições hospitalares (OLIVEIRA; LEDERMAN; BATISTA, 2014, p. 19).

4.3.2 Prevenção do câncer de mama favorecida pelas novas tecnologias

Dentre os tipos de câncer, o melanoma (câncer de pele) é o que mais mata no mundo, mas entre as mulheres, no Brasil, atualmente, o maior índice de mortalidade é o de câncer de mama. Trata-se de um problema que pode ser minimizado, com a prevenção ou com o aumento de sobrevivência das pacientes, quando se realiza um diagnóstico precoce. “A mamografia é, isoladamente, o mais importante método de imagem na detecção das alterações da mama, possuindo uma sensibilidade próxima a 90%”. Se, por um lado, mamas densas e mamas de pacientes jovens (as de maior tecido glandular) interferem na nitidez na resolução da imagem mamográfica; por outro lado, mamas de mulheres próximas à menopausa (cujo tecido parenquimatoso involui e atrofia tornando-se adiposo quanto mais com o passar dos anos) propicia boa qualidade e definição da imagem. Na medicina, a maior parte dos diagnósticos médicos atualmente decorre do uso de imagem. “O processamento de imagens se concentra em desenvolver procedimentos para extrair informações de uma imagem de forma adequada para o processamento computacional” (NETO; RIBEIRO; VALERI, 2004, p. 1).

Quanto mais cedo se detectam as lesões mamárias, mais se reduz a mastectomia e as chances de sobrevivência aumentam. Mesmo considerada como o método “padrão-ouro” na detecção de câncer de mama, a mamografia pode apresentar um falso positivo que leva a biópsias desnecessárias, pois apesar de sua alta sensibilidade, possui baixa especificidade. Wild e Neal foram os pioneiros na proposição do uso da ultrassonografia no exame das mamas. Desde então, ficou provado que as técnicas de imagem por ultrassom são capazes de identificar muitos dos cânceres perdidos pelas técnicas mamográficas, especialmente aquelas que ocorrem em mulheres com seios densos. Ademais, imagens de ultrassom têm a vantagem de não serem invasivas, são de baixo custo e não há necessidade de adição ionizante. Apesar das vantagens acima, depende em grande parte do

desenvolvimento de algoritmos de segmentação eficientes (YAP; EDIRISINGHE; BEZ, 2006, p. 1).

Se, por um lado, a imagem no diagnóstico de doenças como o câncer de mama é imprescindível, por outro, os riscos à saúde da paciente é uma realidade. Daí a motivação no estabelecimento da instauração de métodos dosimétricos relativos à atuação de diagnóstico por imagem, com vista a avaliar danos à saúde do paciente quando submetido a exame que utiliza radiações ionizantes. É sempre necessário contrabalançar riscos e benefícios nesta modalidade diagnóstica procurando garantir que ela é, de fato, escolhida por levar à paciente menos efeitos maléficos à sua saúde e mais segurança no diagnóstico “quando comparada a outras opções diagnósticas” (COSTA *et al.* 2016, p. 176).

Conforme Furquim e Costa (2009, p. 91),

A radiologia diagnóstica tem como função principal diagnosticar patologias. Quando se utilizam imagens obtidas a partir da interação da radiação ionizante com o paciente, espera-se que esta apresente qualidade de modo a minimizar os erros de interpretação e identificação de estruturas, possibilitando diagnóstico mais preciso e com a menor dose. Uma imagem sem a qualidade adequada deve ser repetida e há alguns custos envolvidos neste processo que devem ser evitados, e o principal é a duplicação de dose em um mesmo paciente. Assim, a adoção de conceitos de qualidade em radiologia torna-se muito útil, uma vez que auxilia no controle do processo obtenção de imagem com a redução de erros previsíveis.

É certo que os avanços em diagnósticos e tratamento vêm possibilitando sobrevida progressiva em casos até pouco tempo considerados incuráveis. Inclusive, em muitos países reduziu-se a mortalidade decorrente do câncer de mama com a implantação de programas organizados de rastreamento; no Brasil, entretanto, mesmo com os esforços empregados, a mortalidade dessa doença tem aumentado e a incidência também, em mulheres entre 40 e 50 anos de idade (URBAN, 2017, p. 245).

A radiologia hoje é uma especialidade de suporte diagnóstico e terapêutico, uma vez que inclui desde a radiologia convencional até a mamografia, tomografia computadorizada, ultrassonografia, ressonância magnética, medicina nuclear, tomografia por emissão de pósitrons, radioterapia (FIOCRUZ, 2009, p. 1). Na alçada da radiologia, as tecnologias vêm avançando muito rápido e se manifestando cada vez mais indispensáveis no desenvolvimento contínuo dessa área. Os variados procedimentos requerem do radiologista uma interseção de seus conhecimentos técnico-científicos a outros, como os administrativos para um desempenho efetivo de

sua função. Ele há de ter uma “atuação multiprofissional integrada e complementar, incorporando complexos processos e tecnologias de ponta, com grandes investimentos em equipamentos, técnicas e insumos” (OLIVEIRA; LEDERMAN; BATISTA, 2014, p. 18).

De acordo com Alves (2018, p. 265), nos métodos de inteligência artificial,

Os algoritmos de *deep learning* (DL) têm-se tornado progressivamente mais sofisticados e são especialmente eficazes e superiores ao desempenho humano em tarefas de reconhecimento de padrões, segmentação de imagens e detecção de objetos. O desenvolvimento inicial destas técnicas foi no mercado dos videojogos, mas rapidamente foi adaptado à radiologia.

Tais algoritmos compõem uma série de técnicas; na radiologia, as mais utilizadas são as redes neurais já que o objetivo em radiologia é assenhorar-se de um sistema computacional inteligente, que analise grande quantidade de dados complexos e identifique arquétipos clinicamente significativos e proficientes. A evolução e perscrutação dos métodos de inteligência artificial em radiologia concentram-se “em quatro grandes áreas: detecção, quantificação, segmentação de imagem e classificação diagnóstica” (ALVES, 2018, p. 266).

No seu cotidiano, o uso de programas computadorizados leva o radiologista a manipular aparelhos cada vez mais vanguardistas e sofisticados. Assessorados pela tecnologia da informação, que trouxe em seu bojo a era digital e seus equipamentos em conexão com imagens e laudos arquivados nos computadores, os radiologistas podem distribuir essas informações a qualquer hora e para qualquer local. Tal revolução na área radiológica impõe desafios cada vez maiores ao profissional impelindo-o à aquisição de equipamentos mais aprimorados e onerosos cuja compensação é a “redução do custo do investimento inicial pelo aumento da produtividade e da qualidade das imagens” (OLIVEIRA; LEDERMAN; BATISTA, 2014, p. 18).

Nesse sentido, Alves (2018, p. 268) pondera que numa conjuntura de orçamentos mais comedidos é imprescindível que a aquisição de equipamentos “sejam submetidos a um rigoroso processo de avaliação custo-benefício”. Além disso, comenta que a radiologia de intervenção que cresce incessantemente possibilita terapêuticas minimamente invasivas a partir de meios imagiológicos mais precisos, diferentemente de décadas passadas em que “a radiologia era uma atividade exclusivamente diagnóstica e sem contato clínico”.

Para Alves (2018, p. 269), “A radiologia está a sofrer transformações profundas desde o final do século XX e nos próximos anos prevê-se o aparecimento de múltiplas tecnologias que serão certamente disruptivas, mas que criarão também novas oportunidades”.

4.4 METODOLOGIA

As imagens utilizadas como entrada para os resultados deste trabalho foram baixadas de um banco de imagens. Lee *et. al.* (2017) disponibilizaram uma versão atualizada e padronizada do Banco de Dados Digital para Mamografia de Triagem (DDSM), deste conjunto de dados utilizou-se o CBIS-DDSM (subconjunto de imagens de mama curadas da DDSM) que inclui imagens descomprimidas, seleção e curadoria de dados por mamógrafos treinados. A linguagem utilizada para a geração do algoritmo de *thresholding* foi o Python 3.6.8 no ambiente de desenvolvimento Web Jupiter.

4.4.1 Imagem digital

Uma imagem digital é uma função $f(x, y)$ discretizada tanto em coordenadas espaciais quanto em brilho. Essa função é um produto da luminância pela reflectância em cada ponto (x, y) (FILHO; NETO, 1999).

4.4.2 Histograma

O histograma representa a quantidade de pixels de cada nível de cinza em uma imagem monocromática. Estes valores são normalmente representados por um

gráfico de barras que fornece para cada nível de cinza a contagem de pixels correspondentes na imagem (GONZALEZ; WOODS, 2008).

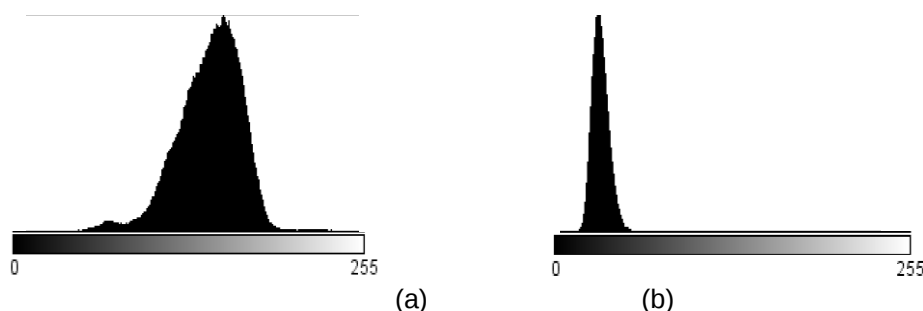


Figura 8– Histogramas: 8a) Imagem clara; 8b) imagem escura.

Fonte: Próprio autor

4.4.3 Limiarização (*Thresholding*)

É uma técnica que consiste em separar as regiões de uma imagem quando esta apresenta duas classes (o fundo e o objeto). Devido ao fato da limiarização produzir como saída uma imagem binária, este processo é denominado, muitas vezes, binarização (JAIN; 1989). Na operação de limiarização, uma imagem de entrada $f(x, y)$ com N tons de cinza produz na saída uma imagem binária $T(x, y)$ segundo a regra:

SE $f(x, y) \geq L$ então $T(x, y) = 1$, caso contrário $T(x, y) = 0$;

no qual L é o ponto de corte ou limiar. Este valor de limiar L pode ser obtido tanto manualmente quanto automático.

4.5 RESULTADOS E DISCUSSÃO

O processo de limiarização das microcalcificações baseia-se em verificar os pixels com alta intensidade luminosa, levando em consideração cada imagem de recorte. A imagem recortada contém as possíveis áreas de concentrações de microcalcificações. Para encontrar essas microcalcificações, é efetuado o histograma da imagem e a partir deste visualiza-se um ponto de corte (limiar) que vai

separar a imagem em dois tons (preto ou branco). Para todas as imagens o limiar foi encontrado de forma manual (supervisionado pelo especialista).

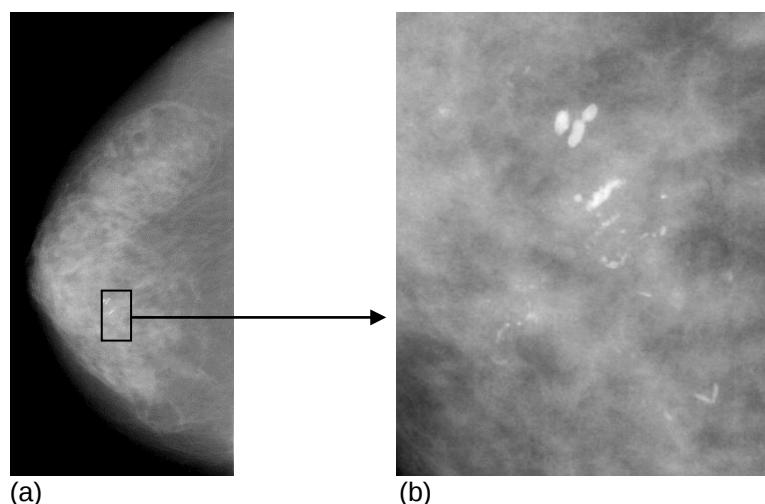


Figura 9– 9a) Mama seio esquerdo; 9b) Recorte da área de interesse.

Fonte: Imagem adaptada de Lee *et al.* (2016).

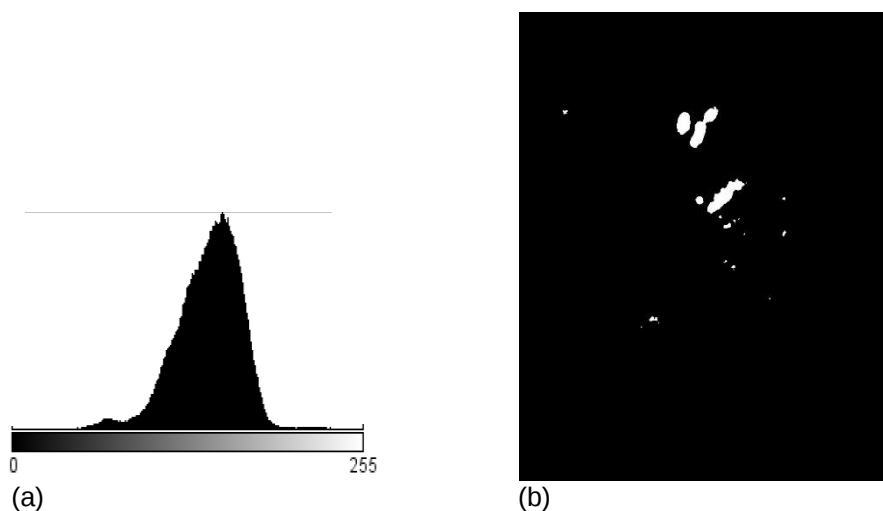


Figura 10– 10a) Histograma da Figura 10b; 10b) Limiar com valor $L = 200$

Fonte: Próprio autor.

A partir da imagem original (Figura 9a) o CBIS-DDSM também fornece as imagens de ROI (Region of Interest) (Figura 9b). A imagem de resultado do algoritmo de limiarização pode ser observada na Figura 10b, no qual é gerada uma imagem binária com um ponto de corte em $L = 200$. Conforme pode ser visto no histograma (Figura 10a), há uma maior frequência de pixels com maior intensidade luminosa (pixels claros) na imagem de recorte (Figura 9b), possivelmente esses pixels com tons mais claros são microcalcificações da imagem original.

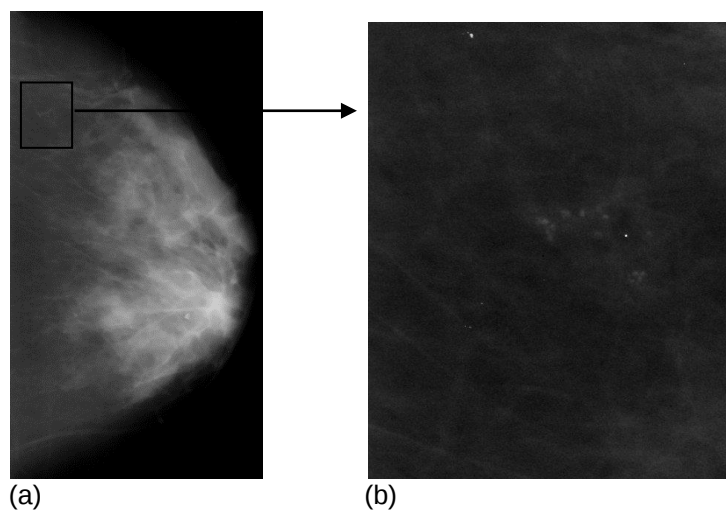


Figura 11– 11a) Mama seio direito; 11b) Recorte da área de interesse.

Fonte: Imagem adaptada de Lee *et al.* (2016).

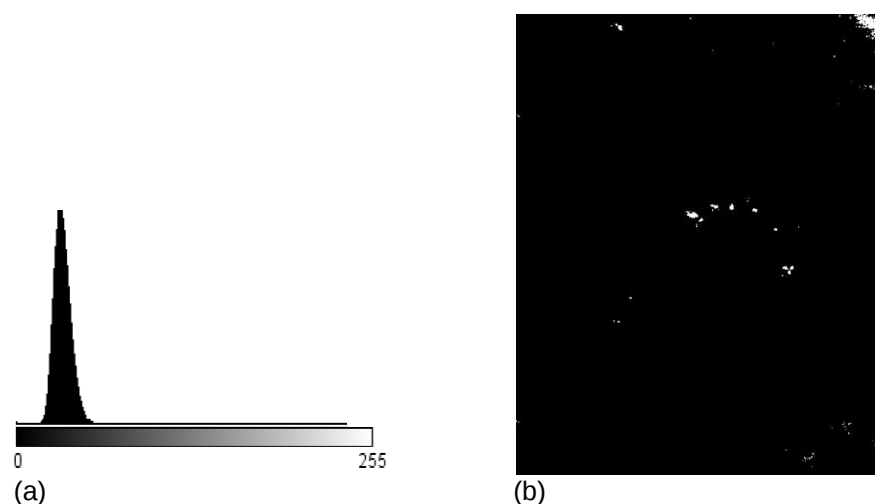


Figura 12– 12a) Histograma da Figura 12b; 12b) Limiar com valor $L = 61$

Fonte: Próprio autor.

No histograma (Figura 12a) que foi gerado pela Figura 11b, há uma maior frequência de pixels de baixa intensidade luminosa (pixels escuros), isso acarreta num ponto de corte baixo, $L = 61$ (Figura 12b). Pode-se destacar que as bordas superior e inferior da imagem estão fora da região central onde se localizam as microcalcificações.

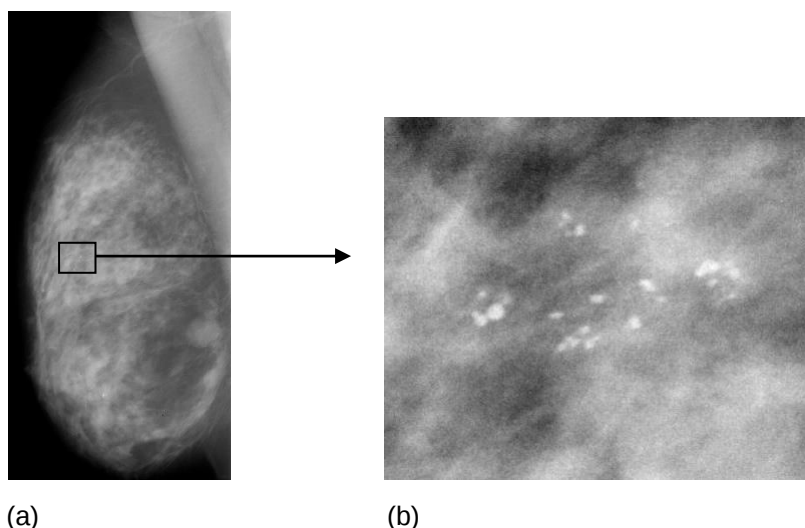


Figura 13– 13a) Imagem Original; 13b) Recorte da área de interesse.

Fonte: Imagem adaptada de Lee *et al.* (2016).

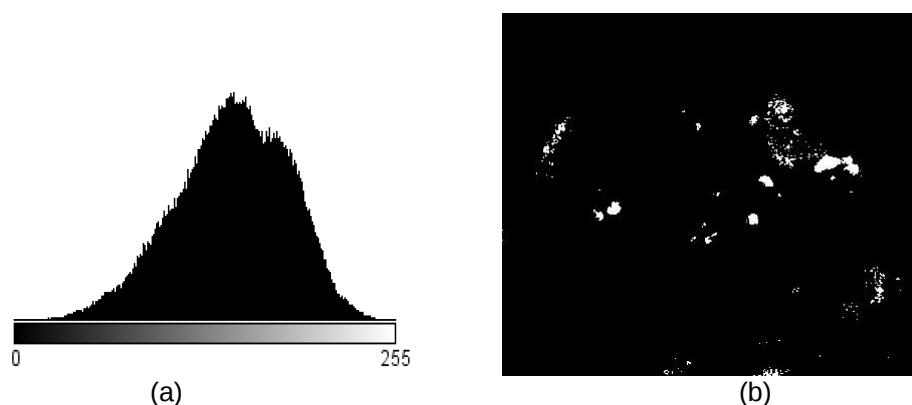


Figura 14– 14a) Histograma da Figura 13b; 14b) Limiar com valor $L=171$

Fonte: Próprio autor.

No histograma (Figura 14a) que foi gerado pela Figura 13b, há um equilíbrio de pixels em baixa e alta intensidade luminosa. O ponto de corte utilizado foi $L = 171$ (Figura 14b). Pode-se destacar que as bordas superior e inferior da imagem estão fora da região central onde se localizam as microcalcificações.

4.6 CONCLUSÕES

A literatura na área da radiologia nos mostrou o seu comprometimento nas transfigurações e melhorias nos rumos da medicina desde a descoberta dos raios X

até os dias atuais, em que as imagens radiológicas se tornam cada vez mais elucidativas proporcionando ao médico radiologista mais segurança em dar um diagnóstico e decidir sobre melhor tratamento para o paciente.

Em relação ao câncer de mama, a pesquisa com o uso da limiarização/thresholding demonstrou que os progressos tecnológicos coligados à radiologia contribuem, de fato, para a tomada de decisões mais firmes e acertadas, evitando mortes prematuras, evolução da doença, tratamentos cirúrgicos traumáticos como a mastectomia; pelo contrário: favorece, a partir de um diagnóstico precoce, uma sobrevida de qualidade para a paciente e até mesmo a cura.

4.7 REFERÊNCIAS

ALVES, P. A Radiologia no Início do Século XXI. **Gazeta Médica**, Lisboa, v. 5, n. 4, p. 265-269, out.-dez. 2018.

BALAN, W. C. **Apostila Instrucional**. 2009. 7 p. Disponível em: <http://twixar.me/tKM1>. Disponível em: 29 ago. 2019.

COLELLA, S. R. L. Processamento de imagens digitais na área da saúde. **Technology Leadership Council Brazil, IBM**, [s.l.], ano 13, n. 318, ago. 2018. Mini Paper Series.

CONCI, A. **A importância do histograma em Análise de Imagens**. [Niterói, RJ]: Universidade Federal Fluminense, 2015. 138 p.

COSTA, P. R. *et al.* Correlação entre dose efetiva e riscos radiológicos: conceitos gerais. **Radiologia Brasileira**, São Paulo, v. 49, n. 3, p. 176-181, São Paulo, 2016.

FIOCRUZ. **Serviços de radiologia**: conceitos básicos de planejamento. 2009. 43 p. Disponível em: encurtador.com.br/gxMW4. Acesso em: 23 set. 2019.

FURQUIM, T. A. C.; COSTA, P. R. Garantia de qualidade em radiologia diagnóstica, **Revista Brasileira de Física Médica**, São Paulo, v. 3, n. 1, p. 91-99, 2009.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. 3 rd. Tennessee: Pearson, 2008.

JAIN, A. K. **Fundamentals of Digital Image Processing**. New Jersey: Prentice Hall, 1989.

LEE, R. S. *et al.* A curated mammography data set for use in computer-aided detection and diagnosis research. **Scientific data**, New York, v. 4, p. 170-177, 2017.

LEE, R. S. *et al.* Curated breast imaging subset of DDSM. **The cancer imaging archive**, Little Rock, v. 8, 2016.

MARQUES FILHO, O.; NETO, H. V. **Processamento digital de imagens**. Rio de Janeiro: Brasport, 1999.

NETO, G. H.; RIBEIRO, G. C.; VALERI, F. V. Processamento e Segmentação de Mamogramas Digitais. *In*: CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 9., 2004, Ribeirão Preto. **Anais [...]**. São Paulo: CBIS, 2004.

OLIVEIRA, A. F. de; LEDERMAN, H. M.; BATISTA, N. A. O aprendizado sobre a tecnologia no diagnóstico por imagem, **Radiologia Brasileira**, São Paulo, v. 45, n. 1, p. 18-22, jan.-fev. 2014.

SANTA CATARINA, A. **Projeto da Disciplina Processamento de Imagens Digitais (PID)**. Cascavel, PR: Unioeste, 2019.

SOARES, A. C. Introdução à imagem digital, definição de pixel e cores digitais. **ESMGA**, [s.l.], ano 12, n.1, p. 10-11, 2010. Disponível em: <https://anasoares1.wordpress.com/2010/11/22/introducao-a-imagem-digital-definicao-de-pixel-e-cores-digitais/>. Acesso em: 10 jun. 2019.

URBAN, L. A. B. D. Recomendações do Colégio Brasileiro de Radiologia e Diagnóstico por Imagem, da Sociedade Brasileira de Mastologia e da Federação Brasileira das Associações de Ginecologia e Obstetrícia para o rastreamento do câncer de mama. **Radiologia Brasileira**, São Paulo, v. 50, n. 4, p. 244–249, jul.-ago. 2017.

YAP, M. H.; EDIRISINGHE, E. A.; BEZ, H. Object Boundary Detection in Ultrasound Images. *In*: CONFERENCE ON COMPUTER AND ROBOT VISION, 3., 2006, Canada. **Proceedings [...]**. [s.l.]:[s.n.], 2006. p. 6.

5 ARTIGO C – SEGMENTAÇÃO AUTOMÁTICA DE ÁREAS DE CALCIFICAÇÕES EM IMAGENS DIGITAIS DE MAMA

5.1 RESUMOS

5.1.1 Resumo em português

Este trabalho tem como objetivo evidenciar que a mamografia – exame de imagem utilizado para o diagnóstico de anormalidades mamárias – pode se tornar mais eficaz quando a mamografia se associa a técnicas de segmentação inteligente, facilitando a interpretação dos achados pelo médico radiologista e, assim, auxiliando o diagnóstico precoce do câncer de mama. Após uma revisão de literatura sobre o câncer de mama (categorias, prevenção envolvendo ambiente e estilo de vida, diagnóstico e rastreamento da doença), descreve-se a metodologia com alguns conceitos de imagem digital e técnicas de Machine Learning (Redes Neurais e Florestas Aleatórias). Para a elaboração dos resultados (na metodologia descrita), utilizou-se um banco de imagens previamente classificado com regiões suspeitas. O software Fiji foi a ferramenta utilizada para extração de regiões suspeitas candidatas em imagens de mamografia. Foram definidas três classes para o resultado da segmentação da imagem. Tanto a Floresta Aleatória quanto a rede neural apresentaram um bom resultado na segmentação no destaque das regiões suspeitas. Realizou-se a detecção dos contornos das regiões, indicando que podem

ser realizados cortes destas regiões segmentadas e posteriormente realizar uma classificação automática dos alvos utilizando um algoritmo de aprendizado.

Palavras-chave: Câncer de mama. Imagem. Machine learning. Prevenção.

5.1.2 Abstract

The aim of this study is to show that mammography - image exam used to diagnose breast abnormalities - is associated with smart segmentation techniques, facilitating the interpretation of the findings by the radiologist and thus helping Early diagnosis of breast cancer. Following a literature review on breast cancer (categories, prevention involving environment and lifestyle, diagnosis and tracking of the disease), the methodology with some digital imaging concepts and Machine Learning techniques (Neural Networks and Random Forests). For the elaboration of the results (in the described methodology), a previously classified image bank with suspicious regions was used. Fiji software was the tool used to extract candidate suspect regions in mammography images. Three classes were defined for the result of image segmentation. Both the Random Forest and the neural network showed a good segmentation result in the highlighting of the suspect regions. The detection of the contours of the regions was performed, indicating that cuts of these segmented regions can be performed and then an automatic classification of the targets can be performed using a learning algorithm.

Keywords: Breast câncer. Image. Machine learning. Prevention.

5.2 INTRODUÇÃO

No Brasil, o grande número de morte de mulheres por câncer de mama, torna a doença um problema vultoso de saúde pública, tanto pela comoção da morbimortalidade, quanto pelo alto custo pessoal e social relativo à doença e a seu tratamento, e mais ainda pelas danosas sequelas físicas e psicológicas nas pacientes. Trata-se de um tumor originado das células de mama que crescem

desordenadamente. 60% a 70% desse tipo de câncer surgem em forma de nódulo “irregular ou esferoide, de margens espiculadas, indistintas ou microlobuladas, sem calcificações. Próximo de 20% dos casos são nódulos com calcificações. Calcificações sem nódulo associado constituem pouco menos de 20% do total de casos” (BOFF *et al.*, 2019, p. 170).

O câncer mamário é bastante assustador tanto pela alta frequência quanto pelos efeitos físicos (seja pela mastectomia total seja pela ressecção segmentar – a cicatriz deixada tem sua importância) e psicológicos (baixa autoestima, redução da sexualidade, estigma) que em geral lesionam os pacientes. É uma das doenças que mais provocam mortificações e transtornos em ampla dimensão: afeta a doente, os familiares próximos e os cuidadores (BOFF *et al.*, 2019, p. 170).

Na verdade, o maior risco dessa doença é o diagnóstico tardio do tumor, o que pode ser evitado com as tecnologias digitais que vêm avançando em ritmo cada vez mais acelerado para apoiar o diagnóstico precoce da doença. A mamografia, por exemplo, é um exame usado para diagnosticar anormalidades na mama; é utilizada em programas para rastreamento em mulheres; por ela se reduz, em mais de 40% o índice de mortalidade decorrente de câncer de mama, obviamente se o diagnóstico se dá ainda em estágio pré-clínico, isto é, fase precoce de evolução dos tumores malignos (BOFF *et al.*, 2019, p. 171).

Pelo fato de o câncer de mama ser o segundo de maior incidência no mundo e sua prevenção e principalmente tratamento serem onerosos para a saúde pública de um país, a limiarização/thresholding é uma técnica mais simples de implementação para processamento de imagens e de velocidade computacional, aplicada com muita frequência no cenário de segmentação de imagens. É também chamada de binarização, pelo fato de a técnica basear-se “na partição do histograma da imagem para converter todos os pixels [picture element] cujo tom de cinza é maior ou igual a um certo valor de limiar T em brancos e os demais em pretos” (SOUTO, 2013, p. 66).

Isso posto, surge a pergunta: dentre as muitas possibilidades reais de prevenção do câncer de mama, qual seria a relevância do threshold nesse empenho? A busca de respostas a esse questionamento propicia alcançar o objetivo desta pesquisa que é o de discorrer sobre vantagens e limites da técnica da limiarização/ threshold aplicada à mamografia na busca das áreas suspeitas de

neoplasia maligna, facilitando a interpretação dos achados pelo médico e, dessa forma, auxiliando o diagnóstico precoce do câncer de mama.

Assim sendo, terminando esta introdução, a pesquisa apresenta a seguir o referencial teórico sobre as duas variáveis do problema: câncer de mama e a técnica threshold. Mais adiante, vem a descrição da metodologia, seguida da análise e discussão e, na sequência, estão as considerações finais; por fim, as referências.

5.3 REVISÃO DE LITERATURA - CÂNCER DE MAMA

Ainda que não haja etiologia delineada, o câncer de mama é estudado como multifatorial, de modo que determinados fatores de risco em interação determinam a doença. Boff *et al.* (2019) discorrem sobre as duas categorias de fatores de risco:

1) *Modificáveis* – sujeitos a modificações, intervenções, controles como, por exemplo, “obesidade, hábito alimentar rico em carboidrato, consumo exagerado de carne vermelha e gorduras, alta ingestão de bebida alcóolica, realização de terapia de reposição hormonal combinada por mais de 5 anos e exposição excessiva à radiação” (BOFF *et al.*, 2019, p. 169). Lê-se em Vieira que “o excesso de peso é considerado fator de risco para o desenvolvimento da doença e isso pode ser explicado pelos níveis estrogênicos elevados resultantes da conversão periférica no tecido adiposo” (VIEIRA, 2017, p. 21-22).

2) *Não modificáveis* – não se modificam e são inerentes ao paciente, tais como gênero (incidência maior no feminino), idade (predominância nas que passam dos 50 anos), raça e etnia (mais frequente entre as brancas não hispânicas e as negras). Ademais,

[...] status reprodutivo e hormonal da mulher sendo a menarca precoce e menopausa tardia um dos mais associados, história familiar, associação quando parentes de primeiro grau (mãe, irmã, filha, pai, irmão e filho) desenvolveram câncer de mama, mutação nos genes BRCA1 e BRCA2, famílias portadoras destas mutações apresentam fortes indicativos para a doença e também a presença de patologia mamária prévia, mulheres com câncer de mama prévio, possuem um risco aumentado em 1,5x em desenvolver novamente o câncer de mama” (BOFF *et al.*, 2019, p. 170).

Como melhor prevenir o câncer de mama? Eis a pergunta que deveria fazer parte do questionamento das mulheres, principalmente, pois as respostas possibilitariam uma prevenção primária da doença, isto é, anterior ao início do processo patológico, evitando, inclusive, a exposição a muitos fatores de risco modificáveis, como o ambiente e estilo de vida, por exemplo, preservando-se, assim, a saúde e reduzindo a mortalidade em decorrência desse mal.

Alguns hábitos deveriam fazer parte da vida de todas as mulheres, a saber: praticar atividades físicas, amamentar lactente, alimentar-se regularmente de frutas, verduras, vegetais, peixe e castanhas, azeite de oliva e, em contrapartida evitar a ingestão de gordura e carne vermelha (o que ainda não está muito claro), alimentos processados, alimentos com alto índice glicêmico, reduzir o uso de açúcar (VIEIRA, 2017, p. 22-23).

São, pois, práticas simples a serem cultivadas, principalmente quando se atesta que, no Brasil, o risco de se ter câncer de mama é de 8%, ao longo da vida, o que significa que uma mulher em cada doze, um risco expressivo do sexo feminino. Em vista disso, a Sociedade Brasileira de Mastologia, Sociedade Brasileira de Cirurgia Oncológica e a Sociedade Brasileira de Radiologia recomendam o rastreamento da doença em mulheres a partir dos quarenta anos submetendo-se à mamografia anualmente (VIEIRA, 2017, p. 25) com o fito de detectar quaisquer anormalidades da mama.

Portanto, o procedimento é o do diagnóstico precoce do estágio pré-clínico, antes mesmo da apresentação de quaisquer sintomas, o que aumenta consideravelmente as chances de cura. Tem-se aí a prevenção secundária, cujo principal objetivo é o rastreamento universal e precoce e a realização da mamografia, aplicada “a grandes populações, em programas de rastreamento, reduz significativamente índices de mortalidade por câncer de mama (redução acima de 40%)” (BOFF, 2019, p. 170).

A prevenção secundária tem por finalidade alterar o curso da doença, uma vez que seu início biológico já aconteceu, por meio de intervenções que permitam sua detecção precoce e seu tratamento oportuno. Para isso, deve haver clara evidência de que a doença em questão possa ser identificada em uma fase precoce, quando ainda não está clinicamente aparente, e que permita uma abordagem terapêutica eficaz, alterando seu curso ou minimizando os riscos associados com a terapêutica clínica. Além disso, a queda resultante em morbidade ou mortalidade deve ser alcançada sem um grande fardo de efeitos adversos gerados pela estratégia adotada. A detecção precoce de uma doença é possível por meio de educação para o

diagnóstico precoce em pessoas sintomáticas ou do rastreamento (triagem) em populações assintomáticas (THULER, 2003, p. 234).

Já a prevenção terciária ocorre numa fase clínica e sintomática da doença em face de achados como nódulo, edema – fase em que a mamografia deixa de ser de rastreio para ser diagnóstica. “É importante salientar que a mamografia apresenta falsos positivos e falsos negativos, uma falha inerente ao método, mas é o melhor método de rastreamento disponível no momento” (VIEIRA, 2017, p. 25).

Há dois grupos de exames de mamografia: o do rastreamento (em pacientes assintomáticos, de prevenção primária, que deve ser realizado anualmente após quarenta anos de idade em período pós-menstrual) e o do diagnóstico (em pacientes sintomáticos sugestivos de câncer de mama ou ainda dos que “necessitam de complementação com um outro exame”) (BOFF *et al.* 2019, p. 171).

É, portanto, fundamental a precocidade no diagnóstico e tratamento da doença para que as consequências sejam menos lesivas, evitando-se ou reduzindo-se mastectomias e o risco de morte. Inclusive, se diagnosticado precocemente, há grandes chances de recuperação e até mesmo de cura. O tratamento geralmente é longo, não menos que um ano. Uma das fases mais críticas é a da quimioterapia, que, embora muito eficaz, seus efeitos colaterais, de modo geral, são perversos: frequência de náuseas, vômitos e mucosites, mudanças somáticas (alopecia, aumento de peso, falência ovariana, redução de hormônios – testosterona e estrógeno provocando menopausa e atrofia vaginal –, várias baixas, como as de libido, lubrificação vaginal, anorgasmia e dispareunia, afetando a sexualidade. Ademais, a revelação da doença aos outros é mais um problema a enfrentar, numa sociedade leiga de onde emergem julgamentos valorativos inconvenientes para o doente, inclusive o sentimento de piedade. Mais sinistro ainda é o sentimento de finitude em face da devastadora doença (SANTOS; SANTOS; VIEIRA, 2014; OLIVEIRA; OSELAME; NEVES, 2014).

Em face de problemas tão nefastos, a insistência na prevenção nunca é demais, como também não é demais insistir nos bons hábitos alimentares e comportamentais, prática de exercícios físicos, vida regrada com padrões de sono, que são conceitos básicos da Saúde Pública. Nessa prevenção, insiste-se, a mamografia é indispensável, ainda que a sua sensibilidade seja de aproximadamente 85% e podendo reduzir-se a 50% em mamas muito densas. Nesse sentido, qualquer alteração identificada no exame deve ser descrita conforme

o Breast Imaging Reporting and Data System (BI-RADS), “um sistema criado para padronização dos laudos dos exames diagnósticos da mama quanto aos termos utilizados, criação de relatório e recomendação de conduta, o que facilita a comunicação entre a equipe multidisciplinar que assiste a paciente” (BOFF *et al.*, 2019, p 163). Também imagens podem ser melhoradas a partir da técnica *Threshold*.

5.4 METODOLOGIA

Para realizar a avaliação da metodologia proposta, foi necessário obter um banco de imagens previamente classificado, e com regiões suspeitas já segmentadas. Lee *et al.* (2017) disponibilizaram uma versão atualizada e padronizada do Banco de Dados Digital para Mamografia de Triagem (DDSM). Este conjunto de dados, o CBIS-DDSM (subconjunto de imagens de mama curadas da DDSM), inclui imagens descomprimidas, seleção e curadoria de dados por mamógrafos treinados. Neste trabalho, a ferramenta utilizada para extração de regiões suspeitas candidatas em imagens de mamografia foi o software Fiji, uma distribuição de software livre do projeto ImageJ. Fiji é um software focado na análise de imagens biológicas e conta com a combinação de poderosas bibliotecas de software com uma ampla variedade de linguagens de script que permitem a prototipagem rápida de algoritmos de processamento de imagem (SCHINDELIN *et al.*, 2012).

5.4.1 Imagem digital

Gonzalez e Woods (2018) definem uma imagem como sendo uma função bidimensional $f(x,y)$ de intensidade da luz, em que x e y denotam as coordenadas espaciais de um ponto e o valor da função f (Figura 15) é proporcional ao brilho da imagem naquele ponto (imagem monocromática). Uma imagem digital é uma função $f(x, y)$ discretizada tanto em coordenadas espaciais quanto em brilho. Essa função é um produto da luminância pela reflectância em cada ponto (x, y) .

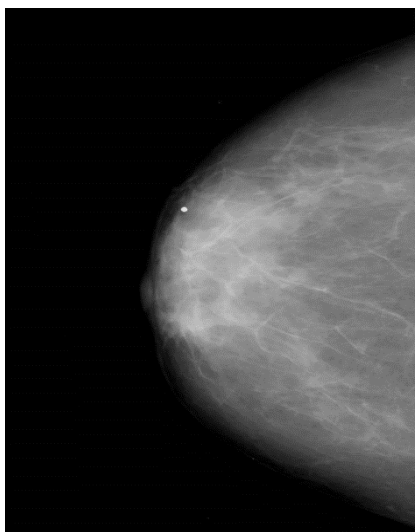


Figura 15– Mamografia

Fonte: Imagem adaptada de Lee *et al.* (2016)

5.4.2 Imagem monocromática

Uma imagem monocromática contém pixels com apenas uma graduação de cinza (tons de cinza). Toda escala tem um valor mínimo e máximo. No caso da escala de cinza (com 8 bits), os pixels que se aproximam de zero são os pixels mais escuros, enquanto os que se aproximam do valor máximo menos um ($L-1$) são os pixels mais claros (GONZALEZ; WOODS, 2018).



Figura 16– Escala de cinza

Fonte: Próprio autor

5.4.3 Limiarização (*Thresholding*)

É uma técnica que consiste em separar as regiões de uma imagem quando esta apresenta duas classes (o fundo e o objeto). Devido ao fato da limiarização produzir como saída uma imagem binária, este processo é denominado, muitas vezes, binarização (GONZALEZ; WOODS, 2018). De forma semelhante à imagem monocromática, uma imagem binária é uma matriz bidimensional com apenas dois valores. Às vezes são denominadas de imagens lógicas: preto corresponde ao valor 0 e branco corresponde ao valor 1 (SOLOMON; BRECKON, 2011).



Figura 17 – Imagem Binária

Fonte: Imagem adaptada de Lee *et. al.* (2016)

5.4.4 Redes neurais

O neurofisiologista McCulloch e o matemático Walter Pitts em 1943 modelaram *redes neurais artificiais simples* usando circuitos elétricos. Segundo McCulloch e Pitts (1943), um neurônio pode ser representado através de lógica binária (0 ou 1). As redes neurais artificiais surgiram da busca de solucionar problemas de maneira análoga ao cérebro.

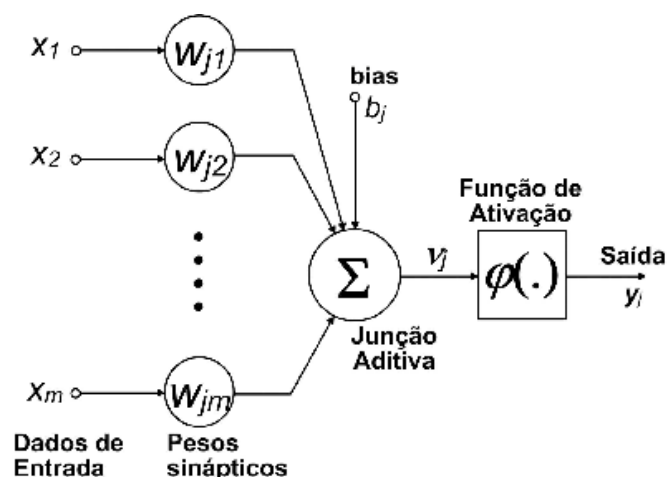


Figura 18– Modelo de um neurônio artificial

Fonte: MARTINS, F. R. In: <http://twixar.me/thV1>.

5.4.5 Random Forest (Floresta Aleatória)

Random Forest é um classificador que se constitui de uma coleção de árvores de decisão. A ideia é que, se uma árvore é boa, uma floresta deve ser ainda melhor, desde que haja variedade suficiente dentro dela. A coisa mais interessante sobre uma floresta aleatória é a maneira como cria aleatoriedade a partir de um conjunto de dados padrão (MARSLAND, 2015).

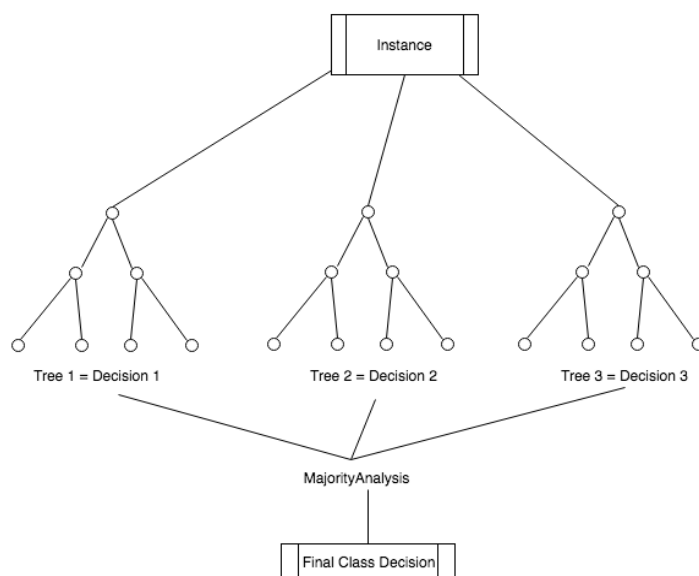


Figura 19– Modelo de uma Floresta Aleatória

Fonte: O'Reilly Medi, Inc. In: <https://bit.ly/2mnoCYB>.

5.5 RESULTADOS E DISCUSSÃO

O processo de segmentação das regiões suspeitas candidatas é baseado em um vetor de características gerado para cada pixel da imagem. Deste modo é possível obter um padrão de atributos que possam distinguir calcificações. Este método consiste em aplicar variados filtros de processamento de imagens sobre a imagem do exame, gerando uma imagem de múltiplos canais, de forma que cada imagem obtida pela aplicação dos filtros será utilizada como atributo para treinamento de um algoritmo de *machine learning* (ARGANDA-CARRERAS et. al., 2017).

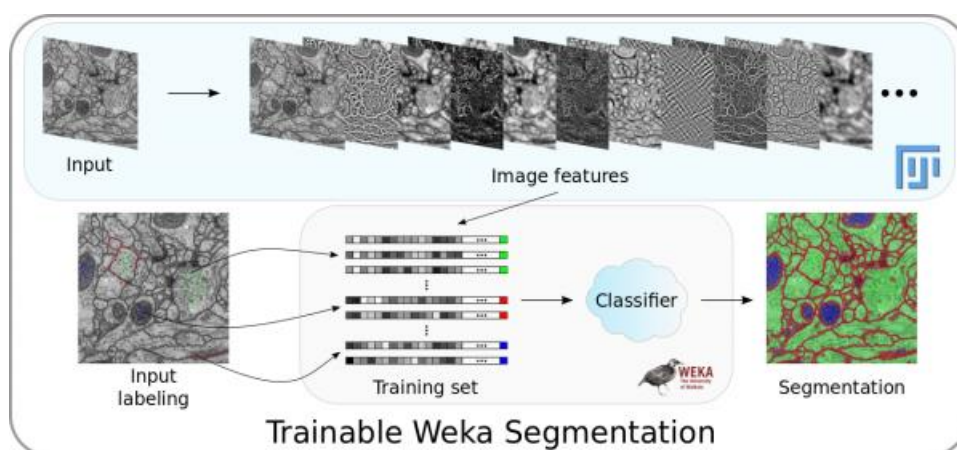


Figura 20– Vetor de atributos de uma imagem aplicado a um algoritmo

Fonte: Arganda-Carreras, I. et al. (2017).

Como se pode observar, na Figura 20, os algoritmos de aprendizado de máquina são fornecidos pelo kit de ferramentas de mineração de dados e aprendizado de máquina Waikato Environment for Knowledge Analysis (WEKA) (HALL et al., 2009).

Através deste *plugin* do Fiji, foram selecionados os filtros de imagem que melhor atendem a extração de características de calcificações em imagens mamográficas. A Figura 21 demonstra a seleção dos filtros e o algoritmo de aprendizado utilizados neste estudo.

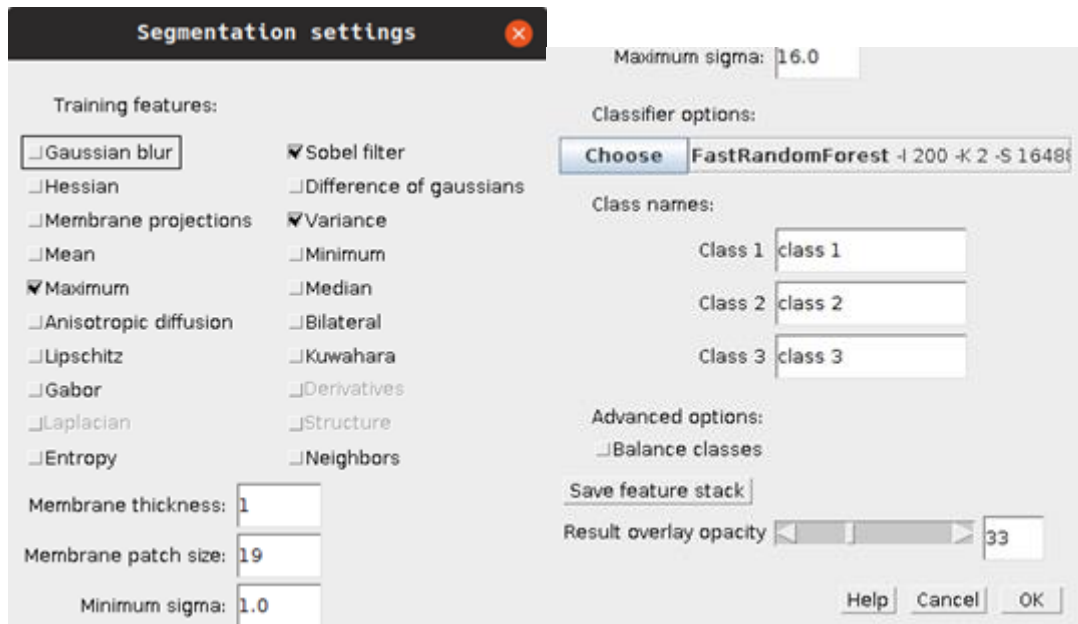


Figura 21– Tela de parâmetros por Segmentação com Random Forest

Fonte: Weka.

Foram selecionados três filtros, um detector de bordas (Sobel), variância e valor máximo. Para cada filtro aplicado são geradas novas imagens, realizando alterações nos atributos requeridos para cada filtro extrator de características. Assim sendo, o *plugin* aplica variadas configurações para cada filtro escolhido.

Além das configurações de definição dos filtros e algoritmo de aprendizado, foram definidas três classes para o resultado da segmentação da imagem.

A Figura 22b representa em três cores as classes: o vermelho são as regiões amostrais de calcificação; as amostras verdes representam demais regiões da imagem; e, por fim, em roxo, as regiões mais escuras a serem ignoradas.

Definidas as amostras de treinamento, os vetores de características serão gerados; os valores das regiões das amostras serão utilizados como dado de treinamento do algoritmo *Random Forest*. Após a fase de treinamento irá ocorrer a predição para cada pixel da imagem que será segmentada.

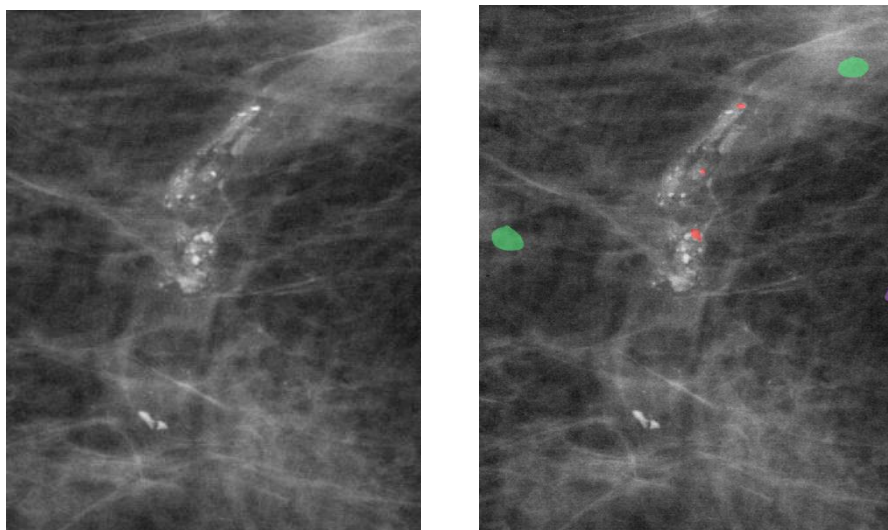


Figura 22- à esquerda: Imagem Original; à direita: Imagem com as classes

Fonte: Imagem adaptada de Lee *et al.* (2016).

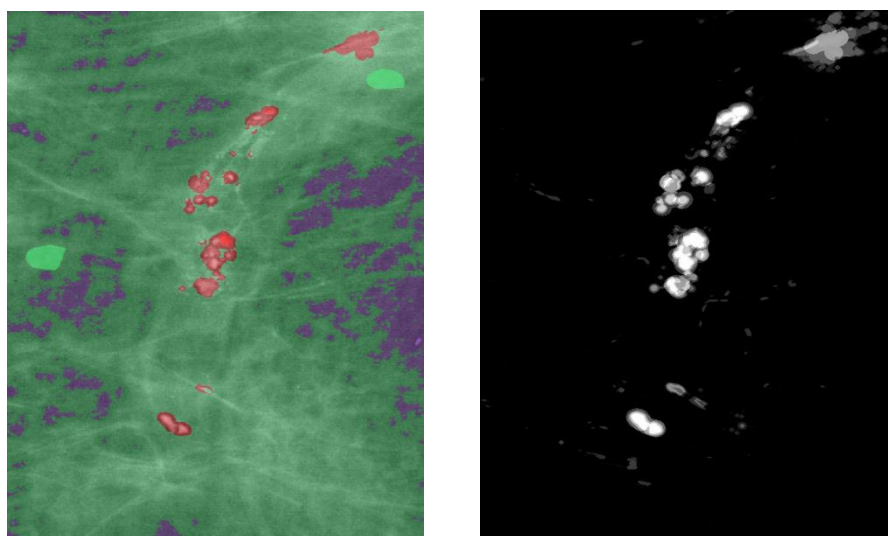


Figura 23- à esquerda: Predição utilizando Random Forest; à direita: Mapa de probabilidade Random Forest

Fonte: Imagem adaptada de Lee *et al.* (2016).

As imagens de resultado do algoritmo de segmentação podem ser observadas na Figura 23: a primeira apresenta segmentação das três classes previamente definidas; e a segunda, o mapa da probabilidade das regiões suspeitas candidatas.

O algoritmo aplicado para segmentação da imagem retorna à probabilidade para cada pixel da imagem pertencer à primeira classe definida, calcificações, deste modo a Figura 23 apresenta tons mais claros para uma maior possibilidade de região suspeita. A probabilidade de cada pixel da imagem tem o valor entre 0 (zero)

e 1 (um), de forma que 1 (um) corresponde a cem por cento de confiabilidade de acerto.

Com a finalidade de avaliar a utilização do mesmo método de extração de características da imagem, foram realizados testes de segmentação utilizando redes neurais como algoritmo de aprendizado. Os filtros de imagem para criar o vetor de características para cada pixel da imagem foi o mesmo aplicado ao algoritmo *Random Forest*, sendo diferente apenas a técnica de aprendizado.

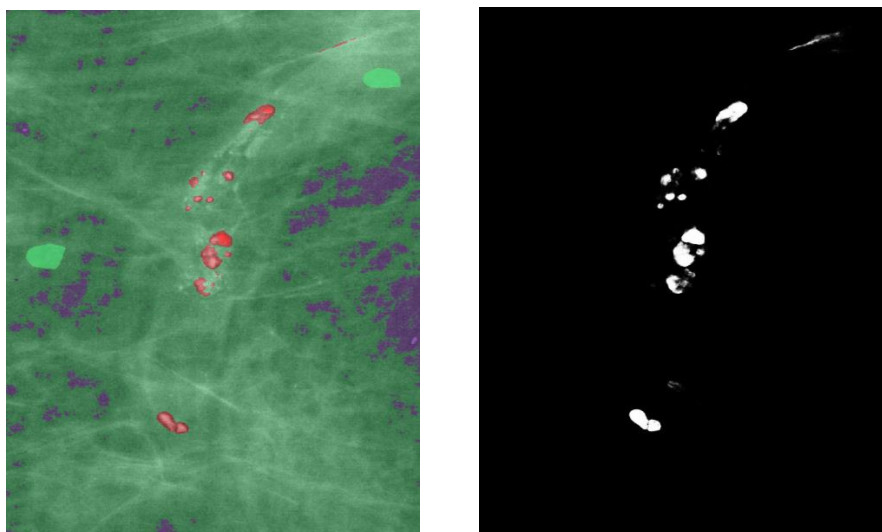


Figura 24– à esquerda: Predição utilizando Rede Neural; à direita: Mapa de probabilidade Rede Neural.

Fonte: Imagem adaptada de Lee *et al.* (2016).

A rede neural apresentou um bom resultado nesta metodologia de segmentação. Nota-se que a imagem do mapa de probabilidade apresenta valores mais próximos ao branco; sendo assim, menos ruídos de regiões suspeitas.

Os resultados da segmentação, apresentados nas Figuras 23 e 24, demonstram que é possível destacar as regiões suspeitas. Através destas imagens de probabilidade é possível realizar a detecção dos contornos das regiões; deste modo, podem ser realizados cortes destas regiões segmentadas e, posteriormente, realizar uma classificação automática dos alvos utilizando um algoritmo de aprendizado.

5.6 CONCLUSÕES

Evidenciou-se, neste estudo, relevância da técnica de segmentação inteligente associada ao exame da mamografia no diagnóstico do câncer mamário, pela eficácia de processamento de imagens e identificação da região suspeita. Nesse sentido, alcançou-se o objetivo inicial da pesquisa de averiguar que, de fato, essa técnica facilita a interpretação dos achados pelo médico auxiliando sobremaneira o diagnóstico precoce do câncer mamário.

Os resultados foram satisfatórios no teste da metodologia de segmentação, que procura solucionar problemas de modo similar ao cérebro humano, o que, no contorno de regiões suspeitas segmentadas, possibilitou cortes e, a partir dos quais, classificar os alvos, de modo automático, com suporte em um algoritmo de aprendizado.

O algoritmo Random Forest combina vários critérios de decisão para obter uma predição com maior acurácia, logo, permite achados mais confiáveis, o que facilita a interpretação por parte dos médicos, uma vez que regiões suspeitas são mais facilmente identificadas.

Mais estudos nessa área são necessários para o aperfeiçoamento da técnica, porém já é possível sonhar com um futuro em que máquinas e homens trabalham juntos para o favorecimento de uma população, principalmente quando falamos em câncer, uma doença cuja mortalidade ainda é grande e pouco se sabe sobre ela. Unir forças para o diagnóstico precoce é o maior desafio.

5.7 REFERÊNCIAS

ARGANDA-CARRERAS, I. *et al.* Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. **Bioinformatics**, Oxford, v. 33, n. 15, p. 2424-2426, 2017.

BOFF, R. A. *et al.* **Pocketbook da Mastologia**: um guia prático. 2. ed. Caxias do Sul, RS: São Miguel, 2019.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. 4th ed. Cranbury: Pearson, 2018.

HALL, M. *et al.* The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, New York, v. 11, n. 1, p. 10-18, 2009.

LEE, R. S. *et al.*. Curated breast imaging subset of DDSM. **The cancer imaging archive**, Little Rock, v. 8, 2016.

LEE, R. S. *et al.*. A curated mammography data set for use in computer-aided detection and diagnosis research. **Scientific data**, Londres, v. 4, p. 170-177, 2017.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, Cham, v. 5, p. 115-133, 1943.

OLIVEIRA, K. D.; OSELAME, G. B.; NEVES, E. B. Infertilidade após o tratamento oncológico. **Revista Médica Saúde**, Brasília, ano 1, n. 3, p. 72-84, 2014.

SANTOS, D. B.; SANTOS, M. A.; VIEIRA, E. M. Sexualidade e câncer de mama: uma revisão sistemática da literatura. **Saúde Soc.**, São Paulo, ano 4, n. 23, p. 1342-1355, 2014.

SCHINDELIN, J. *et al.* Fiji: an open-source platform for biological-image analysis. **Nature methods**, New York, v. 9, n. 7, p. 676, 2012.

SILVA, T. D. C. A; TAVARES, J. M. R. S. Algoritmos de segmentação de imagem e sua aplicação em imagens do sistema cardiovascular. *In*: CONGRESSO IBEROAMERICANO DE ENGENHARIA MECÂNICA, 10., 2011, Porto, Portugal. **Actas** [...]. Porto, Portugal: FEUP, 2011.

SOLOMON, C.; BRECKON, T. **Fundamentals of digital image processing: a practical approach with examples in matlab**. New Jersey: Wiley, 2011.

SOUTO, L. P. M. **Mineração de imagens para a classificação de tumores de mama**. 2013. Dissertação (Mestrado em Ciência da Computação) - Universidade do Estado do Rio Grande do Norte. Universidade Federal Rural do Semiárido, Mossoró-RN, 2013.

THULER, L. C., 1Considerações sobre a prevenção do câncer de mama feminino. **Revista Brasileira de Cancerologia**, Rio de Janeiro, v. 49, n. 4, p. 227-238, 2003.

VIEIRA, M. M. *et al.* Correlação clínica, mamográfica e histopatológica do câncer mamário em mulheres com idade entre 50 e 70 anos. **Mastology**, Campinas, v. 27, n. 2, p. 96-101, 2017.

VIEIRA, S. C. (org.). **Câncer de mama: Consenso da Sociedade Brasileira de Mastologia**, Regional Piauí. Teresina: EDUFPI, 2017.

6 CONSIDERAÇÕES FINAIS

A partir da elaboração dos três Artigos (A, B e C), tornou-se possível assentir o problema traçado na Introdução geral deste texto dissertativo sobre a cooperação com o diagnóstico precoce do câncer mamário utilizando-se das tecnologias digitais da IA. Com efeito, o processamento de imagens computacionais tornou-se um grande suporte ao médico, um caminho sem volta integrado às áreas Informática/Tecnologia e Medicina, uma vez que o avanço nessa interseção tem possibilitado diagnósticos precisos, contribuindo sobremaneira para o tratamento precoce da doença, que afeta um número considerável de mulheres no mundo, alcançando em muitos casos a cura e evitando a estas tratamentos traumáticos, muitas vezes, desnecessários.

O Artigo A destaca os Estados Unidos como o país de maior número de publicações a respeito desse tema, mas o número ainda é bem limitado – o que nos leva a inferir a necessidade da intensificação de mais estudos sobre a temática. O ano de 2006 sobressaiu-se quantitativamente em publicações a esse respeito, porém de lá para cá, houve um recuo nesse sentido.

O Artigo B aponta os avanços tecnológicos relativos à radiologia, coligada à limiarização/thresholding, que contribuem para as decisões médicas sobre diagnóstico e prevenção do câncer de mama. Nesse sentido, o tratamento da doença pode se iniciar precocemente, evitando o seu agravamento e possibilitando inclusive a cura da paciente.

O artigo C conseguiu demonstrar relevância da técnica de segmentação inteligente associada ao exame da mamografia no diagnóstico do câncer mamário, pela eficácia de processamento de imagens e identificação da região suspeita. Esse é um passo importante no diagnóstico precoce do câncer de mama, pois, de fato,

essa técnica facilita a interpretação dos achados pelo médico, auxiliando sobremaneira o diagnóstico e minimizando os erros. Claro que outros estudos nessa área são necessários para o aperfeiçoamento da técnica, porém um futuro em que máquinas e homens trabalham juntos não é mais algo tão distante.

A despeito de todo esse upgrade, são poucas as publicações científicas nesta seara, principalmente no Brasil, demandando, portanto, da academia científica maior visibilidade desta questão, inclusive para que ocorra um incessante refinamento das técnicas ora utilizadas em benefício de uma grande população vitimada e/ou vulnerável nessa conjuntura.

REFERÊNCIAS

American College of Radiology (ACR). **Atlas BI-RADS do ACR: Sistema de laudos e registro de dados de imagem da mama**. Tradução Angela Caracik. 2. ed. São Paulo, Colégio Brasileiro de Radiologia, 2016.

CALAS, M. J. G.; GUTFILEN, B.; PEREIRA, W. C. de A. CAD e mamografia: por que usar esta ferramenta? **Radiologia Brasileira**, São Paulo, v. 45, n. 1. p. 46-52, jan.-fev. 2012.

CUBUK, E. D. *et al.* AutoAugment: Learning Augmentation Strategies From Data. *In: PROCEEDINGS CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION*. 2018, Portugal. **Proceedings [...]**. Portugal: [s.n.], 2018, p. 113-123.

D'ORSI, C. J.; SICKLES, E. A.; MENDELSON, E. B.; MORRIS, E. A. **ACR BIRADS® Atlas, Breast Imaging Reporting and Data System**. Reston, Virginia: American College of Radiology, 2013.

HALL, M. *et al.* The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, New York, v. 11, n. 1, p. 10-18, 2009.

HUBBARD R. A.; KERLIKOWSKI, K.; FLOWERS, C. I.; YANKASKAS, B. C.; ZHU, W.; MIGLIORETTI, D. L. Cumulative Probability of False-Positive Recall or Biopsy Recommendation After 10 Years of Screening Mammography. *Ann Intern Med.*, Philadelphia, PA, v. 155, n. 8, p. 481-492, 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3209800/>. Acesso em: 20 jun. 2018.

Instituto Nacional do Câncer (INCA). **Conceito e Magnitude do câncer de mama**. 19 nov. 2018. Disponível em: <https://bit.ly/2VEsRit>. Acesso em: 28 abr. 2019.

Instituto Nacional do Câncer (INCA). **Abordagens básicas para o Controle do Câncer**. Rio de Janeiro: INCA, 2011. Disponível em: <https://is.gd/EB7zzz>. Acesso em: 24 jul. 2019.

OLIVEIRA, M. F. de. **Metodologia científica**: um manual para a realização de pesquisas em Administração. Catalão, UFG, 2011.

PULTZ, B. dos A. *et al.* **Investigação dos níveis séricos de EGFR1, HER-2, TSP-1 e RANKL em pacientes com câncer de mama inicial e avançado.** 91 f. 2015. Dissertação (Mestrado em Imunologia e Parasitologia Aplicadas) - Instituto de Ciências Médicas, Universidade Federal de Uberlândia, Uberlândia, MG, 2015.

RIBEIRO, K. R.; CRESPO, R. S.; MANHÃES, F. C. Indicadores bibliométricos da produção nacional e internacional sobre o tema 'evasão escolar' na base Scopus. 2016. Programa de Pós-graduação em Cognição e Linguagem. In: CONGRESSO LATINOAMERICANO EM HUMANIDADES, 12., 2016, Campos dos Goytacazes, RJ. **Anais [...]**. Campos dos Goytacazes, RJ,: [s.n.], 2016.

SANTANA, P. do C. **Desenvolvimento de metodologia para avaliação automatizada da qualidade de imagens digitalizadas em mamografia.** 113 f. 2010. Dissertação (Mestrado em Ciência e Tecnologia das Radiações, Minerais e Materiais) - Centro de Desenvolvimento de Tecnologia Nuclear (CDTN) Belo Horizonte, MG, 2010.

SIEGEL, R. L.; MILLER, K. D.; JEMAL, A. **Cancer Statistics.** 2018. Disponível em: <https://bit.ly/2u7nawL>. Acesso em: 12 jun. 2018.

SILVA, M. R.; LEAL, M. R. R.; LIMA, F. M. Predição do Câncer de Mama com aplicação de modelos de Inteligência Computacional. **Tendências em Matemática Aplicada e Computacional**, São Carlos, v. 20, n. 2, maio-ago. 2019.