



**UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**



KEILA LUCAS DOS SANTOS

**DEEPMAMO-BR: CLASSIFICAÇÃO INTELIGENTE DE
LESÕES MAMÁRIAS UTILIZANDO O PROTOCOLO
BI-RADS**

MOSSORÓ

2022

KEILA LUCAS DOS SANTOS

**DEEPMAMO-BR: CLASSIFICAÇÃO INTELIGENTE DE
LESÕES MAMÁRIAS UTILIZANDO O PROTOCOLO
BI-RADS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação - associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Marcelino Pereira dos Santos Silva

MOSSORÓ

2022

© Todos os direitos estão reservados a Universidade do Estado do Rio Grande do Norte. O conteúdo desta obra é de inteira responsabilidade do(a) autor(a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei nº 9.279/1996 e Direitos Autorais: Lei nº 9.610/1998. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu(a) respectivo(a) autor(a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

Catálogo da Publicação na Fonte.
Universidade do Estado do Rio Grande do Norte.

S237d Santos, Keila Lucas dos
DEEPMAMO-BR: Classificação Inteligente de Lesões Mamárias Utilizando o Protocolo BI-RADS. / Keila Lucas dos Santos. - Mossoró, 2022.
85p.

Orientador(a): Prof. Dr. Marcelino Pereira dos Santos Silva.

Dissertação (Mestrado em Programa de Pós-Graduação em Ciência da Computação). Universidade do Estado do Rio Grande do Norte.

1. Câncer de Mama. 2. BI-RADS. 3. Aprendizado Profundo. 4. Rede Neural Convolutacional. 5. Deep Cross-Training. I. Silva, Marcelino Pereira dos Santos. II. Universidade do Estado do Rio Grande do Norte. III. Título.



PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO



KEILA LUCAS DOS SANTOS

DEEPMAMO-BR: CLASSIFICAÇÃO INTELIGENTE DE LESÕES MAMÁRIAS UTILIZANDO O PROTOCOLO BI-RADS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação para a obtenção do título de Mestre em Ciência da Computação.

APROVADA EM: 27/07/2022

Marcelino Pereira dos Santos Silva:71864091487 Assinado digitalmente por Marcelino Pereira dos Santos Silva:71864091487
Dados: 2022.08.02 13:59:58 -03'00'

Prof. Dr. Marcelino Pereira dos Santos Silva
Orientador e Presidente

ARAKEN DE MEDEIROS SANTOS:00739803409 Assinado digitalmente por ARAKEN DE MEDEIROS SANTOS:00739803409
Data: 2022.08.02 11:45:00-03'00'

Prof. Dr. Araken de Medeiros Santos
Membro Interno - Universidade Federal Rural do Semi-Árido - UFERSA

Prof. Dr. Wellington Candeia de Araújo
Membro Externo - Universidade Estadual da Paraíba - UEPB

Aos meus pais e meus irmãos.

AGRADECIMENTOS

Ao Professor Dr. Marcelino Pereira por toda a orientação, dedicação e incentivo durante essa fase tão importante da minha formação acadêmica.

Aos professores, Dr. Araken de Medeiros e Dr. Wellington Candeia, membros da Banca Examinadora, por terem atendido ao convite, contribuindo significativamente na avaliação deste trabalho.

À Dra. Joana Rosa Nogueira pela disponibilidade e gentileza de esclarecer minhas dúvidas e de me orientar no estudo das mamografias, transmitindo seu importante conhecimento para o sucesso deste trabalho.

À minha mãe, Inês Lucas, ao meu pai, José Nunes, e aos meus irmãos, Raniery, Yuri e Uriel por todo suporte familiar, sem vocês não sou nada.

Aos meus colegas de mestrado, que compartilharam comigo essa experiência intensa da pós-graduação, especialmente ao meu amigo José Neto.

A todos os professores do PPGCC UERN/UFERSA que contribuíram imensamente na ampliação dos meus conhecimentos, vocês têm minha admiração e respeito.

À Universidade do Estado do Rio Grande do Norte pela concessão ao uso dos recursos computacionais do Laboratório LORDI. E ao técnico Mizaél Clístion por toda assistência remota durante os testes da pesquisa.

EPÍGRAFE

O mundo não mudou tanto quanto eu gostaria, mas eu sei que algum dia as coisas serão diferentes e as pessoas serão livres, mais livres do que são agora.

(Nora Helmer, Casa de Bonecas – Parte 2)

RESUMO

O câncer de mama é a neoplasia maligna mais comum entre as mulheres (exceto câncer de pele não melanoma), e o envelhecimento é um dos principais fatores de risco para o desenvolvimento da doença. De acordo com a Organização Mundial da Saúde, a estimativa mundial para o câncer de mama em 2020 foi definida em 2,3 milhões de mulheres com diagnóstico positivo e um total de 685.000 mortes. Na população brasileira, o risco estimado corresponde a 61,61 novos casos por ano a cada 100 mil mulheres. O rastreamento da doença por meio dos exames mamográficos favorece a detecção precoce dos primeiros sinais da doença e influencia na redução da taxa de mortalidade. No entanto, a detecção de lesões em imagens mamográficas é um processo complexo, no qual a precisão dos resultados pode ser afetada por fatores técnicos e biológicos. A ocorrência de falso-positivos e falso-negativos é uma das principais preocupações do imageamento por mamografia. Estima-se que os radiologistas ainda percam entre 10% e 30% dos nódulos com câncer. O desenvolvimento de sistemas *Computer-Aided Diagnosis* (CAD) integrados com as redes neurais profundas permite uma análise objetiva das imagens de mamografias e podem auxiliar o profissional médico na tomada de decisão do diagnóstico do câncer de mama. Dessa forma, esse trabalho tem como objetivo desenvolver uma metodologia e um modelo de Aprendizado Profundo para a classificação de mamografias, de acordo com as categorias de 1 a 5 do protocolo de avaliação BI-RADS, visando a identificação de lesões na mama. Os experimentos de trabalho foram realizados com as imagens de mamografias de duas bases de dados públicas, o CBIS-DDSM e o INbreast. A partir dos recursos das Redes Neurais Convolucionais na abordagem de classificação com *Deep Cross-Training* (DCTr) foi possível alcançar nos resultados de predição, uma acurácia média de 0,745 e Precisão de 0,752 nos modelos baseados na arquitetura ResNet-18, com uso das ROIs de imagens do banco CBIS-DDSM.

Palavras-chave: Câncer de Mama. BI-RADS. Aprendizado Profundo. Rede Neural Convolucional. Deep Cross-Training.

ABSTRACT

Breast cancer is the most common malignancy among women (except for non-melanoma skin cancer), and aging is one of the main risk factors for the development of the illness. According to the World Health Organization, the world estimate for breast cancer in 2020 was set at 2.3 million women with a positive diagnosis and a total of 685,000 deaths. In the Brazilian population, the estimated risk corresponds to 61.61 new cases per year per 100,000 women. The screening of the disease through mammographic exams favors the early detection of the first signs of the disease and influences the reduction of the mortality rate. However, the detection of lesions in mammographic images is a complex process, in which the precision of the results can be affected by technical and biological factors. The occurrence of false positives and false negatives is one of the main concerns of mammography imaging. It is estimated that radiologists still miss between 10% and 30% of cancerous nodules. The development of Computer-Aided Diagnosis (CAD) systems integrated with deep neural networks allow an objective analysis of mammogram images and can help the medical professional in the decision-making of the diagnosis of breast cancer. Thus, this work aims to develop a methodology and a Deep Learning model for classifying mammograms, according to categories 1 to 5 of the BI-RADS assessment protocol aimed at the identification of lesions in the breast. The work experiments were carried out with mammography images from two public databases, CBIS-DDSM and INbreast. From the resources of Convolutional Neural Networks in the classification approach with Deep Cross-Training (DCTr), it was possible to achieve the prediction results, an average accuracy of 0.745 and precision of 0.752 in models based on the ResNet-18, using the ROIs of images from the CBIS-DDSM database.

Keywords: Breast Cancer. BI-RADS. Deep Learning. Convolutional Neural Network. Deep Cross-Training.

LISTA DE FIGURAS

Figura 1 – Anatomia da mama feminina	18
Figura 2 – Evolução do câncer de mama	22
Figura 3 – Forma e margem de uma massa mamária	26
Figura 4 – Distribuição das calcificações	27
Figura 5 – Neurônio biológico	31
Figura 6 – Neurônio <i>Perceptron</i>	31
Figura 7 – Representação da Rede Neural	33
Figura 8 – Efeitos de transformações geométrica em imagens	34
Figura 9 – Operação de Convolução	36
Figura 10 – Arquitetura LeNet-5	36
Figura 11 – Arquitetura AlexNet	37
Figura 12 – Estrutura do Banco CBIS-DDSM	49
Figura 13 – Multiclasse - desenvolvimento do modelo (bloco pontilhado é opcional)	53
Figura 14 – DCTr - desenvolvimento do modelo (bloco pontilhado é opcional) . .	54
Figura 15 – Acurácia e Perda de Treinamento dos Modelos BI-RADS Multiclasse .	57
Figura 16 – Matriz de Confusão dos Modelos BI-RADS Multiclasse	58
Figura 17 – Curva ROC - Classificação BI-RADS com DCTr (ROI) [ResNet-18] .	65

LISTA DE TABELAS

Tabela 1 – Estágios do Câncer de Mama	21
Tabela 2 – Categorias de Avaliação BI-RADS	28
Tabela 3 – Strings de busca para a RSL	41
Tabela 4 – Aplicação dos Critérios de Exclusão da RSL	41
Tabela 5 – Descrição dos <i>Datasets</i>	51
Tabela 6 – Arquiteturas CNN's dos Modelos BI-RADS	52
Tabela 7 – Descrição dos Modelos para DCTr - Definição das classes	53
Tabela 8 – Classificação BI-RADS (Multiclasse)	58
Tabela 9 – Classificação BI-RADS com DCTr (<i>img full</i>)	61
Tabela 10 – Classificação BI-RADS com DCTr (<i>img ROI</i>)	62
Tabela 11 – Acurácia dos modelos - Classificação BI-RADS	62
Tabela 12 – Teste Friedman-Nemenyi	64
Tabela 13 – Pontuação F1 da Classificação BI-RADS com DCTr (<i>img ROI</i>)	66
Tabela 14 – Classificação BI-RADS com DCTr - ResNet-18	67
Tabela 15 – Comparação com estudos semelhantes	67
Tabela 16 – AlexNet - Teste Complementar 1	79
Tabela 17 – AlexNet - Teste Complementar 2	79
Tabela 18 – ResNet-18 - Teste Complementar 1	79
Tabela 19 – ResNet-18 - Teste Complementar 2	79
Tabela 20 – ResNet-18 - Teste Complementar 3	80
Tabela 21 – ResNet-18 - Teste Complementar 4	80
Tabela 22 – ResNet-18 - Teste Complementar 5	80
Tabela 23 – ResNet-34 - Teste Complementar 1	80
Tabela 24 – AlexNet - Teste Complementar 1 (DCTr)	80
Tabela 25 – AlexNet - Teste Complementar 2 (DCTr)	81
Tabela 26 – AlexNet - Teste Complementar 3 (DCTr)	81
Tabela 27 – AlexNet - Teste Complementar 4 (DCTr)	81
Tabela 28 – AlexNet - Teste Complementar 5 (DCTr)	81
Tabela 29 – ResNet-18 - Teste Complementar 1 (DCTr)	81
Tabela 30 – ResNet-18 - Teste Complementar 2 (DCTr)	82
Tabela 31 – ResNet-18 - Teste Complementar 3 (DCTr)	82

Tabela 32 – ResNet-18 - Teste Complementar 4 (DCTr)	82
Tabela 33 – ResNet-18 - Teste Complementar 5 (DCTr)	82
Tabela 34 – ResNet-18 - Teste Complementar 6 (DCTr)	82
Tabela 35 – ResNet-18 - Teste Complementar 7 (DCTr)	83
Tabela 36 – ResNet-18 - Teste Complementar 8 (DCTr)	83
Tabela 37 – ResNet-18 - Teste Complementar 9 (DCTr)	83
Tabela 38 – ResNet-18 - Teste Complementar 10 (DCTr)	83
Tabela 39 – ResNet-34 - Teste Complementar 1 (DCTr)	83
Tabela 40 – ResNet-34 - Teste Complementar 2 (DCTr)	84
Tabela 41 – ResNet-34 - Teste Complementar 2.2 (DCTr)	84
Tabela 42 – ResNet-34 - Teste Complementar 3 (DCTr)	84
Tabela 43 – ResNet-34 - Teste Complementar 4 (DCTr)	84
Tabela 44 – ResNet-34 - Teste Complementar 5 (DCTr)	84
Tabela 45 – ResNet-34 - Teste Complementar 6 (DCTr)	85
Tabela 46 – ResNet-34 - Teste Complementar 7 (DCTr)	85
Tabela 47 – ResNet-34 - Teste Complementar 8 (DCTr)	85
Tabela 48 – ResNet-34 - Teste Complementar 9 (DCTr)	85
Tabela 49 – ResNet-34 - Teste Complementar 10 (DCTr)	85

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	17
1.1.1	<i>Objetivo Geral</i>	17
1.1.2	<i>Objetivos Específicos</i>	17
1.2	Organização do Trabalho	17
2	REFERÊNCIAL TEÓRICO	18
2.1	Câncer de Mama e o Rastreamento Mamográfico	18
2.1.1	<i>O Câncer de Mama</i>	19
2.2	Mamografia	22
2.3	Classificação BI-RADS	25
2.4	Aprendizado de Máquina	29
2.4.1	<i>Aprendizado Profundo</i>	32
2.4.1.1	<i>Rede Neural Convolucional</i>	35
2.5	Considerações Finais	38
3	TRABALHOS RELACIONADOS	40
3.1	Considerações Finais	46
4	METODOLOGIA	47
4.1	Estudos Iniciais	47
4.2	Bases de Dados e Pré-Processamento	48
4.3	Arquiteturas dos Modelos	51
4.4	Desenvolvimento dos Modelos	52
4.5	Considerações Finais	55
5	RESULTADOS	56
5.1	Classificação BI-RADS Multiclasse	56
5.2	Classificação BI-RADS com <i>Deep Cross-Training</i> (DCTr)	59
5.2.1	<i>Teste de Friedman - Análise dos Classificadores</i>	63
5.2.2	<i>ResNet-18 - Análise em Curva ROC</i>	65
5.3	Considerações Finais	68
6	CONCLUSÕES	69
	REFERÊNCIAS	71

APÊNDICES	79
APÊNDICE A – Resultados dos Testes Complementares	79

1 INTRODUÇÃO

O câncer de mama é uma neoplasia maligna que se desenvolve nas células de revestimento dos ductos ou lóbulos no tecido glandular da mama. O comportamento da doença é semelhante aos outros tipos de câncer, no qual se observa a divisão celular contínua e desordenada, com capacidade de se disseminar e invadir outros órgãos (INCA, 2019b).

Em 2020, de acordo com a Organização Mundial da Saúde (OMS), a estatística mundial para o câncer de mama foi 2,3 milhões de mulheres com diagnóstico da doença, e um total de 685.000 mortes (WHO, 2021). No Brasil, as estimativas do câncer de mama entre os anos de 2020-2022 são de 66.280 novos casos da doença, segundo o Instituto Nacional do Câncer (INCA). O risco estimado equivale a 61,61 casos novos por ano a cada 100 mil mulheres (INCA, 2019a).

O câncer de mama é o tipo de câncer mais comum entre as mulheres (exceto câncer de pele não melanoma), podendo se desenvolver em qualquer idade após a puberdade, embora o envelhecimento seja um dos principais fatores de risco para a evolução da doença. De acordo com as Estatísticas Globais do Câncer em 2020, consolidada pela Agência Internacional de Pesquisa sobre Câncer (IARC), dos 2,26 milhões de novos casos estimados mundialmente, 70,5% incidem principalmente em mulheres acima dos 50 anos (IARC, 2021).

A detecção precoce das lesões é o principal objetivo das estratégias de controle do câncer de mama, pois quanto mais cedo um tumor invasivo for detectado, maior será a probabilidade de cura, e menos agressivos serão os tratamentos indicados. Uma importante medida de conscientização que permite o diagnóstico precoce é orientação para que as mulheres prestem atenção nas alterações suspeitas na mama, e busquem esclarecimento médico sobre as anormalidades o mais rápido possível (INCA, 2019b).

O Ministério da Saúde define em suas diretrizes que o rastreamento do câncer de mama seja executado através da realização de mamografias para mulheres de 50 a 69 anos, a cada dois anos (INCA, 2020). A mamografia é atualmente uma das melhores técnicas de detecção precoce de lesões não palpáveis na mama após a menopausa. Antes do período indicado, as mamas são mais densas e a sensibilidade da mamografia é reduzida, prejudicando a definição do diagnóstico.

O diagnóstico mamográfico é padronizado pelo protocolo *Breast Imaging-Reporting*

and Data System (BI-RADS). O protocolo BI-RADS foi desenvolvido em 1993 pelo Colégio Americano de Radiologia (ACR) com o objetivo de minimizar os erros de interpretação dos laudos de mamografias, ultrassonografia e ressonância magnética (CBR, 2016).

A interpretação das mamografias para a definição de um diagnóstico é um processo complexo, no qual a precisão dos resultados pode ser afetada por fatores técnicos (qualidade da imagem, contraste, erros na condução do exame), fatores biológicos (densidade da mama, distorção arquitetônica da lesão), além do nível de experiência e habilidade do radiologista. As altas taxas de falso-negativos (FNs) e falso-positivos (FPs) despertam bastante preocupação. Estima-se que devido às limitações da mamografia, os radiologistas ainda percam entre 10% e 30% dos nódulos com câncer e que, em média, 80% das lesões (BI-RADS 4) biopsiadas sejam benignas, ou seja, o procedimento só seria realmente necessário em 20% dos casos (EKPO *et al.*, 2018; NAZÁRIO *et al.*, 2015).

O procedimento de biópsia é um exame cirúrgico que consiste na retirada de um pequeno nódulo suspeito, em qualquer parte do corpo, para análise laboratorial. Geralmente é uma intervenção simples, mas ainda se trata de um procedimento invasivo que causa ansiedade e desconforto ao paciente (MOORE-HIGGS; LIGIBEL, 2009). Quando biópsias são requisitadas em virtude de falso-positivos, além dos transtornos causados ao paciente, custos adicionais são gerados ao processo de rastreamento do câncer de mama, tornando-se um desperdício de recursos financeiros na área da saúde pública.

A discussão em torno dos casos falso-positivos e da requisição de biópsias dispensáveis proporciona base para diversos estudos recentes (AGARWAL *et al.*, 2020; CARNEIRO *et al.*, 2017; COGAN *et al.*, 2019) que aplicam as técnicas da Inteligência Artificial (IA) para o desenvolvimento de sistemas que auxiliam no diagnóstico do câncer. Esses sistemas são definidos como aplicações de diagnóstico auxiliado por computador (*Computer-Aided Diagnosis - CAD*), que apresentam recursos de suporte à análise de imagens médicas, permitindo a identificação de padrões, gerando um alerta para regiões suspeitas detectadas nas imagens (CALAS *et al.*, 2012; DOI, 2007). No que se refere ao câncer de mama, um sistema CAD, em termos gerais, atua com um método quantitativo auxiliar que complementa a análise médica do radiologista, ampliando a qualidade do diagnóstico das lesões mamárias.

Em virtude das limitações dos sistemas CAD tradicionais, os pesquisadores investigam melhorias na detecção auxiliada por computador aplicando métodos de Aprendizado

Profundo (*Deep Learning* - DL), com foco principalmente nas Redes Neurais Convolucionais (*Convolutional Neural Network* - CNN), que são uma classe de algoritmos de Aprendizado Profundo capazes de extrair recursos visuais, ajustando pesos para a diferenciação entre imagens. No contexto das imagens médicas, o aprendizado das CNNs em mamografias apresenta resultados notáveis na detecção de lesões suspeitas e na classificação do risco de câncer de mama, fornecendo uma análise quantitativa que auxilia os radiologistas a definirem diagnósticos mais precisos (ABDELHAFIZ *et al.*, 2019).

Tendo em vista a relevância desta temática para a saúde, bem como o conjunto de esforços para concepção de sistemas capazes de rastrear o câncer de mama, há evidências de que abordagens utilizando métodos computacionais de Processamento Digital de Imagens e Inteligência Artificial podem avançar o estado da arte nesta área.

1.1 Objetivos

1.1.1 *Objetivo Geral*

Este trabalho tem como objetivo desenvolver uma metodologia e um modelo de Aprendizado Profundo, com recursos das Redes Neurais Convolucionais, para a classificação de lesões na mama em imagens de mamografias, de acordo com as categorias de 1 a 5 da avaliação BI-RADS para o diagnóstico do câncer de mama.

1.1.2 *Objetivos Específicos*

Destaca-se como objetivos específicos deste trabalho:

- Investigar aspectos relevantes do Processamento Digital de Imagens para a análise de mamografias;
- Desenvolver um modelo inteligente para a classificação das imagens de mamografias, conforme o protocolo BI-RADS.
- Avaliar o aprendizado das Redes Neurais Convolucionais em relação à classificação BI-RADS;
- Identificar a melhor abordagem de classificação para a predição das imagens de mamografias de campo completo e de regiões de interesse.
- Avançar no estado da arte em soluções de Aprendizado Profundo para a classificação BI-RADS de lesões na mama.

1.2 Organização do Trabalho

No Capítulo 2 é apresentado o Referencial Teórico acerca do câncer de mama, Protocolo BI-RADS, Aprendizado Profundo e Redes Neurais Convolucionais para a classificação de mamografias. No Capítulo 3 descreve-se os trabalhos relacionados com os objetivos de pesquisa, avaliados por Revisão Sistemática da Literatura. No Capítulo 4 é descrita a metodologia de trabalho, com destaque para as etapas de pré-processamento das imagens e desenvolvimento dos modelos de classificação. No Capítulo 5 são apresentados os resultados do trabalho, com as respectivas discussões sobre os experimentos de Classificação Multiclasse e da Classificação com *Deep Cross-Training* (DCTr) das imagens de mamografias. Por fim, as considerações finais são relatadas no Capítulo 6.

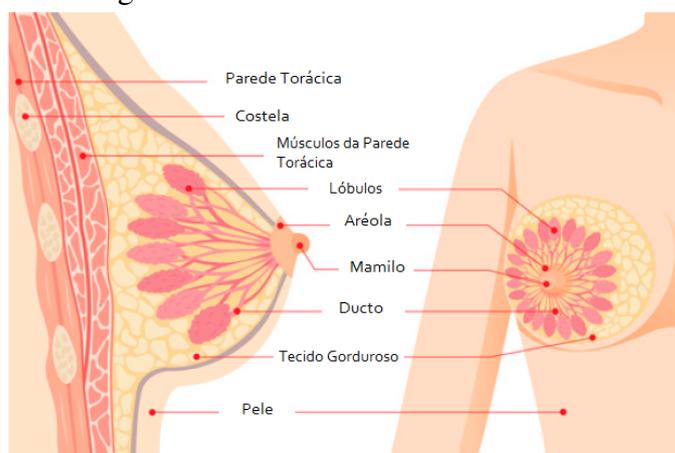
2 REFERÊNCIAL TEÓRICO

Neste capítulo serão apresentados os conteúdos que fundamentam a proposta deste trabalho, destacando-se os aspectos relacionados ao câncer de mama, exame de mamografia, sistema de padronização de laudos mamográficos (Protocolo BI-RADS), Aprendizado de Máquina e Aprendizado Profundo.

2.1 Câncer de Mama e o Rastreamento Mamográfico

A estrutura anatômica da mama feminina é formada por tecido adiposo, glândulas produtoras de leite, denominadas de lóbulos, ductos que transportam o leite até o mamilo, e aréola, a região em torno do mamilo. Na Figura 1 está representada a ilustração dessas partes biológicas, destacando-se a parede torácica, ossos das costelas, músculos da parede torácica, lóbulos, aréola, mamilo, ductos, tecido adiposo e pele (SENKUS-KONEFKA *et al.*, 2018).

Figura 1 – Anatomia da mama feminina



Fonte: Senkus-Konefka *et al.* (2018)

O processo de desenvolvimento das mamas humanas se inicia na fase embrionária, em torno da quinta semana de formação do feto, quando uma "estria de leite" se desenvolve ao longo do tronco em ambos os lados da axila à virilha. Em muitos mamíferos, uma série de glândulas mamárias se desenvolve ao longo dessa estria, mas no gênero masculino ela regride até o local definitivo do mamilo adulto. Durante a fase da puberdade ocorre a maior diferenciação na formação das glândulas mamárias. Nesse período, os ductos começam a proliferar e suas terminações formam massas sólidas de células, dando origem aos lóbulos

mamários (ELLIS; MAHADEVAN, 2013; JATOI *et al.*, 2005).

Durante a fase da gravidez, as mamas femininas sofrem novas transformações anatômicas. A partir da fecundação, toda a estrutura mamária se prepara para a produção de leite. Os tecidos ficam mais espessos, aparecem os alvéolos secretores, ocorre o brotamento ductal e a proliferação lobular, com aumento da pigmentação mamilar e areolar (ELLIS; MAHADEVAN, 2013).

Naturalmente, em torno dos 50 anos de idade, a mulher atinge a menopausa, e assim se iniciam outras transformações no corpo feminino. Na menopausa ocorre a interrupção da menstruação, pois os ovários deixam de produzir os hormônios estrogênio e progesterona. Nesse período, observa-se nas mamas que o tecido glandular atrofia, o tecido conjuntivo torna-se menos celular e ocorre uma diminuição na quantidade de colágeno (JAMESON *et al.*, 2020).

O período pós-menopausa requer um acompanhamento sistemático da saúde feminina, em especial as mamas. A partir dos 50 anos, é muito importante que as mudanças na região mamária sejam periodicamente avaliadas, pois o envelhecimento é um dos mais importantes fatores para o desenvolvimento de anormalidades, principalmente do câncer de mama, que é uma neoplasia maligna que todos os anos afeta milhões de mulheres no mundo.

2.1.1 O Câncer de Mama

O câncer de mama é uma doença caracterizada pelo crescimento anormal de células da mama, que se multiplicam formando tumores no tecido mamário. Esse é o tipo de câncer mais comum no gênero feminino, com maior incidência em mulheres acima de 50 anos de idade. De acordo com o relatório do GLOBOCAN 2020 produzido pela IARC, foram estimados cerca de 19,3 milhões de novos casos de câncer (18,1 milhões excluindo câncer de pele não melanoma) no mundo, destacando-se 2,3 milhões (11,7%) de novos casos de câncer de mama feminino, como o mais comumente diagnosticado, ultrapassando os casos de câncer de pulmão. Além disso, a mortalidade por câncer de mama foi estimada em 684.996 novos casos (SUNG *et al.*, 2021).

Os dados consolidados pela IARC indicam que o câncer de mama é o mais diagnosticado em 159 países, seguido do câncer cervical (câncer do colo do útero) que é o mais diagnosticado em 23 países. O Brasil é um dos países no qual o câncer de mama é

mais incidente entre as mulheres, sendo umas das principais causas de morte prematura. As estimativas da doença para cada ano do triênio 2020-2022 são de 66.280 novos casos da doença (29,7% de incidência dentre os dez tipos de câncer mais comuns), segundo o Instituto Nacional do Câncer (INCA). Observando-se um risco estimado de 61,61 casos novos a cada 100 mil mulheres (INCA, 2019a).

Analisando pela perspectiva regional é possível identificar que a incidência do câncer de mama se distribui em maior número de casos na região Sudeste (36.470), seguida pelas regiões Nordeste (13.190), Sul (10.890), Centro-Oeste (3.760) e Norte (1.970). Além do câncer de mama, nessas regiões também se destacam os casos de câncer de colo do útero e colorretal, mesmo em menor número, que também têm impacto importante na população feminina (INCA, 2019a).

A partir dos 40 anos de idade, a incidência do câncer de mama tende a crescer progressivamente, assim como a taxa de mortalidade pela doença, que geralmente está associada ao diagnóstico tardio. De acordo com o Sistema de Informações sobre Mortalidade (SIM) do Ministério da Saúde, o número de mortes por câncer de mama em 2019 foi 18.295, sendo 18.068 mulheres e 227 homens. As regiões Sudeste e Sul apresentaram as maiores taxas, de 16,14 e 15,08 óbitos/100.000 mulheres, respectivamente (SIM, 2019).

De acordo com a Histopatologia, o câncer de mama é dividido em duas categorias principais, dependendo do envolvimento do sistema ductal-lobular da mama: Não Invasivos e Invasivos. Os cânceres de mama não invasivos ficam dentro dos ductos de leite e dos lóbulos da mama. Enquanto os cânceres invasivos se espalham além dessas áreas e invadem outros tecidos externos à região mamária, provocando um quadro clínico conhecido como metástase. O principal tipo não invasivo é o Carcinoma Ductal In Situ (CDIS), no qual o câncer se desenvolve nos ductos que transportam o leite para o mamilo. Semelhante ao CDIS, ocorre o Carcinoma Lobular In Situ (CLIS), que começa nas glândulas produtoras de leite, denominadas de lóbulos (AYDINER *et al.*, 2019).

Em relação aos carcinomas invasivos, eles podem ser classificados como "nenhum tipo específico" quando as características morfológicas dos tumores não são suficientes para serem determinadas como um tipo histológico característico, e também podem ser classificados como um "tipo especial", nos quais se apresentam características distintivas suficientes, como comportamento celular e molecular particular. Dentre os tipos especiais mais comuns de câncer de mama se destacam o carcinoma medular, carcinoma metaplásico,

carcinoma apócrino, carcinoma mucinoso, carcinoma cribriforme, carcinoma tubular, carcinoma neuroendócrino, carcinoma lobular clássico e carcinoma lobular pleomórfico (NASCIMENTO; OTONI, 2020).

O câncer de mama invasivo evolui e se espalha para fora dos ductos (Ductal Invasivo) ou dos lóbulos (Lobular Invasivo). Na Tabela 1 observa-se os estágios de evolução do câncer de mama e a descrição referente à morfologia do tumor. Percebe-se que a partir do estágio T1 (T1a, T1b e T1c) o tumor vai aumentando em centímetros até atingir o estágio T4 (câncer avançado), também conhecida como fase metastática, no qual o câncer se espalha para outros locais do corpo, como fígado, pulmões, ossos, cérebro ou outros tecidos.

Tabela 1 – Estágios do Câncer de Mama

Tumor Primário (T)	Descrição
T0	Sem evidência de tumor primário
TIS	Carcinoma In Situ
T1	$Tumor \leq 2cm$
T1a	$0.1cm < Tumor \leq 0.5cm$
T1b	$0.5cm < Tumor \leq 1cm$
T1c	$1cm < Tumor \leq 2cm$
T2	$2cm < Tumor \leq 5cm$
T3	$Tumor > 5cm$
T4	Extensão da parede torácica, inflamação, lesões satélites, ulcerações

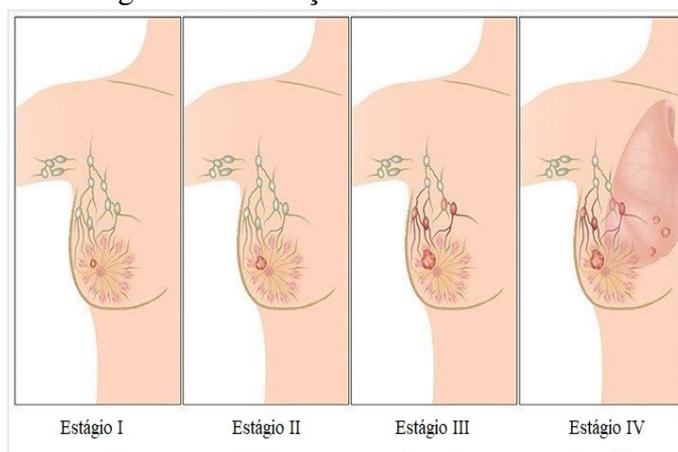
Fonte: Jameson *et al.* (2020)

Na Figura 2 é apresentado o processo de evolução do câncer através dos tecidos da mama, se espalhando dos ductos, aglomerando nos gânglios linfáticos, até atingir outros tecidos fora da mama.

Os sintomas mais comuns do câncer são alterações nas mamas, geralmente caracterizadas por ondulação ou espessamento da pele, secreção do mamilo, vermelhidão, dor ou desconforto recorrente, e também pode ser percebida a presença de um caroço. É difícil definir a causa principal para o surgimento do câncer de mama, pois vários fatores de risco podem influenciar o surgimento da doença. Dentre os principais fatores de risco destacam-se o gênero feminino, aumento da idade, predisposição genética e obesidade (SENKUS-KONEFKA *et al.*, 2018).

Segundo Pasqualini (2008), as influências genéticas são responsáveis por pelo

Figura 2 – Evolução do câncer de mama



Fonte: EDH (2018)

menos 5% dos casos de câncer de mama. O histórico familiar é um importante indicador da pré-disposição ao desenvolvimento da doença. A ocorrência de casos em outros parentes de primeiro grau pode destacar a existência de um fator genético que amplia a probabilidade do câncer de mama. O risco genético é evidenciado principalmente pela mutação dos genes BRCA1 e BRCA2, que aumentam muito as chances de se desenvolver o câncer de mama e de ovário. Uma mulher portadora de uma mutação BRCA1 tem um risco de 65 - 95% de desenvolver o câncer de mama (PALUCH-SHIMON *et al.*, 2016).

Observar as primeiras alterações físicas, ou fazer um mapeamento genético, permite o reconhecimento do câncer antes do agravamento do caso. Com auxílio médico, as investigações iniciais do câncer de mama se iniciam através de um exame físico, geralmente por mamografia ou ultrassom. Em alguns casos específicos, as mamas também podem ser avaliadas por ressonância magnética.

O exame de mamografia é a base do rastreamento nacional do câncer de mama, de acordo com as diretrizes do Ministério da Saúde. A execução desse exame é recomendada para mulheres de 50 a 69 anos, a cada dois anos, e atualmente a mamografia é uma das melhores técnicas para detecção precoce de lesões não palpáveis na mama (INCA, 2020).

2.2 Mamografia

Existe uma ampla variedade de técnicas de imageamento que são usadas para fins de triagem e detecção prévia de doenças. Dentre os principais tipos de imagem diagnóstica se destacam o raio-X, ultrassom, tomografia computadorizada e ressonância magnética. No contexto do câncer de mama, o diagnóstico não invasivo por meio de imagem é realizado

principalmente através da mamografia, que é uma modalidade de raio-X, e do ultrassom mamário (BANSAL *et al.*, 2013).

Os primeiros experimentos de exame radiográfico da mama ocorreram em meados de 1913, quando o cirurgião Albert Salomon relatou em suas investigações o uso de radiografia em espécimes de mastectomia para demonstrar a disseminação do tumor para os linfonodos axilares. Em 1931, Walter Vogel relatou para a *Leipzig Medical Society* uma classificação radiográfica de lesões benignas da mama, e os aspectos que as tornavam diferentes do carcinoma. E anos depois, na década de 1960, Robert Egan apresentou os primeiros 53 casos ocultos de câncer de mama que foram detectados em 2.000 mamografias (GOLD *et al.*, 1990).

A mamografia é uma imagem de raio-X utilizada para detectar os primeiros sinais do câncer de mama. Esse exame permite a avaliação de pacientes com achados clínicos anormais, como um nódulo na mama ou secreção no mamilo, e em casos assintomáticos a mamografia permite a detecção precoce do câncer, até três anos antes que ele possa ser sentido (CDC, 2020).

A obtenção da imagem de mamografia envolve a exposição das mamas a uma pequena dose de radiação ionizante para captar o interior da região mamária. Durante o procedimento, um feixe de raios-X incide sobre a mama pressionada no compressor do mamógrafo, sendo parte dele transmitido através do tecido e captado por um receptor de imagem. Outra parte do feixe sofre um processo de espalhamento, não contribuindo para a formação da imagem. E a fração restante é absorvida pelos tecidos da mama (INCA, 2019b).

Quando a mamografia foi introduzida como modalidade de imageamento, eram usados como receptor de imagem os filmes radiográficos de exposição direta, sem tela intensificadora, e a imagem era gerada por processo químico, semelhante a negativo de fotografia. Posteriormente, os filmes radiográficos foram associados às telas fluorescentes de alta resolução para melhorar a qualidade da imagem. A combinação tela-filme ainda é bem comum para a realização das mamografias (INCA, 2019b).

Com o avanço da tecnologia foram desenvolvidos os sistemas de radiografia computadorizada (*Computerized Radiography* - CR) e os sistemas de radiografias digitais (*Digital Radiology* - DR). Nos sistemas CR utiliza-se placas de material fotoestimulável, à base de fósforo para o registro da imagem, que será digitalizada para um computador

por meio de uma leitora CR. Nos sistemas DR a imagem é armazenada por meio de sinais eletrônicos e transferidas diretamente para um computador (JALES, 2015).

Os sistemas DR apresentam maior qualidade da imagem com menor dose de radiação, em comparação com os sistemas CR. As mamografias produzidas em sistemas DR, também conhecidas como mamografia digital de campo total (*Full-Field Digital Mammography* - FFDM), além de apresentar melhor qualidade de imagem, são processadas com mais rapidez, e são mais fáceis de serem manipuladas e armazenadas. No entanto, o custo elevado para implantar e realizar a manutenção de um centro de mamografias digitais é uma grande limitação para expandir o rastreamento do câncer de mama com essa modalidade de imagem (KÖRNER *et al.*, 2007; LBN Medical, 2020).

O exame mamográfico é usado em dois tipos de abordagens, para o rastreamento do câncer e para o diagnóstico da doença. A mamografia de rastreamento é realizada em mulheres sem sinais ou sintomas do câncer de mama, com o principal objetivo de detectar precocemente pequenas lesões. A mamografia diagnóstica é indicada para as mulheres que já apresentam sintomas do câncer, e a realização do exame serve para se confirmar ou afastar a suspeita da doença (INCA, 2019b).

De acordo com a Sociedade Americana de Câncer (*American Cancer Society*), a mamografia é a melhor técnica para o rastreamento do câncer de mama na atualidade (ACS, 2019). No entanto, os falso-positivos e os falso-negativos decorrentes da análise são a maior limitação do exame. Em uma mamografia falso-positiva é indicada a existência de anormalidade, embora não haja anormalidade de fato naquela região mamária. Enquanto em uma mamografia falso-negativa, existe a anormalidade, mas ela não foi percebida. Essa situação é bastante preocupante, pois a evolução do câncer sem ser percebido prejudica as chances de sucesso em tratamentos futuros.

O rastreamento do câncer de mama a cada dois anos reduz o risco de morte pela doença, porque o diagnóstico precoce favorece o início do tratamento no momento adequado, com maiores chances de cura, e sem a necessidade de métodos muito agressivos, como a quimioterapia. Entretanto, nas mamografias de rastreamento geralmente ocorre falha de 1 em 5 casos de câncer de mama. A maior probabilidade de ser definido um resultado falso-negativo ocorre principalmente em seios densos, pois existe uma maior proporção de tecido fibroglandular que pode ocultar as anormalidades, e em mulheres com mutação no gene supressor do câncer mamário, o BRCA 1 e BRCA 2. Ademais, é possível

que o câncer se desenvolva entre o intervalo bienal das mamografias sem ter sido detectado no primeiro exame (ACS, 2019).

As mamografias com anormalidades geralmente exigem exames complementares, como ultrassonografia, ressonância magnética ou até mesmo a realização de biópsia da mama para se confirmar o diagnóstico. Resultados falso-positivos podem causar transtornos de ansiedade e impulsionar a realização de um procedimento de biópsia, que ocasionará desconforto físico e custos financeiros extras. Quando existem mamografias anteriores disponíveis para comparação é possível reduzir as suas chances de ser definido um resultado falso-positivo, pois o médico consegue observar a simetria das mamas através de diferentes imagens da mesma região (HAYWARD *et al.*, 2016).

É importante ressaltar também o sobrediagnóstico (*overdiagnosis*) de um câncer. Essa definição médica se refere à detecção de um câncer que nunca vai evoluir ou se espalhar, ou seja, uma mulher recebe um diagnóstico de câncer de mama que nunca causará efeitos em sua saúde. Mesmo que esse tipo de câncer não cause danos, os médicos não conseguem distingui-lo dos outros tipos que evoluem e se espalham, e de alguma forma a paciente será submetida a um tratamento e sofrerá os efeitos colaterais dele (ACS, 2019).

O diagnóstico da mamografia é orientado pela classificação BI-RADS, que foi desenvolvida pelo Colégio Americano de Radiologia em 1993, formalizando uma linguagem padronizada para a interpretação de exames de imagens de mamografia, ultrassonografia e ressonância magnética. Após a criação inicial do BI-RADS em 1993, mais quatro edições foram publicadas em 1995, 1998, 2003 e 2013, respectivamente. A edição mais recente (5ª edição) fornece a especificação de termos e representações gráficas da anatomia mamária e de patologias para auxiliar no processo de avaliação do câncer de mama.

2.3 Classificação BI-RADS

A classificação BI-RADS é utilizada na maioria dos países que possuem o rastreamento do câncer de mama implementado no sistema de saúde. A linguagem padronizada pelo protocolo reduz a variabilidade entre os radiologistas na definição dos laudos em mamografia, ultrassonografia ou ressonância magnética (BALLEYGUIER *et al.*, 2007).

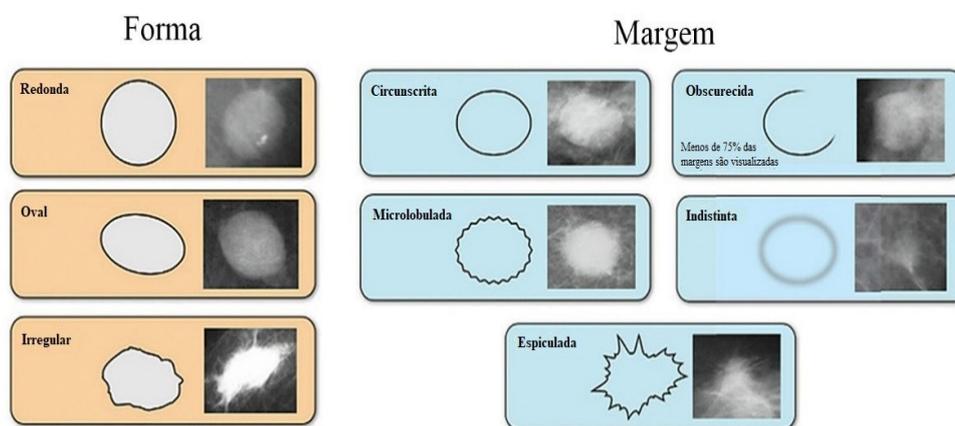
A 5ª edição do BI-RADS foi desenvolvida a partir da revisão de trabalhos anteriores, com o objetivo de esclarecer termos precedentes, e de incluir descritores de ultrassom e

ressonância magnética para ampliar a coerência entre as modalidades. O manual publicado apresenta diagramas descritivos das anomalias, recomendações para o relatório de avaliação e exemplos de casos mamográficos (SPAK *et al.*, 2017).

No manual BI-RADS, a visão geral do léxico de imagens da mama em mamografias apresenta descrições acerca da composição da mama e dos achados que são denominados de massas, calcificações, distorção arquitetural, assimetrias, linfonodo intramamário, lesão na pele e ducto dilatado solitário. Os achados mamários podem ser identificados em ambas as mamas, por isso que a mamografia de rastreamento segue o imageamento da lateralidade esquerda e direita, de acordo com as visualizações Craniocaudal (CC) e a Mediolateral-Oblíqua (MLO) (CBR, 2016).

Uma massa é um agrupamento celular tridimensional que ocupa espaço. Esse tipo de achado é definido em relação à sua forma, margem e densidade. A forma é categorizada em redonda, oval ou irregular. Enquanto a margem pode ser definida como circunscrita, obscurecida, microlobulada, indistinta e espiculada, como representado na Figura 3.

Figura 3 – Forma e margem de uma massa mamária



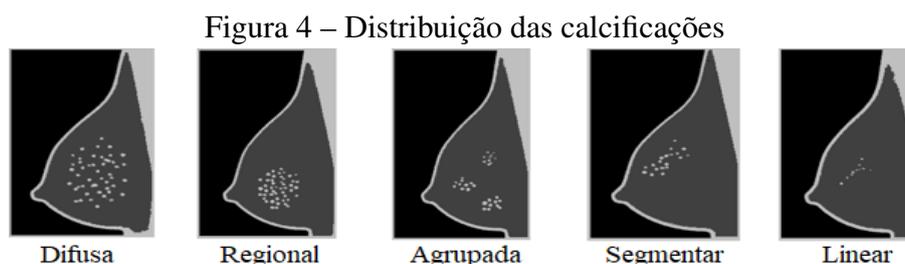
Fonte: Adaptada de RADASS-UZ (2015)

A densidade de uma massa em relação ao tecido mamário fibroglandular pode ser descrita como densidade alta, densidade igual, densidade baixa e contendo gordura. Quanto mais irregular for a forma de massa, maior a probabilidade de malignidade. As margens espiculadas e indistintas também possuem maior probabilidade de malignidade, assim como as massas com densidade mais alta (TANG *et al.*, 2001; BALLEYGUIER *et al.*, 2007).

As calcificações são pequenos depósitos de cálcio que aparecem nas mamografias como manchas ou pontos brancos brilhantes no tecido mamário. Essas estruturas são

avaliadas como benignas quando geralmente são maiores, mais grosseiras, redondas com margens regulares, sendo mais facilmente vistas do que as calcificações malignas. As calcificações associadas à malignidade (e também muitas calcificações benignas) são muito pequenas, e requerem o uso de ampliação para serem visualmente identificadas (CBR, 2016).

A distribuição das calcificações na mama é classificada como difusa, regional, agrupada, segmentar e linear, conforme representado na Figura 4. As calcificações difusas são distribuídas aleatoriamente por toda a mama. As regionais ocupam uma grande parte do tecido mamário, enquanto as agrupadas se apresentam em pequenas porções do tecido. As lineares são dispostas em linha, e podem levantar suspeita de malignidade, pois o depósito pode estar ocorrendo em um ducto, assim como as calcificações segmentares, pois sugerem que os depósitos são em um ducto ou ductos e suas ramificações (CBR, 2016).



Fonte: Adaptada de Campos *et al.* (2013)

Nas descrições do laudo mamário, de acordo com as recomendações do manual BI-RADS, é preciso que seja indicada a composição geral da mama e a avaliação acerca dos achados visualizados na mamografia. A composição da mama, também denominada de densidade, é classificada em quatro tipos: Mamas predominantemente adiposas, mamas com densidades fibroglandulares esparsas, mamas heterogeneamente densas (pode ocultar pequenos nódulos) e mamas extremamente densas (diminui a sensibilidade da mamografia).

A densidade mamográfica é perceptível pela luminosidade relativa de uma mamografia. As mamas densas bloqueiam os raios-X em maior extensão, por isso ficam bem mais claras na imagem, em comparação com as mamas adiposas. Essa característica tende a camuflar o câncer de mama e as microcalcificações, que geralmente aparecem como áreas mais claras nas mamografias, ampliando a taxa de interpretações falso-positivas e falso-negativas. A densidade da mama varia ao longo do ciclo menstrual, apresentando-se com maior ou menor densidade de acordo com as alterações hormonais, e também varia

de acordo com idade da mulher, pois mulheres mais jovens possuem mamas mais densas (IMNRC, 2005).

Em relação a avaliação dos achados em mamografia, no protocolo BI-RADS são padronizadas sete categorias (0, 1, 2, 3, 4, 5 e 6) de avaliação, as quais possuem percentagem representativa em relação à suspeita de malignidade. Cada categoria possui uma conduta médica vinculada para o acompanhamento do câncer de mama, conforme descrito na Tabela 2. A interpretação médica acerca da distinção das lesões entre categorias representativas não é um procedimento simples, principalmente para a categoria 4, que é definida por subcategorias (4A, 4B e 4C) de suspeita de malignidade.

Tabela 2 – Categorias de Avaliação BI-RADS

Avaliação	Conduta	Probabilidade de Câncer
Categoria 0: Incompleta	Avaliação de imagem adicional e / ou comparação com mamografias anteriores.	Avaliação incompleta, sendo recomendada a realização de avaliações adicionais
Categoria 1: Negativa	Rastreamento mamográfico de rotina	0% de probabilidade de malignidade
Categoria 2: Achado(s) benigno(s)	Rastreamento mamográfico de rotina	Essencialmente 0% de probabilidade de malignidade, achado benigno.
Categoria 3: Achado(s) provável - mente benignos	Seguimento de curto prazo (6 meses) ou acompanhamento mamográfico periódico.	Entre 0-2% de probabilidade de malignidade
Categoria 4: Achado(s) suspeito(s)		Entre 2-95% de probabilidade de malignidade
Categoria 4A: Baixa suspeita de malignidade	Diagnóstico tecidual	Entre 2 e 10% de probabilidade de malignidade
Categoria 4B: Moderada suspeita de malignidade		Entre 10 e 50% de probabilidade de malignidade
Categoria 4C: Alta suspeita de malignidade		Entre 50 e 95% de probabilidade de malignidade
Categoria 5: Achado(s) altamente sugestivo(s) de malignidade	Diagnóstico tecidual	>= 95% de probabilidade de malignidade
Categoria 6: Malignidade comprovada por biópsia	Excisão cirúrgica quando clinicamente apropriado	100% de probabilidade de malignidade

Fonte: Adaptada de CBR (2016)

Durante uma avaliação mamográfica vários parâmetros são utilizados para avaliar os sinais do câncer de mama em um paciente. Detectar se existe ou não uma lesão, verificar o tamanho e a capacidade proliferativa dos tumores, analisar o status dos linfonodos axilares, entre outros, são procedimentos que tornam o processo de definição do diagnóstico do câncer de mama uma tarefa complexa (AYDINER *et al.*, 2019).

Muitos fatores perceptivos e interpretativos limitam a avaliação radiológica. Falhas técnicas e/ou humanas impulsionam a determinação de falso-positivos e falso-negativos. Além disso, as características biológicas inerentes ao tumor, assim como a composição da mama, também são destacáveis desafios para o diagnóstico do câncer na mama (GIESS *et*

al., 2012).

Uma alternativa para reduzir as falhas avaliativas e tornar o processo de diagnóstico mais preciso é utilizar técnicas computacionais para aprimorar o reconhecimento das lesões nas imagens de mamografias. Os recursos da IA, especialmente os diversos métodos do Aprendizado de Máquina e do Aprendizado Profundo, proporcionam que ferramentas de diagnóstico médico sejam desenvolvidas para tornar as avaliações de imagens mais precisas, possibilitando uma redução nas taxas de falso-positivos e falso-negativos durante o processo clínico.

2.4 Aprendizado de Máquina

O Aprendizado de Máquina é uma área da Inteligência Artificial que investiga métodos para viabilizar que sistemas computacionais adquiram conhecimento a partir de dados e sejam capazes de melhorar automaticamente com a experiência. O direcionamento de trabalho na área do Aprendizado de Máquina se concentra na utilização de técnicas e métodos que permitem a extração de recursos dos dados, conduzindo o processo para a descoberta de conhecimento. E o conhecimento extraído de bases de dados auxilia a tomada de decisão em diversos contextos, como a detecção de fraudes, reconhecimento de imagens, previsão de eventos críticos, recomendação de produtos, entre outros (HAN *et al.*, 2011).

O Aprendizado de Máquina é caracterizado em quatro paradigmas de aprendizado: Supervisionado, Não-Supervisionado, Semissupervisionado e Reforço. O aprendizado supervisionado é o processo pelo qual a máquina aprende a partir de dados rotulados. Nesse paradigma, o algoritmo recebe um conjunto de entradas e suas respectivas saídas corretas, e aprende os valores de acordo com os padrões existentes, além de comparar os valores aprendidos com os valores reais, pontuando os erros da aprendizagem. A regressão e a classificação são os dois tipos de tarefas mais comuns do aprendizado supervisionado (HAN *et al.*, 2011).

No aprendizado não-supervisionado o algoritmo aprende a partir de dados não rotulados, executando a extração de recursos, e avaliando os padrões específicos do próprio conjunto de dados para definir as comparações. O algoritmo descobre o que está sendo apresentado e vai explorando todos os subconjuntos para identificar as similaridades. O Agrupamento (*Clustering*) e a Associação (*Association rule*) são tarefas comumente

aplicadas na aprendizagem não-supervisionada (XIN *et al.*, 2018).

O aprendizado semissupervisionado é aplicado para conjuntos de dados rotulados e não rotulados. Nesse aprendizado, os exemplos rotulados são usados para aprender modelos de classe e exemplos não rotulados são usados para refinar os limites entre as classes. No aprendizado por reforço o algoritmo aprende por meio de tentativa e erro que geram recompensas ou punições como *feedback* das próprias ações. Esse tipo de aprendizado é fortemente difundido na área de jogos (SHAO *et al.*, 2019).

Os problemas clássicos de Aprendizado de Máquina são altamente relacionados à mineração de dados, que é uma etapa do processo de descoberta de conhecimento em bancos de dados (*Knowledge Discovery in Database - KDD*). No processo de mineração de dados, executa-se a extração automatizada de padrões que representam um determinado conjunto de dados, associado a um contexto em análise. Diversos tipos de algoritmos podem ser aplicados para desenvolver a mineração de dados, de acordo com uma abordagem específica para o problema em análise (HAN *et al.*, 2011).

Alguns dos algoritmos mais utilizados na aprendizagem supervisionada são Regressão Linear, Regressão Logística, Árvore de Decisão, Floresta Aleatória, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)* e *Naïve Bayes*. Para a aprendizagem não-supervisionada é bastante usual a aplicação do *K-Means*, *Fuzzy K-Means*, Agrupamento Hierárquico e Mistura Gaussiana. Além desses citados, existem as redes neurais artificiais, que são algoritmos aplicados tanto para o aprendizado supervisionado e não-supervisionado, como para aprendizado por reforço, dependendo dos objetivos de aprendizagem (AGGARWAL, 2015).

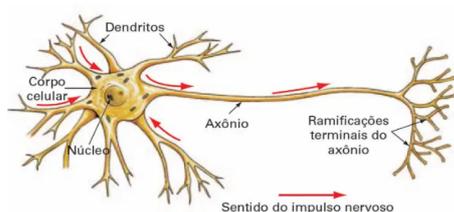
As redes neurais artificiais foram desenvolvidas a partir da motivação de simular os sistemas biológicos correspondentes ao cérebro humano para criar estruturas de Aprendizado de Máquina. No cérebro humano, a rede neural é constituída por neurônios, terminais axônios, regiões sinápticas e outras estruturas celulares. O neurônio é uma célula neural responsável pela condução de impulsos nervosos, e o axônio é um prolongamento dessa célula que transmite o impulso até os dendritos (ramificações) de outro neurônio. As regiões sinápticas são áreas de proximidade entre a extremidade de um neurônio e uma célula vizinha (LODISH *et al.*, 2000).

Na rede neural artificial, uma unidade de computação básica também é denominada de neurônio. A arquitetura mais básica da rede neural é um *Perceptron*, que é uma unidade

formada por um conjunto de nós de entrada, pesos distribuídos para cada uma dessas entradas, uma função somatória, uma função de ativação e um nó de saída (PATTERSON; GIBSON, 2017). O desenvolvimento do *Perceptron* ocorreu em 1958 por Frank Rosenblatt utilizando como base a modelagem matemática desenvolvida por Warren McCulloch e Walter Pitts em 1943 (GOLDEN, 2001).

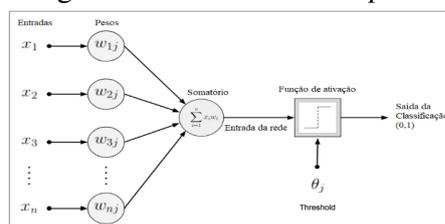
Nas Figuras 5 e 6 é possível visualizar as representações de um neurônio biológico e do neurônio artificial *Perceptron*.

Figura 5 – Neurônio biológico



Fonte: DSA (2021)

Figura 6 – Neurônio *Perceptron*



Fonte: Adaptada de Patterson e Gibson (2017)

O algoritmo de aprendizagem *Perceptron* altera os pesos até que todas as entradas sejam classificadas corretamente. Este procedimento funciona corretamente para problemas linearmente separáveis, como o problema da porta lógica OR e AND. No entanto, quando o problema não é linearmente separável, tal como o problema da porta lógica XOR, o *Perceptron* simples não consegue finalizar a classificação. Um problema linearmente separável se refere a um conjunto de dados para o qual se pode encontrar os valores de um hiperplano que será dividido corretamente por uma linha, formando duas classes de dados (PATTERSON; GIBSON, 2017).

A limitação do *Perceptron* com os problemas não linearmente separáveis foi vista na época como uma falha para o campo das redes neurais, e só foi superada anos depois quando foi desenvolvido o *Multilayer Perceptron* (MLP), que é uma rede com mais de uma camada de *Perceptrons*. O MLP pertence à categoria das redes neurais *feedforward*, que são redes neurais em que as conexões entre as unidades não formam um ciclo, e a informação só é transmitida para frente na rede, sem *loops*, passando pelos nós de entrada, depois pelos nós ocultos (se houver) e, finalmente, pelos nós de saída (AGGARWAL, 2018).

O treinamento de uma rede MLP depende do ajuste de vários parâmetros que são essenciais para refinar a aprendizagem no conjunto de dados. Dentre os principais

parâmetros da rede destaca-se a taxa de aprendizagem, a função de perda, a função de ativação, o número de neurônios e o número de camadas. A taxa de aprendizagem dimensiona o tamanho das atualizações em cada iteração ajustando a rede para o mínimo de uma função de perda. A função de perda quantifica o erro de aprendizagem da rede, indicando a proximidade para os valores ideais. Em associação com a taxa de aprendizado é também possível definir o *momentum*, um parâmetro que aumenta a velocidade de treinamento da rede, reduzindo o perigo de instabilidade, caso o algoritmo fique preso em determinados locais no espaço de busca (PATTERSON; GIBSON, 2017).

A função de ativação é o parâmetro da rede que controla o acréscimo progressivo dos valores de saída ao longo das camadas da rede, de acordo os valores máximos e mínimos de um determinado intervalo. As principais funções de ativação utilizadas nas redes neurais são a Sigmoide, a Tangente Hiperbólica (TanH), a Unidade Linear Retificada (ReLU), a Unidade Linear Exponencial (ELU) e a Unidade Linear Retificada com Vazamento (AGGARWAL, 2018).

Com relação ao número de camadas e ao número de neurônios em cada camada da rede, esses parâmetros precisam ser ajustados de acordo com a complexidade do problema. O limite de decisão do conjunto de dados influencia o ajuste da quantidade de neurônios e de camadas para se obter uma aprendizagem eficiente. A partir dessa percepção que surgiu o paradigma da aprendizagem profunda, que corresponde ao processo de Aprendizado de Máquina para problemas de complexidade elevada utilizando redes neurais profundas.

2.4.1 Aprendizado Profundo

O Aprendizado Profundo é uma categoria do aprendizado de máquina, no qual as redes neurais de várias camadas de neurônios artificiais aprendem representações dos dados em diferentes níveis de abstração. As redes de várias camadas, também denominadas de redes profundas, são constituídas de camadas ocultas adicionais que permitem a otimização e o refinamento da aprendizagem. As multicamadas de neurônios exploram os dados, extraíndo recursos e transmitindo-os para os níveis superiores até que o processamento seja finalizado (LECUN *et al.*, 2015).

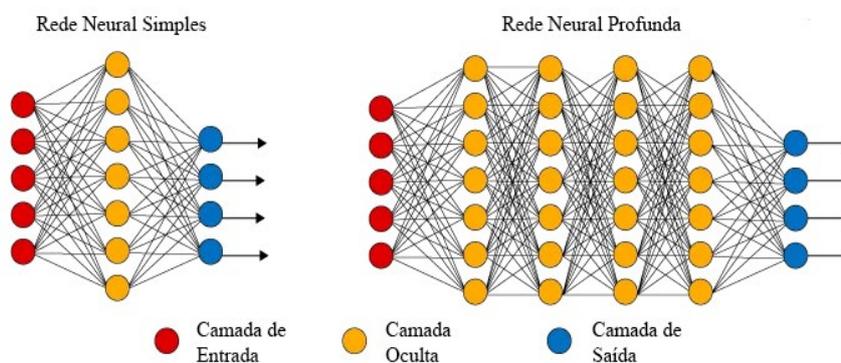
O termo *Deep Learning* (DL) foi citado a primeira vez em 1986 por Rina Dechter em seu estudo *Learning While Searching in Constraint-Satisfaction Problems*, e a partir

disso tornou-se um importante tema de discussão na área da Inteligência Artificial, tendo sua maior expansão após a interseção das redes neurais profundas com as *Graphics Processing Unit* (GPU) para acelerar o processamento (DECHTER, 1986; MADHAVAN; JONES, 2021).

O DL permite o processamento de imagens, vídeos, linguagem natural, sendo integrado em diversas aplicações como assistentes virtuais, recomendação de entretenimento, robótica, detecção de fraude, carros autônomos, sistemas de reconhecimento facial, modelagem de investimento, modelagem de diagnóstico médico, dentre outras (YOUNG *et al.*, 2018).

A estrutura geral de uma rede neural profunda é definida pela existência de uma camada de entrada, pelas camadas ocultas, que variam de acordo com a arquitetura da rede, e uma camada de saída, na qual cada nó correspondente a uma das saídas válidas da rede. Na Figura 7 está representada a variação estrutural de uma rede neural simples para uma rede neural profunda, sendo evidenciada a diferença no número de camadas ocultas.

Figura 7 – Representação da Rede Neural



Fonte: Adaptada de DSA (2021)

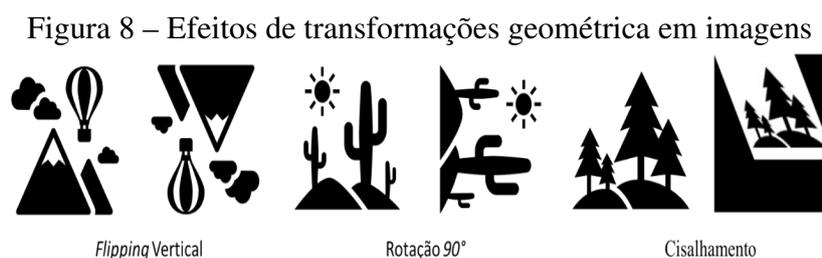
Nas redes neurais profundas, todo o aprendizado dos pesos é feito automaticamente pelo algoritmo de retropropagação, que usa programação dinâmica para calcular o gradiente de uma função de composição, com base na regra da cadeia de cálculo diferencial. Esse algoritmo possui duas fases principais, conhecidas como fases *forward* (avanço) e *backward* (retrocesso). Na fase *forward* ocorre o cálculo dos valores de saída e os derivados locais em vários nós, e na fase *backward* é determinado o acúmulo dos produtos desses valores locais em todos os pontos do nó para a saída (AGGARWAL, 2018).

As redes de retropropagação necessitam de milhares de iterações para aprender as representações dos dados de forma precisa. A dimensão do conjunto de dados influencia

diretamente na capacidade da rede construir as estruturas de aprendizado. Quando o conjunto de dados é pequeno, com pouca variabilidade representativa, ou quando a rede não está estruturalmente bem definida para o problema, podem ocorrer dois tipos de problema com a aprendizagem: O *underfitting* (aprendizagem insuficiente) e o *overfitting* (sobreajuste da aprendizagem), ambos prejudiciais para os objetivos do Aprendizado de Máquina (ZHANG *et al.*, 2019).

No problema do *underfitting*, o modelo não consegue capturar as características dos conjuntos de treinamento e, conseqüentemente, não realiza adequadamente o mapeamento no conjunto de teste. No *overfitting*, o modelo aprende excessivamente a partir dos dados do conjunto de treinamento e não consegue criar uma generalização para as amostras de teste ou validação. Também pode ser consequência de uma má distribuição representativa entre os subconjuntos de treinamento, teste e validação (AGGARWAL, 2018).

Quando é necessário realizar o processamento de imagens nas redes neurais profundas, uma das principais estratégias para evitar o *overfitting* é adicionar mais amostras ao conjunto de dados, criando-se um aumento de dados com uso de transformações geométricas (inversão - "*flipping*", rotação, translação, corte, escala, cisalhamento), transformações de espaço de cores, adição de ruídos, filtros de núcleo, combinação de imagens e apagamento de áreas aleatórias (CLARO *et al.*, 2020). Na Figura 8 é possível visualizar o efeito do *flipping* vertical, da rotação 90° e do cisalhamento nas imagens.



Fonte: Elaborada pela autora (2022)

Também é possível gerar aumento de dados utilizando recursos das próprias redes neurais profundas. O Aumento do Espaço de Recursos (*Feature Space Augmentation*) pode ser realizado a partir das Redes Neurais Convolucionais (*Convolutional Neural Network* - CNN), dos *Autoencoders* e das Redes Adversariais Generativas (*Generative Adversarial Networks* - GANs). Esses tipos de redes fazem parte das principais arquiteturas de Aprendizado Profundo, incluindo as Redes Neurais Recorrentes, as Redes Recursivas e

as Redes *Long Short Term Memory* (LSTM) (PATTERSON; GIBSON, 2017).

Na última década, o processamento de imagens médicas através das Redes Neurais Profundas, mais especificamente das Redes Convolucionais, tem apresentado resultados muito relevantes para o aprimoramento do diagnóstico de câncer, proporcionando uma redução nas taxas de falso-positivos e falso-negativos durante o processo clínico (AGGARWAL, 2018; ABDELHAFIZ *et al.*, 2019).

Uma CNN é um tipo de arquitetura multicamada que foi inspirada pelo córtex visual animal e é particularmente muito utilizada para processamento de imagem. A extração de recursos realizada pela rede permite a segmentação e detecção de objetos em diversos tipos de imagens, como fotografias de faces, paisagens naturais e imagens médicas.

2.4.1.1 Rede Neural Convolutiva

A arquitetura CNN foi proposta por Lecun *et al.* (1999) e é definida por uma combinação de três ideias arquitetônicas, que garantem algum grau de mudança, escala e invariância de distorção em campos receptivos locais, pesos compartilhados e subamostragem espacial. O plano de entrada da rede recebe imagens de objetos cujo tamanho deve ser normalizado.

O termo convolutiva é decorrente de uma operação matemática chamada convolução que é realizada durante o treinamento da rede. Nesta operação um filtro é usado para mapear as ativações de uma camada para a próxima. O produto escalar entre todos os pesos no filtro de qualquer região espacial (do mesmo tamanho do filtro) em uma camada determina o valor do estado oculto na próxima camada. Para uma entrada 2D, ou seja, uma imagem comum, o valor de cada pixel é recalculado, com base na média ponderada dos *pixels* em torno dele (AGGARWAL, 2018).

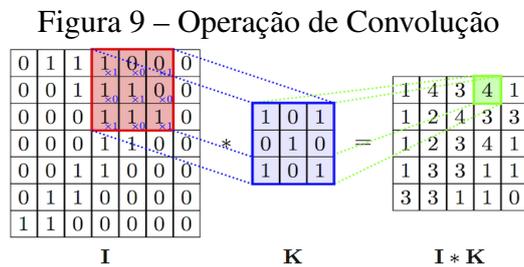
A operação de convolução é normalmente indicada por um asterisco. Na função (1), $x(t)$ é a entrada e w corresponde aos pesos do *kernel* (GOODFELLOW *et al.*, 2016).

$$s(t) = (x * w)(t) \quad (1)$$

As convoluções aplicadas em mais de um eixo de cada vez são representadas pela equação (2), na qual I corresponde a uma imagem bidimensional de entrada, K é um *kernel* bidimensional, m e n indica uma faixa de valores válidos (GOODFELLOW *et al.*, 2016).

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (2)$$

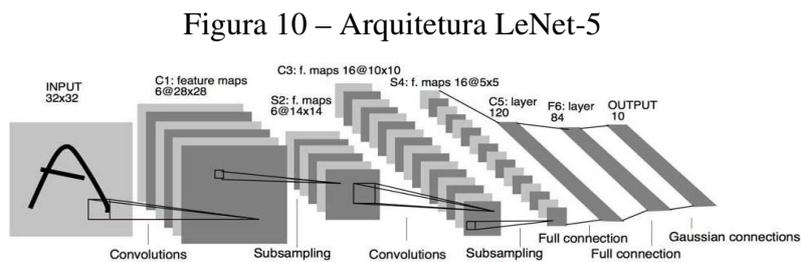
Na Figura 9 se destaca uma representação da etapa de convolução em uma determinada imagem em processamento na rede.



Fonte: TDS (2021a)

Os atributos de uma imagem de entrada correspondem a uma estrutura com largura (*pixels*), altura (*pixels*) e profundidade (Canais RGB). Dessa forma, a arquitetura de uma CNN corresponde a um volume tridimensional de neurônios, tornando-se um aspecto significativo de como as CNNs evoluíram das redes *feedforward* anteriores (PATTERSON; GIBSON, 2017).

As especificações estruturais de uma CNN variam de acordo com o modelo de implementação. Na Figura 10 destaca-se a arquitetura da CNN LeNet-5 desenvolvida por Lecun *et al.* (1999), na qual representa-se a imagem de entrada, os mapas de recursos entre as camadas de convolução, as subamostragens espaciais, as camadas totalmente conectadas e a camada de saída.

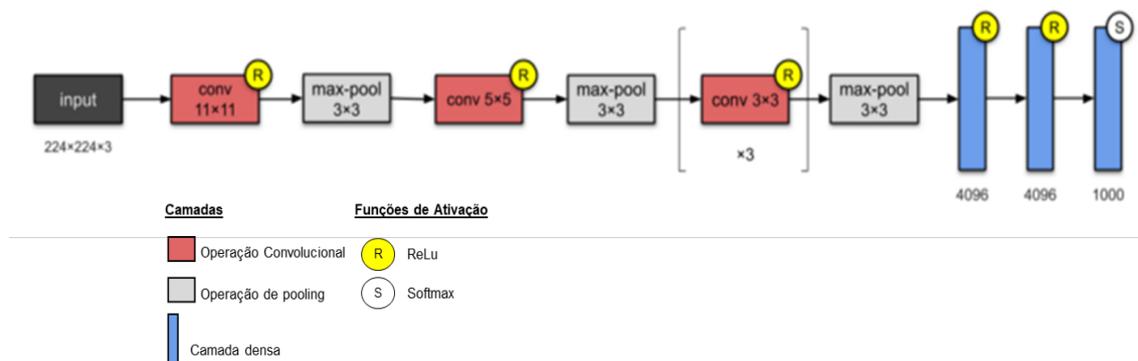


Fonte: Lecun *et al.* (1999)

Outras arquiteturas de CNN como a AlexNet (KRIZHEVSKY *et al.*, 2012a), a VGG-16 (SIMONYAN; ZISSERMAN, 2014), a Inception v1 (SZEGEDY *et al.*, 2015), a Residual Neural Network (ResNet-50) (HE *et al.*, 2015), entre outras, seguem diferentes implementações para as camadas de entrada, camadas convolucionais, filtros, mapas de ativação, camadas de *pooling* e camadas totalmente conectadas. As características de cada arquitetura criam um padrão ordenado de transformações nas imagens, construindo progressivamente recursos de ordem superior (PATTERSON; GIBSON, 2017).

Na Figura 11 é possível visualizar as características específicas na distribuição de camadas da arquitetura AlexNet, constituída por cinco camadas convolucionais, camadas de *pooling* e três camadas densas.

Figura 11 – Arquitetura AlexNet



Fonte: Adaptada de TDS (2021)

Nas camadas convolucionais, os filtros (*Kernel*) são aplicados em toda a largura e altura do campo receptivo da imagem de entrada em uma janela deslizante, controlando os mapas de ativação que serão transferidos para as próximas camadas. Nessas camadas também ocorre o compartilhamento dos hiperparâmetros para minimizar o total de recursos necessário para aprender o conjunto de dados de treinamento (PATTERSON; GIBSON, 2017).

Principais hiperparâmetros de uma CNN:

- Tamanho do filtro – Determina a largura e altura do filtro, geralmente definido como 5x5 ou 3x3. Para as imagens em cores RGB é necessário indicar a profundidade do filtro, destacando que existem 3 canais de cores;
- Profundidade de saída - Controla a contagem de neurônios na camada convolucional que está conectada à mesma região do volume de entrada;
- *Stride* – Determina o tamanho do passo realizado pela janela deslizante do filtro;
- *Zero-padding* – É um preenchimento de zero utilizado para controlar o tamanho espacial dos volumes de saída.

As camadas de *pooling* são inseridas entre camadas convolucionais para reduzir progressivamente o tamanho espacial (largura e altura) da representação de dados. Essa redução é realizada por meio de filtros em toda dimensão espacial dos dados, simplificando as informações de saída da camada de acordo com uma estatística resumida dos valores

agrupados. O *pooling* médio (avg-pool) e o *pooling* máximo (max-pool) são duas das principais técnicas utilizadas para o procedimento de *pooling* em uma rede convolucional. No *pooling* médio, os *pixels* são agrupados, e a média desses valores é adotada como saída. Enquanto no *pooling* máximo, é adotada como saída o valor máximo do agrupamento de *pixels* (AGGARWAL, 2018).

As últimas camadas de uma CNN são as camadas totalmente conectadas. Elas são responsáveis pelos cálculos finais que determinam as saídas da rede, e funcionam exatamente como uma rede densa *feedforward* tradicional, na qual cada neurônio está conectado com todos os neurônios da próxima camada. As dimensões do volume de saída nessas camadas são $1 \times 1 \times N$, ou seja, um vetor com N número de classes. As funções de ativação utilizadas nessas camadas dependem da natureza da aplicação, e podem ser função logística, tangencial, *softmax* ou linear (PATTERSON; GIBSON, 2017).

O desenvolvimento das CNNs proporcionou inúmeros avanços no processo de extração de recursos de imagem e na visão computacional. Devido à propriedade de "invariância espacial", as Redes Convolucionais possuem a capacidade de reconhecer as características da imagem, independentemente da posição, permitindo a análise de imagens para diversos objetivos. Além disso, as atualizações nas arquiteturas e a ampliação do poder computacional de hardware tornam o processamento de dados realizado pelas CNNs mais rápido e eficiente.

No processo de classificação de imagens, especialmente de imagens médicas como mamografias, o desempenho das CNN's proporcionou avanços significativos na identificação de padrões correspondentes às lesões malignas na mama. E a integração dessas redes em sistemas médicos fornece recursos que auxiliam aos radiologistas aperfeiçoar análises quantitativas precisas no diagnóstico do câncer.

2.5 Considerações Finais

Neste capítulo foram apresentados os principais conceitos acerca do câncer de mama, do exame de mamografia e da classificação BI-RADS que padroniza os diagnósticos do câncer de mama. Explanou-se os principais conceitos referentes ao Aprendizado de Máquina e ao Aprendizado Profundo, além das definições conceituais sobre as Redes Neurais Convolucionais e as aplicações no processamento de imagens.

No próximo capítulo serão discutidos os trabalhos relacionados, obtidos por meio

de uma Revisão Sistemática da Literatura, que serviram como referencial técnico para os desenvolvimentos dos experimentos realizados neste trabalho.

3 TRABALHOS RELACIONADOS

Os direcionamentos técnicos desta pesquisa foram definidos a partir da revisão de outros trabalhos relacionados com o objetivo de aplicar Aprendizado Profundo por meio de Redes Neurais Convolucionais para o desenvolvimento de um modelo de classificação de mamografias. Os trabalhos mais relevantes, que serão citados neste capítulo foram identificados por meio de uma Revisão Sistemática da Literatura (RSL) iniciada com a seleção de 176 publicações do ano de 2015 a 2021.

Foram consultadas quatro bases de publicações científicas: *Science Direct*, *IEEE Xplore*, *ACM Digital Library* e Periódicos CAPES. As três primeiras Bases citadas foram selecionadas para as buscas de trabalhos porque são os principais repositórios de publicações internacionais, e apresentam as pesquisas mais recentes da área da Computação. Em relação ao Periódicos Capes, está é a principal fonte de trabalhos nacionais em Língua Portuguesa, que indexa os trabalhos desenvolvidos nos programas de Pós-Graduação do Brasil.

A revisão bibliográfica dos trabalhos retornados em cada base foi refinada a partir da aplicação dos seguintes Critérios de Exclusão:

- CE1 - Não aborda manipulação de mamografias.
- CE2 - Não tem foco na classificação de massas ou calcificações (aborda sobre densidade e texturas de tecidos).
- CE3 - Trabalhos que são um resumo ou Revisão Sistemática de Literatura.
- CE4 - Trabalhos escritos em outro idioma além do Inglês ou Português.
- CE5 - Trabalhos que não abordam aplicações com rede neural para classificação de imagens ou os resultados do trabalho são pouco descritivos.
- CE6 - Trabalhos iguais ou que possuam versão atualizada (deve ser mantido apenas o mais recente).
- CE7 - Trabalhos que não estão disponíveis para visualização ou download no momento da pesquisa.

Na Tabela 3 estão descritas as *strings* de busca em cada Base, com a indicação do total de trabalhos que foram retornados na consulta. A primeira consulta de publicações foi realizada em novembro de 2020 e, posteriormente, a revisão foi atualizada em janeiro de 2022 para inclusão de novos trabalhos. Na Tabela 4 descreve-se o refinamento da RSL a partir da aplicação dos Critérios de Exclusão citados anteriormente. Por fim, foram

incluídos para a avaliação completa 36 trabalhos, 30 publicações coletadas na Base *Science Direct*, três na *IEEE Xplore*, um na *ACM DL* e dois no repositório da *CAPES*.

Tabela 3 – Strings de busca para a RSL

Base	String	Total
Science Direct	TITLE-ABSTR-KEY ("Breast cancer"OR "câncer de mama") AND ("Mamografia" OR "Mammography") AND ("Rede neural" OR "neural network") AND ("Convolutacional" OR "Convolutional") AND ("BI-RADS")	148
IEEE Xplore	("Breast cancer"OR "câncer de mama") AND ("Mamografia" OR "Mammography") AND ("Rede neural" OR "neural network") AND ("Convolutacional" OR "Convolutional") AND ("BI-RADS")	6
ACM DL	("Breast cancer"OR "câncer de mama") AND ("Mamografia" OR "Mammography") AND ("Rede neural" OR "neural network") AND ("Convolutacional" OR "Convolutional") AND ("BI-RADS")	6
Periódicos CAPES	Opção marcada: Título + Contém: ("Breast cancer"OR "câncer de mama") AND Opção marcada: Qualquer + Contém: ("Mamografia" OR "Mammography") AND ("Rede neural" OR "neural network") AND ("Convolutacional" OR "Convolutional") AND ("BI-RADS")	16

Fonte: Elaborada pela autora (2022)

Tabela 4 – Aplicação dos Critérios de Exclusão da RSL

Base	Total	CE1	CE2	CE3	CE4	CE5	CE6	CE7	Incluídos
Science Direct	148	36	14	42	0	21	1	4	30
IEEE Xplore	6	0	2	0	0	1	0	0	3
ACM DL	6	2	1	2	0	0	0	0	1
Periódicos CAPES	16	1	2	4	0	0	1	6	2

Fonte: Elaborada pela autora (2022)

Com base na revisão completa dos 36 trabalhos incluídos na RSL, foi possível identificar as publicações com os melhores resultados decorrente da aplicação do Aprendizado Profundo, especificamente do treinamento das Redes Neurais Convolucionais para classificar lesões de massa ou calcificações em mamografias. Também foi possível conhecer as técnicas de pré-processamento de dados adotadas por outros autores e os principais métodos de avaliação de desempenho dos modelos de classificação.

Dos 36 trabalhos incluídos na revisão, dez publicações apresentam a classificação das mamografias em abordagem de multiclasse. Enquanto os outros 26 trabalhos descrevem uma predição binária das imagens mamográficas, sendo a maioria direcionada para a distinção das lesões em benigna ou maligna.

Dentre as dez publicações que abordam a classificação multiclasse das mamografias, cinco trabalhos (CARNEIRO *et al.*, 2017, KAUR *et al.*, 2019, SUJATHA *et al.*, 2020, BAKALO *et al.*, 2021 e WANG *et al.*, 2021) realizaram a predição das imagens em três categorias: Benigno, Maligno e Normal. Essas categorias ressaltam a patologia das lesões

aparentes nas imagens (Benigno ou Maligno), ou a ausência de anormalidade na região mamária (Normal).

No trabalho de Shen *et al.* (2019) o objetivo de análise das mamografias foi definido a partir do procedimento de classificação de regiões de interesse (ROI) conforme cinco categorias: Fundo (*Background*), Massa Benigna, Massa maligna, Calcificação Benigna e Calcificação Maligna. A predição dos ROI's para as classes citadas foi realizada a partir dos classificadores baseados em Resnet50 e VGG16, usando como conjunto de teste as imagens do *Curated Breast Imaging Subset of DDSM (CBIS-DDSM)*. Um método de *bootstrapping* com 3000 execuções foi utilizado para derivar os intervalos de confiança de 95% para a precisão, que foi registrada em 0,89 [0,88, 0,90] no modelo Resnet50 pré-treinado e 0,84 [0,83, 0,85] no modelo VGG16, também com pré-treinamento.

Na avaliação de desempenho dos modelos para a predição das classes, Shen *et al.* (2019) relatam que a classe de fundo foi a mais fácil de classificar, enquanto as calcificações malignas foram as mais difíceis de se definir predições corretas, sendo mais propensas a serem classificadas erroneamente como calcificação benigna, seguida de massa maligna.

Seguindo o direcionamento de predição multicategórica de mamografias, outras quatro publicações foram identificadas na RSL, e correspondem a trabalhos que desenvolveram a classificação das imagens conforme o léxico do protocolo BI-RADS. No trabalho de Hernández-López e Gómez-Flores (2020) foi realizada a classificação de massas conforme o léxico BI-RADS para Formato, Margem e Densidade. Essas características histopatológicas estão associadas diretamente ao desenvolvimento de uma lesão benigna ou maligna. Os tumores benignos são comumente de morfologia regular com margens bem definidas e circunscritas, enquanto os tumores malignos tendem a ter formas irregulares e margens mal definidas.

A classificação das massas foi realizada a partir de dois modelos híbridos, o Perceptron Multicamada (MLP) e Neurônio Morfológico Dendrítico (DMN) para classificação de padrões, e a arquitetura AlexNet pré-treinada foi utilizada para extração de características. No procedimento de avaliação dos modelos, Hernández-López e Gómez-Flores (2020) aplicaram o método de validação cruzada, considerando um $k\text{-fold} = 10$, no qual em cada experimento, um grupo é o conjunto de teste e os nove grupos restantes compõem o conjunto de treinamento. Os resultados de classificação relatam que os modelos neurais híbridos obtiveram uma precisão de 91% para predição dos atributos de forma e densidade

e uma precisão de 85% para predição de atributos de margem. Os autores também executaram o teste estatístico de Wilcoxon para afirmar que não foram observadas diferenças significativas entre os modelos híbridos.

No trabalho de Chougrad *et al.* (2020), os resultados da análise das mamografias foram obtidos a partir da classificação multiclasse em 8 rótulos (Densidade: BI-RADS I - IV, Massa, Calcificação, Benigno e Maligno), sendo quatro deles definidos pelo léxico BI-RADS para a classificação da densidade mamária. A densidade mamária é considerada um marcador independente de risco do câncer de mama, pois a composição do tecido mamário pode afetar a sensibilidade da mamografia em virtude do “efeito de mascaramento”, ocasionando um aumento do risco de câncer de intervalo (WINKEL *et al.*, 2015).

As imagens utilizadas nos experimentos do trabalho foram coletadas de quatro conjuntos de dados públicos, o CBIS-DDSM, BCDR, INbreast e o MIAS. O pré-processamento realizado no conjunto de imagens foi constituído por um processo de recorte das regiões de interesse, normalização de contraste global e aumento de dados executado através de uma série de transformações aleatórias. Para os experimentos, os autores adotaram o particionamento do conjunto de dados em 70% para treinamento, 10% para validação e 20% para teste.

Chougrad *et al.* (2020) utilizaram a arquitetura CNN VGG16 com pré-treinamento em *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC), adaptada na camada final totalmente conectada para produzir uma saída 8-dimensional indicando a probabilidade prevista sobre os rótulos de classe. Quatro variantes do modelo VGG (VGG-Sc, VGG-2FT, VGG-AIIFT, VGG-FTED) foram desenvolvidas, e o melhor desempenho foi alcançado pelo modelo VGG-FTED, que foi treinado usando uma estratégia de ajuste fino com SGD (com *momentum*) e taxa de aprendizagem exponencialmente decrescente para personalizar todos os pesos pré-treinados, otimizando o modelo para o conjunto de dados de mamografias. No conjunto de dados *Digital Database for Screening Mammography* (DDSM), observou-se uma pontuação F1 de 0.935 ± 0.019 e precisão média de 0.895 ± 0.017 .

Segundo Chougrad *et al.* (2020), o ajuste fino com a taxa de aprendizagem com declínio exponencial permitiu uma aprendizagem mais adaptativa e natural, controlando melhor a mudança dos pesos em cada parte da rede. No entanto, os autores ressaltam que a análise das imagens de mamas altamente densas é mais desafiadora em comparação com

os outros tipos de densidade que têm maior probabilidade de serem anotadas corretamente.

Considerando a análise de mamografias conforme o léxico BI-RADS para a avaliação (*assessment*) das lesões na mama, identificou-se na RSL dois trabalhos que realizaram a classificação das imagens em seis categorias (1-6) da avaliação BI-RADS, Domingues, Abreu e Santos (2018) e Siddeeq *et al.* (2021).

No trabalho de Domingues, Abreu e Santos (2018) foi desenvolvida a classificação das imagens mamográficas do Banco de Dados INbreast em seis categorias BI-RADS (1-6) por meio de um classificador profundo baseado em AlexNet com pré-treinamento. Nos experimentos do trabalho, os autores executaram a divisão do conjunto de dados a partir da seleção aleatória de 75% dos pacientes por classe para treinamento e os 25% restantes foram usados para teste.

Foram propostas duas estratégias de pré-processamento, aumento de dados e pirâmide de escalas. No aumento de dados foi utilizado transformações geométricas de translação, rotação, escala e cisalhamento entre todas as imagens de treinamento para gerar variabilidade. E para capturar e destacar as informações em diferentes escalas foi adotada como entrada para a CNN uma matriz de três canais composta por imagens de Diferença de Gaussianas (DoG).

O desempenho dos modelos foi avaliado usando o Erro Médio Absoluto (*Mean Absolute Error* - MAE), além da análise da acurácia máxima. Os melhores resultados foram alcançados a partir do aprimoramento multiescala baseado em Diferença de Gaussianas, combinado com a técnica de aumento de dados, que ampliou a precisão em mais de 33% quando comparado com os resultados sem o aumento. Quando a imagem DoG foi fornecida como entrada e não foi aplicado balanceamento de dados, observou-se uma acurácia máxima de $83,4 \pm 3,6$ e MAE de $0,4 \pm 0,0$.

Domingues, Abreu e Santos (2018) ressaltam que numa avaliação mamográfica, quando mais de um achado estiver presente na mamografia, o BI-RADS geral no laudo médico corresponde ao achado com o BI-RADS mais alto. Portanto, em trabalhos futuros os autores considerarão a incorporação das características de ordinalidade e valor máximo para a classificação das imagens.

Seguindo o mesmo objetivo de classificação em seis níveis da avaliação BI-RADS, o trabalho de Siddeeq *et al.* (2021) propôs a classificação multiclasse de mamografias por meio de uma rede neural personalizada baseada em ResNet (RN-BCNN), combinando as

técnicas de aumento de dados por deformação elástica e pirâmide de escalas no conjunto de dados INbreast desbalanceado.

Os autores particionaram o conjunto de dados em três subconjuntos para treinamento, teste e validação. O conjunto de treinamento foi definido por 76% dos dados coletados aleatoriamente de cada classe e o conjunto de teste foi formado pelos 24% restantes de cada classe. O conjunto de validação corresponde a 10% do conjunto direcionado para o treinamento. O aumento de dados foi utilizado para gerar artificialmente um novo conjunto de dados de treinamento, aplicando a técnica de deformação elástica, em combinação com transformações de rotação aleatória e zoom.

Siddeeq *et al.* (2021) utilizaram na compilação do modelo RN-BCNN o otimizador *Stochastic Gradient Descent* com a taxa de aprendizado inicial de $1e-2$, em declínio no final de cada época, e a *Categorical Crossentropy* como função de perda. Para a avaliação de desempenho foram observadas a precisão e a pontuação F1, além da matriz de confusão para entender as classes que foram classificadas erroneamente.

O melhor resultado da metodologia proposta apresenta precisão de 85,9% no conjunto de teste, caracterizando um melhor desempenho em comparação com a precisão de 83,4% publicada em um estudo anterior produzido pelos autores. Para trabalhos futuros, Siddeeq *et al.* (2021) pretendem expandir a avaliação das mamografias incluindo os subníveis da categoria 4 do BI-RADS (4A, 4B E 4C), em combinação com a análise *multi-views* das imagens.

Em relação aos procedimentos de pré-processamento executados em imagens de mamografias para o desenvolvimento de modelos de Aprendizado Profundo, com a revisão da literatura foi possível identificar no trabalho de Cogan *et al.* (2019) a execução de um mapeamento de cores para o espaço RGB, com Equalização do Histograma Adaptativo com Limitação de Contraste (*Contrast Limited Adaptive Histogram Equalization - CLAHE*) e redimensionamento das imagens para uma dimensão máxima de 1.024 *pixels*. Estes procedimentos foram realizados nas imagens dos bancos DDSM e INbreast para o treinamento e avaliação de um modelo ResNet-101 Faster R-CNN, desenvolvido com o objetivo de localizar os tumores nas imagens mamográficas.

Ainda na fase de pré-processamento das imagens, Cogan *et al.* (2019) realizaram o procedimento de aumento de dados nas imagens de treinamento com o objetivo de ampliar a variabilidade do conjunto, evitar o *overfitting* da aprendizagem e aumentar a precisão

geral da validação do modelo. Foram aplicadas transformações de rotação e espelhamento para não alterar as classificações das imagens, e para cada mamografia foram geradas sete novas imagens de treinamento usando as seguintes transformações: rotação de 90 graus, rotação de 180 graus, rotação de 270 graus, espelho de imagem, espelho de imagem seguido de rotação de 90 graus, espelho de imagem seguido de rotação de 180 graus e imagem espelho seguido de rotação de 270 graus. Também foram utilizadas inversões horizontais aleatórias, inversões verticais e *jitter* de caixa delimitadora.

Em suas considerações finais, Cogan *et al.* (2019) ressaltam a importância de um grande conjunto de imagens de mamografias para refinar o treinamento da rede neural, assim como Agarwal *et al.* (2020) também evidenciam que o uso de grandes conjuntos de dados para treinar modelos de detecção em imagens médicas pode garantir uma aprendizagem ótima das massas e calcificações.

Diante dos trabalhos citados, este trabalho tem como principal contribuição propor a classificação das imagens de mamografias em duas abordagens: Classificação Multiclasse em cinco níveis BI-RADS (1-5) e a Classificação com *Deep Cross-Training* para a predição dos pares: BIRADS-1/Não BIRADS-1, BIRADS-2/Não BIRADS-2, BIRADS-3/Não BIRADS-3, BIRADS-4/Não BIRADS-4 e BIRADS-5/Não BIRADS-5. Busca-se desenvolver um modelo para a classificação das categorias BI-RADS, capaz de proporcionar uma redução na taxa de falso-positivos das avaliações e auxiliar o médico radiologista na tomada de decisão em relação ao diagnóstico do câncer de mama.

3.1 Considerações Finais

Neste capítulo foram apresentadas as principais referências de trabalhos relacionados com o objetivo desta pesquisa, obtidas e avaliadas por meio de uma Revisão Sistemática da Literatura. Foram ressaltados os direcionamentos mais importantes para o desenvolvimento de modelos de classificação de mamografias com recurso das Redes Neurais Convolucionais, evidenciando-se a aplicação de técnicas de pré-processamento de imagens para otimizar a aprendizagem profunda.

No próximo capítulo será descrita a metodologia de trabalho, com destaque para o pré-processamento das imagens e a descrição da implementação dos modelos de classificação BI-RADS.

4 METODOLOGIA

A metodologia deste trabalho foi baseada no processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Database - KDD*) (FAYYAD *et al.*, 1996), uma metodologia consolidada para a mineração de dados e extração automatizada de padrões que representam um determinado conjunto de dados. O objeto principal de estudo são as imagens de mamografias de dois Bancos de imagens públicas, o *Curated Breast Imaging Subset of DDSM* (CBIS-DDSM) (LEE *et al.*, 2017) e o INbreast (MOREIRA *et al.*, 2012).

4.1 Estudos Iniciais

Os objetivos desse trabalho seguem o direcionamento dos trabalhos iniciais realizados por Souto (2014) e por Silva (2020), que desenvolveram análise de imagens de mamografias através do processo de segmentação para a extração de métricas morfológicas e da aplicação de algoritmos de Aprendizado de Máquina para a classificação de tumores da mama, de acordo com o protocolo BI-RADS.

No trabalho de Souto (2014) foi realizado um processamento de imagens com o objetivo de extrair atributos geométricos em regiões de interesse de tumores mamários para definir uma classificação benigna ou maligna. A análise das métricas de perímetro, área, complexidade de forma, compactação, circularidade e contiguidade permitiu que o software MAMOCAD realizasse a uma interpretação eficaz dos tumores, contribuindo para a diminuição de casos de falso-positivo e falso-negativo no processo de tomada de decisão do especialista.

Silva (2020) propôs uma metodologia (MAMOCAD-BR) estruturada em técnicas de Inteligência Artificial para o processamento e mineração de imagens de mamografias, objetivando a classificação de nódulos na mama de acordo com o protocolo BI-RADS. Os experimentos foram realizados com as imagens da Base *BREAST CANCER DIGITAL REPOSITORY* (BCDR) (LOPEZ *et al.*, 2012), de acordo com as etapas de pré-processamento, segmentação, extração de atributos e classificação dos nódulos nas categorias do BI-RADS (2, 3, 4A, 4B, 4C e 5). A mineração das imagens foi executada com a aplicação dos algoritmos *Classification and Regression Trees* (CART), *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP) e *k-Nearest Neighbors* (KNN).

Com base nos estudos anteriormente citados, este trabalho avança na análise de imagens de mamografias através da aplicação do Aprendizado Profundo e das Redes Neurais Convolucionais para construção de um modelo de classificação de lesões mamárias conforme as categorias (1, 2, 3, 4 e 5) do protocolo BI-RADS.

4.2 Bases de Dados e Pré-Processamento

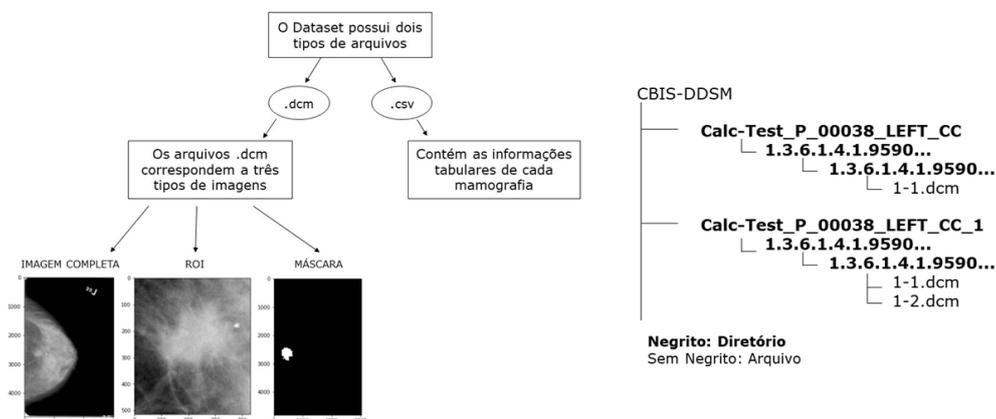
A fase inicial do trabalho de análise foi constituída por um processo de coleta, sistematização e pré-processamento das imagens do CBIS-DDSM e do INbreast. O Banco de imagens CBIS-DDSM, consolidado por Lee *et al.* (2017), é uma versão atualizada da Base *Digital Database for Screening Mammography* (DDSM) (HEATH *et al.*, 1998), sendo composto por 6.671 casos avaliados de massas e calcificações marcadas com BI-RADS 0, 1, 2, 3, 4 e 5. Esta Base é disponibilizada pelo *The Cancer Imaging Archive* (TCIA), um serviço que desidentifica e hospeda imagens médicas de câncer, liberando os arquivos para download público. Os arquivos de imagens são padronizados em formato *Digital Imaging and Communications in Medicine* (DICOM) (Dicom Standard, 2022), com metadados especificados em arquivos *Comma-Separated Values* (CSV), referentes às massas e calcificações na mama.

A Base INbreast tem um total de 115 casos (410 imagens), sendo 90 casos de mulheres com ambas as mamas afetadas (quatro imagens por caso) e 25 casos de pacientes mastectomizadas (duas imagens por caso). Nas imagens disponibilizadas, as lesões registradas são do tipo massas, calcificações, assimetrias e distorções, rotuladas com BI-RADS 1, 2, 3, 4, 5 e 6. A Base INbreast também possui os arquivos de imagens padronizados em formato DICOM, com metadados em CSV e arquivos *eXtensible Markup Language* (XML) para referência de consulta dos casos em imagens de campo completo (FULL) e de Regiões de Interesse (ROI).

No processo de manipulação das imagens do CBIS-DDSM identificou-se que os arquivos de imagens são do tipo “Full” (mamografia completa), “ROI” (corte da região de interesse) e “Mask” (máscara da mamografia completa). Originalmente os arquivos da base não possuem um rótulo único de identificação, pois todas as imagens são identificadas apenas como “1-1” ou “1-2”, organizadas em diretórios e subdiretórios nomeados com o ID do estudo e a marcação do tipo de imagem, conforme é apresentado na Figura 12.

A estrutura original do CBIS-DDSM não favorece a separação das imagens de

Figura 12 – Estrutura do Banco CBIS-DDSM



Fonte: Adaptada de TDS (2021b)

acordo com as categorias BI-RADS, por isso foi necessário executar um procedimento de extração e de nova rotulação dos arquivos para que fossem criados seis subconjuntos (BI-RADS 0, BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4 e BI-RADS 5) de imagens. A sistematização dos arquivos foi executada seguindo as marcações dos metadados, e assim as imagens ficaram corretamente organizadas para as próximas etapas de trabalho.

É importante ressaltar que nos experimentos do trabalho não foi considerado nenhum caso com marcação de BI-RADS 0, pois essa avaliação da mamografia representa um diagnóstico incerto acerca das lesões, conforme definido no manual BI-RADS (citado na Seção 2.3). Em relação aos casos marcados com BI-RADS 1, a Base CBIS-DDSM só possui três imagens de massas marcadas com essa avaliação, porém a patologia descrita nos metadados indica que essas lesões foram diagnosticadas como malignas, e pela definição do protocolo BI-RADS a categoria 1 tem 0% de probabilidade de malignidade, então as imagens foram descartadas dos experimentos. Para incluir casos da categoria BI-RADS 1 e balancear o conjunto de dados, foram coletadas 400 imagens com diagnóstico “Normal” da Base DDSM (HEATH *et al.*, 1998), assumindo que a Base CBIS-DDSM é derivada dessa referida Base de mamografias. As imagens citadas foram obtidas por meio da plataforma *Kaggle*, disponibilizadas no projeto *The Complete Mini-DDSM* (CHEDDAD, 2021).

Foram utilizadas as imagens em campo completo de ambas as Bases de imagens (CBIS-DDSM e INbreast) e os ROI's das imagens pertencentes ao CBIS-DDSM. Todas as imagens foram convertidas do formato DICOM para o formato *Portable Network Graphics* (PNG), mesmo procedimento de conversão realizado por Yala *et al.* (2021) e Shen *et al.* (2019). Além da conversão de formato, as imagens foram também submetidas à Equaliza-

ção do Histograma Adaptativo com Limitação de Contraste (*Contrast Limited Adaptive Histogram Equalization - CLAHE*) para aumentar o contraste local das mamografias, estratégia também executada por Cogan *et al.* (2019).

No processo de estruturação dos *datasets* foram consideradas mamografias sem anormalidades (Normais), e mamografias com incidência de massas e calcificações. Todas as imagens selecionadas para o trabalho foram ordenadas em um conjunto principal sistematizado por cinco classes balanceadas, de acordo com as categorias de 1-5 da avaliação BI-RADS, e posteriormente foi subdividido em 70% para treinamento, 10% para validação e 20% para teste dos modelos, seguindo como referência Agarwal *et al.* (2020).

Na finalização da etapa de pré-processamento das imagens foi realizado um procedimento de aumento de dados no subconjunto de treinamento, principalmente para os modelos de classificação multiclasse. O aumento foi executado com transformações geométricas de deslocamento vertical e horizontal, inversão vertical e horizontal, zoom e rotação. Na Tabela 5 é apresentada a descrição do tamanho dos 11 *datasets* que foram estruturados para os experimentos.

O Dataset_01 corresponde à sistematização das imagens de campo completo em cinco classes BI-RADS (1-5) e recebeu o aumento de dados (A.D) com a seguinte configuração: $Width_shift_range = 0.01$, $height_shift_range = 0.01$, $zoom_range = 0.05$, $horizontal_flip = True$, $vertical_flip = True$, $rotation_range = 5$. Os Datasets_(02-06) foram utilizados para o treinamento/validação e teste dos modelos desenvolvidos com a abordagem *Deep Cross-Training*. Eles foram rotulados como SET-B1_NB1, SET-B2_NB2, SET-B3_NB3, SET-B4_NB4 e SET-B5_NB5, e cada um foi organizado com duas classes. Nesses *datasets*, se uma imagem é positiva para um determinado BI-RADS na primeira classe, o modelo precisa diferenciá-la dos outros BI-RADS que estão mesclados na segunda classe. Ou seja, a classe Não BI-RADS 1 do dataset SET-B1_NB1 recebe as imagens BI-RADS 2, BI-RADS 3, BI-RADS 4, BI-RADS 5, e assim por diante.

Por fim, os outros cinco Datasets_(07-11) possuem a mesma sistematização dos Datasets_(02-06), porém eles são compostos apenas pelos ROI's das imagens pertencentes à Base CBIS-DDSM.

Tabela 5 – Descrição dos *Datasets*

	Num de classes	Número de Imagens		
		Treinamento	Validação	Teste
Dataset_01	5	9600 (A.D)	225	455
Dataset_(02-06)	2	420	90	182
Datasets_(07-11)	2	668	94	192

Fonte: Elaborada pela autora (2022)

4.3 Arquiteturas dos Modelos

As arquiteturas AlexNet (KRIZHEVSKY *et al.*, 2012b), ResNet-18 (HE *et al.*, 2015) e a ResNet-34 (HE *et al.*, 2015) foram utilizadas como base para o desenvolvimento dos modelos CNN responsáveis pela classificação das imagens conforme as categorias BI-RADS.

A arquitetura AlexNet é uma CNN de 5 camadas convolucionais, que foi a vencedora do concurso *ImageNet Large Scale Visual Recognition Challenge 2012 - LSVRC2012*. Essa rede recebe como entrada imagens coloridas de tamanho 224x224x3 ou 227x227x3, que são processadas através de 11 camadas (convolucionais, *max-pool* e totalmente conectadas), com mais de 60 milhões de parâmetros treináveis. A descrição dessa arquitetura é apresentada na Tabela 6.

As redes ResNet-18 e ResNet-34 são arquiteturas de Redes Residuais Profundas (*Deep Residual Network - ResNet*) (HE *et al.*, 2015), que permitem o treinamento de centenas de camadas de neurônios, com destacável desempenho na tarefa de classificação, e nos últimos sete anos possibilitaram o aperfeiçoamento dos estudos na área de Visão Computacional.

O treinamento baseado em mapeamento residual é definido pelo método de conexões de atalho ou conexões de salto que desviam os dados de uma camada para outra, sem prejudicar a capacidade de aprendizado do modelo. Dessa forma, torna-se possível lidar com a degradação de aprendizagem de forma mais otimizada (HE *et al.*, 2015). Ao saltar determinadas camadas durante o treinamento, o modelo retém as camadas úteis para a aprendizagem, favorecendo o cálculo do gradiente, sem adicionar parâmetros extras ou complexidade computacional (CHEN *et al.*, 2018).

As arquiteturas ResNet-18 e ResNet-34 possuem respectivamente 18 e 34 camadas (convolucionais, filtros 3x3 e *poolings*). A ResNet-18 possui mais de 11 milhões de

parâmetros treináveis, enquanto os parâmetros treináveis da ResNet-34 excedem mais de 21 milhões. Ambas as arquiteturas recebem como entrada imagens de dimensão 224x224x3. Na Tabela 6 também está descrita a distribuição de camadas de cada uma das arquiteturas citadas e o tamanho das saídas.

Tabela 6 – Arquiteturas CNN's dos Modelos BI-RADS

AlexNet			ResNet	18-camadas	34-camadas
Filtro	Tamanho		Tamanho		
Tamanho/passo	saída		saída		
Conv1	11x11/4	96x55x55	Conv1	112x112	7x7, 64, passo 2
max-pool1	3x3/2	96x27x27	Conv2_x	56x56	3x3 max pool, passo 2
Conv2	5x5/1	256x27x27			$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$ $\left \begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
max-pool2	3x3/2	256x13x13	Conv3_x	28x28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$ $\left \begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
Conv3	3x3/1	384x13x13	Conv4_x	14x14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$ $\left \begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
Conv4	3x3/1	384x13x13	Conv5_x	7x7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$ $\left \begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
Conv5	3x3/1	256x13x13		1x1	average pool, 5-d fc, softmax
max-pool5	3x3/2	256x6x6	FLOPs		1.8x10 ⁹ 3.6x10 ⁹
Fc-6		4096			
Fc-7		4096			
Fc-8		5			

4.4 Desenvolvimento dos Modelos

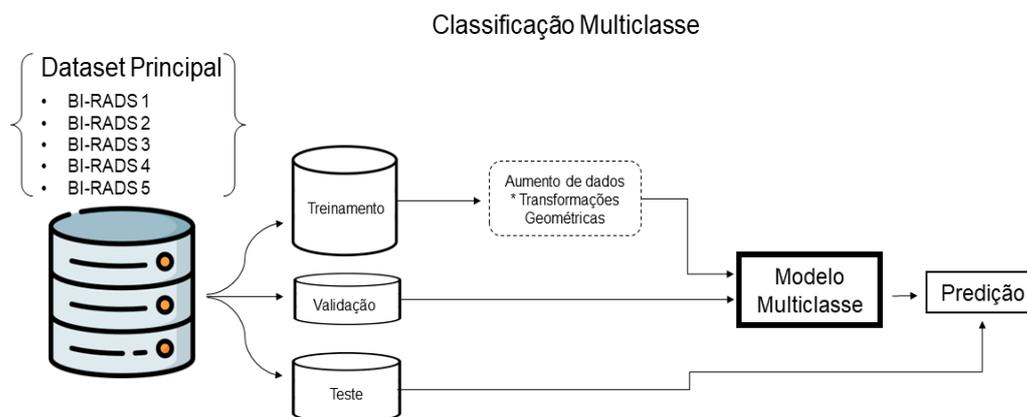
O desenvolvimento e avaliação dos modelos de classificação BI-RADS foram definidos conforme o fluxo de processo descrito na Figura 13 (abordagem Multiclasse) e na Figura 14 (abordagem *Deep Cross-Training* - DCTr), nas quais são ilustradas a fase de treinamento/validação do modelo e a execução da predição das imagens na fase de teste.

A Figura 13 ilustra o desenvolvimento do modelo Multiclasse, no qual o Dataset principal, composto por cinco classes (BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4 e BI-RADS 5) foi particionado em três conjuntos: Treinamento, Validação e Teste, obedecendo à porcentagem de divisão de 70/10/20 citada na Seção 4.2.

No conjunto de treinamento (70%), o aumento de dados com transformações geométricas foi aplicado como suporte para ampliar a variabilidade do conjunto de imagens e impedir o *overfitting* da aprendizagem no desenvolvimento do classificador. E na etapa de avaliação de desempenho, as imagens do conjunto de Teste (20%) foram submetidas à predição das classes BI-RADS e cada modelo foi avaliado conforme as métricas de

acurácia, precisão, *recall* e pontuação F1.

Figura 13 – Multiclasse - desenvolvimento do modelo (bloco pontilhado é opcional)



Fonte: Elaborada pela autora (2022)

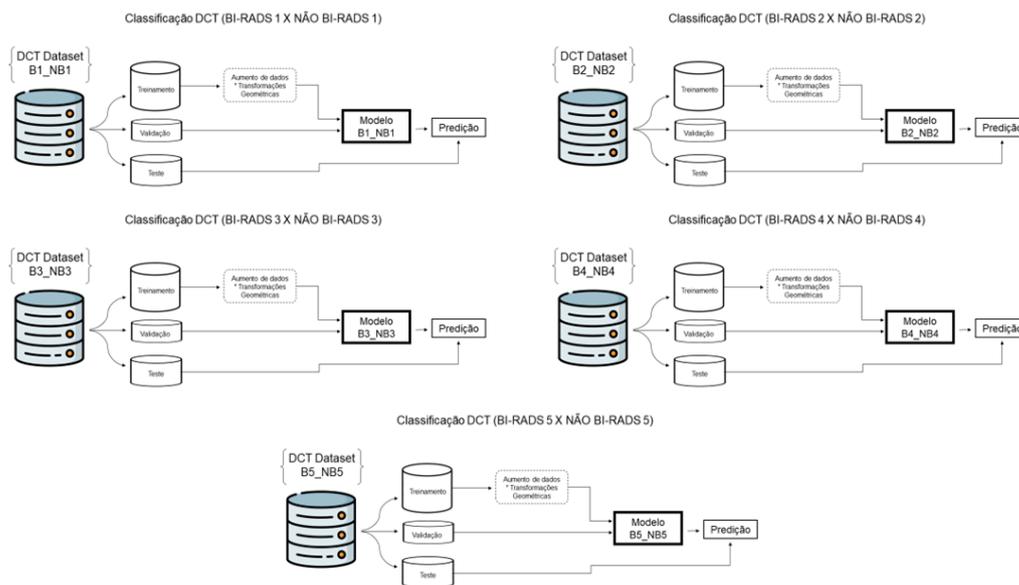
O treinamento e o teste dos modelos DCTr foram realizados também com as imagens dos Bancos CBIS-DDSM e INbreast, que foram sistematizadas inicialmente em cinco classes BI-RADS, e depois foram redefinidas em cinco *datasets* de estrutura binária, nos quais as classes foram organizadas para o treinamento de cada modelo específico, conforme é descrito na Tabela 7. Na classificação DCTr, se uma imagem é positiva na primeira classe, o modelo precisa diferenciá-la de imagens correspondentes às outras quatro categorias que estão mescladas na segunda classe.

Tabela 7 – Descrição dos Modelos para DCTr - Definição das classes

Modelo	Dataset	Classe 1	Classe 2
B1_NB1	SET-B1_NB1	BI-RADS 1	NÃO BI-RADS 1 (BI-RADS 2, BI-RADS 3, BI-RADS 4, BI-RADS 5)
B2_NB2	SET-B2_NB2	BI-RADS 2	NÃO BI-RADS 2 (BI-RADS 1, BI-RADS 3, BI-RADS 4, BI-RADS 5)
B3_NB3	SET-B3_NB3	BI-RADS 3	NÃO BI-RADS 3 (BI-RADS 1, BI-RADS 2, BI-RADS 4, BI-RADS 5)
B4_NB4	SET-B4_NB4	BI-RADS 4	NÃO BI-RADS 4 (BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 5)
B5_NB5	SET-B5_NB5	BI-RADS 5	NÃO BI-RADS 5 (BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4)

A Figura 14 apresenta as ilustrações referentes ao desenvolvimento dos cinco modelos de classificação da abordagem DCTr. Nas variações dos modelos DCTr, os conjuntos de dados binários (citados na Seção 4.2) foram particionados para treinamento, validação e teste, com aplicação de aumento de dados no respectivo conjunto de treinamento. Esses modelos também foram avaliados conforme as métricas de desempenho: Acurácia, precisão, *recall* e pontuação F1.

Figura 14 – DCTr - desenvolvimento do modelo (bloco pontilhado é opcional)



Fonte: Elaborada pela autora (2022)

Todo o carregamento das imagens para o treinamento foi feito com a randomização dos arquivos e as arquiteturas receberam ajuste na camada de saída para que o modelo treinado convergisse em cinco neurônios (classificação Multiclasse) ou em dois neurônios (classificação DCTr) conforme o objetivo do teste.

Toda a implementação das CNN's foi desenvolvida na Linguagem Python, utilizando-se as bibliotecas Tensorflow (ABADI *et al.*, 2015), Keras (CHOLLET *et al.*, 2015) e o framework Pytorch (Pytorch Community, 2022). O processamento das imagens foi realizado em uma máquina de 24 núcleos de CPU e 24GB de memória RAM, integrada ao Cluster Altix XE 1300 da Universidade do Estado do Rio Grande do Norte (UERN).

4.5 Considerações Finais

Neste capítulo foi apresentada a metodologia do trabalho, ressaltando-se os procedimentos de pré-processamento dos dados e as estratégias para desenvolvimento dos modelos de classificação das categorias BI-RADS.

No próximo capítulo serão apresentados os resultados do trabalho, juntamente com as discussões a respeito dos experimentos correspondentes à abordagem de classificação Multiclasse e à abordagem de classificação com DCTr.

5 RESULTADOS

Os experimentos do trabalho foram definidos pelo desenvolvimento e avaliação dos modelos Multiclasse e dos modelos com DCTr conforme os objetivos da classificação BI-RADS. Os testes foram organizados em três principais direcionamentos de classificação das imagens de mamografias: Classificação BI-RADS Multiclasse, Classificação BI-RADS com DCTr (imagem completa) e Classificação BI-RADS com DCTr (imagem ROI).

Durante a fase de treinamento foram realizados ajustes na taxa de aprendizagem (LR) do otimizador Adam e no número de épocas de treinamento. A função de perda foi definida como Entropia Cruzada Categórica e o desempenho dos modelos foi avaliado por meio das métricas de acurácia, precisão, *recall* e pontuação F1.

5.1 Classificação BI-RADS Multiclasse

Os modelos BI-RADS Multiclasse foram desenvolvidos para classificar as imagens de mamografias em cinco categorias, BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4 e BI-RADS 5. Estes modelos foram treinados com taxa de aprendizado 0,001 (*default* do otimizador Adam), em alguns testes foi aplicada a redução da taxa para 0,0001, porém não houve melhoria significativa no desempenho dos modelos. O ajuste no número de épocas de treinamento foi realizado conforme observação do nível de acurácia e de perda (*loss*) do modelo, seguindo o objetivo de ampliar o desempenho sem causar *overfitting* da aprendizagem.

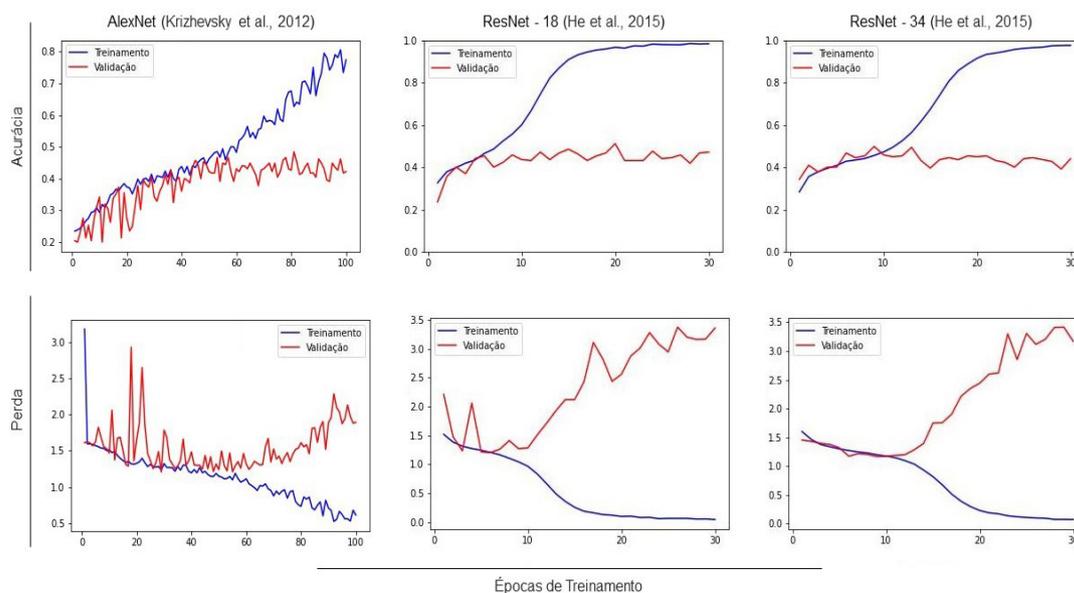
Parte dos testes foram realizados com modelos pré-treinados em *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*, ou seja, os pesos da rede neural foram inicializados no conjunto ILSVRC e depois foram transferidos para o treinamento com as mamografias. Esse processo é conhecido como Transferência de Aprendizagem. No entanto, nesse primeiro estudo, não foi possível obter uma aprendizagem eficiente em nenhuma das arquiteturas CNN. Identificou-se que a taxa de acurácia não estava evoluindo ao longo das épocas de treinamento, mesmo com ajuste na taxa de aprendizado. Portanto esses resultados não serão discutidos nesse trabalho, porém em trabalhos futuros serão realizados novos testes para avaliar essa estratégia de aprendizado para a classificação BI-RADS.

Por outro lado, nos modelos desenvolvidos sem pré-treinamento foi possível verifi-

car uma aprendizagem adequada do conjunto de dados, observando-se que a acurácia dos modelos estava se ampliando de forma gradual durante a fase de treino. Cinco modelos de cada arquitetura CNN foram treinados e avaliados nos testes da classificação Multiclasse. Na Figura 15 é possível visualizar o melhor comportamento da curva de acurácia e da função de perda durante as épocas de treinamento, de acordo com as arquiteturas.

Os modelos AlexNet foram treinados por 100 épocas, enquanto os modelos ResNet-18 e ResNet-34 foram treinados por 30 épocas, todos sem pré-treinamento. Os resultados correspondentes aos experimentos complementares (Apêndice A) com as arquiteturas CNN serviram de suporte para determinar o melhor desempenho dos modelos que serão apresentados na Tabela 8, Tabela 9 e Tabela 10.

Figura 15 – Acurácia e Perda de Treinamento dos Modelos BI-RADS Multiclasse



Fonte: Elaborada pela autora (2022)

O desempenho dos modelos foi avaliado conforme as métricas de Acurácia, Precisão, *Recall* e Pontuação F1, agrupadas por arquitetura. Essas métricas de desempenho relatam a relação entre as taxas de Verdadeiro Positivo (TP), Verdadeiro Negativo (TN), Falso Positivo (FP) e Falso Negativo (FN) na predição das amostras.

A Acurácia (*Accuracy*) determina quantos exemplos da amostra foram de fato classificados corretamente, independentemente da classe. A Precisão (*Precision*) é definida pela razão entre a quantidade de exemplos classificados corretamente como positivos e o total de exemplos classificados como positivos. A taxa de *Recall*, também conhecida como Sensibilidade ou Taxa de Verdadeiro Positivo (TPR) é descrita pela razão entre

a quantidade de exemplos classificados corretamente como positivos e a quantidade de exemplos que são de fato positivos. E a Pontuação F1 (*F1-score*) é definida pela média harmônica entre a Precisão e o *Recall*.

$$\text{Acurácia (Acc)} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precisão (Prec)} = \frac{TP}{TP+FP}$$

$$\text{Recall (Rec)} = \frac{TP}{TP+FN}$$

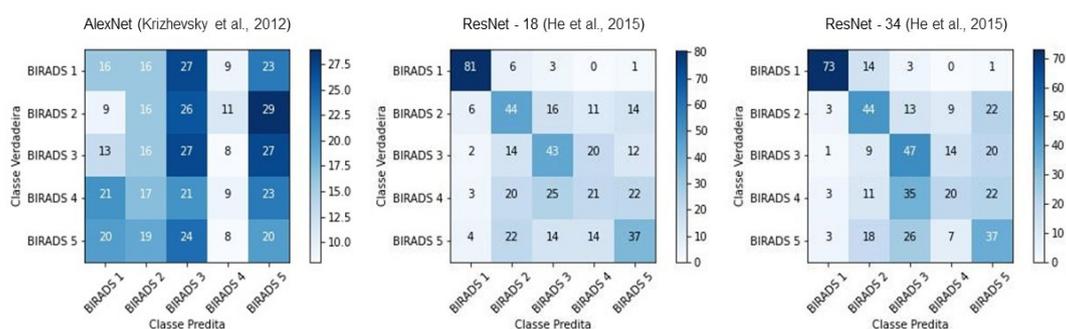
$$\text{Pontuação F1 (F1)} = 2 \times \frac{\text{Prec} \times \text{Rev}}{\text{Prec} + \text{Rev}}$$

Na Tabela 8 destaca-se a média e o desvio padrão das métricas de desempenho, sendo possível observar que o desempenho dos modelos foi insuficiente para a classificação BI-RADS em cinco classes. Mesmo nas arquiteturas ResNet, o melhor resultado alcançado ainda foi inferior a 50% nas métricas. Na Figura 16 é possível visualizar a matriz de confusão dos modelos estruturados em cada arquitetura, sendo destacável a elevada taxa de erros na predição das classes BI-RADS.

Tabela 8 – Classificação BI-RADS (Multiclasse)

	Acurácia	Precisão	Recall	Pontuação F1
AlexNet	0,207 ±0,014	0,138 ±0,092	0,206 ±0,013	0,146 ±0,070
ResNet-18	0,489 ±0,022	0,498 ±0,022	0,488 ±0,022	0,484 ±0,021
ResNet-34	0,485 ±0,011	0,498 ±0,022	0,484 ±0,011	0,476 ±0,018

Figura 16 – Matriz de Confusão dos Modelos BI-RADS Multiclasse



Fonte: Elaborada pela autora (2022)

Considerando que para a abordagem Multiclasse, mesmo com desempenho insuficiente, as redes ResNet ainda apresentaram os melhores resultados em comparação à AlexNet, então foi decidido reavaliar o desempenho dos modelos ResNet-18 e ResNet-34 por meio da validação cruzada, um procedimento de avaliação que é realizado a partir da repetição computacional de treinamento e teste em diferentes subconjuntos ou divisões

(*k-fold*) do conjunto de dados original escolhidos aleatoriamente. Na tentativa *i*, o *i*-ésimo subconjunto de dados é usado como conjunto de teste e o restante dos dados é usado como conjunto de treinamento (GOODFELLOW *et al.*, 2016).

Para a validação cruzada dos modelos ResNet-18 e ResNet-34 utilizou-se um *k-fold* = 5, ou seja, o *dataset* de imagens foi dividido em cinco partes e cada modelo foi retreinado cinco vezes, obedecendo-se a distribuição sequencial de 1/5 para teste e 4/5 para treinamento. Nos resultados obtidos com a validação cruzada, os modelos da arquitetura ResNet-18 tiveram uma acurácia média de $0,46261 \pm 0,040$, e os modelos ResNet-34 demonstraram uma acurácia média de $0,44851 \pm 0,017$. Portanto é possível concluir que o desempenho dos modelos Multiclasse foi insuficiente para a classificação precisa das mamografias em cinco categorias BI-RADS, nesse primeiro momento de testes.

Considerando os resultados obtidos com os modelos Multiclasse, afirma-se que em futuros experimentos serão realizados outros ajustes das redes e do pré-processamento de dados, para que seja possível atingir um desempenho elevado para a classificação das mamografias, de acordo com as categorias BI-RADS.

Em continuidade aos objetivos desse trabalho, e em decorrência dos resultados produzidos com a abordagem Multiclasse, foram realizados novos experimentos de classificação, aplicando-se a abordagem *Deep Cross-Training* na fase de treinamento dos modelos, visando ampliar o desempenho dos classificadores em relação à predição das cinco categorias da avaliação BI-RADS.

5.2 Classificação BI-RADS com *Deep Cross-Training* (DCTr)

Levando em consideração que na primeira fase de testes os modelos multiclasse apresentaram um desempenho insuficiente na classificação das mamografias, foram realizados novos experimentos, com modelos binários, treinados com DCTr para ampliar a precisão de classificação das mamografias conforme as cinco categorias BI-RADS.

Com base na Revisão Sistemática da Literatura (RSL) que foi realizada para a fundamentação teórica desse trabalho, identificou-se que a maioria dos estudos publicados entre 2015 e 2021 se concentram no desenvolvimento da classificação de imagens mamográficas com abordagem binária, geralmente com classes benigno e maligno, massa ou não massa. Nos estudos que desenvolveram a classificação binária das imagens é possível constatar que as Redes Neurais Convolucionais possibilitam o alcance de resultados com

precisão elevada na classificação das mamografias.

Considerando o estado da arte no que diz respeito à aplicação do Aprendizado Profundo para a análise do câncer de mama, neste trabalho foi desenvolvida uma abordagem de classificação cruzada, denominada de *Deep Cross-Training* (DCTr), a partir do treinamento de 30 modelos baseados nas arquiteturas AlexNet (KRIZHEVSKY *et al.*, 2012b), ResNet-18 (HE *et al.*, 2015) e ResNet-34 (HE *et al.*, 2015), com o objetivo de aumentar a precisão na predição das imagens de mamografias em comparação aos resultados da abordagem Multiclasse.

A abordagem DCTr é baseada na estratégia de classificação *One Vs Rest* (OvR) (SCIKIT-LEARN, 2022), na qual algoritmos projetados para classificação binária podem ser adaptados para problemas multiclasse. O fundamento da estratégia OvR consiste em definir um classificador para cada classe. Dessa forma, em cada classificador, uma classe específica é ajustada em relação a todas as outras classes e, simultaneamente, essa classe torna-se representada pelo respectivo classificador. Na abordagem DCTr, o procedimento OvR é executado através das CNN's, objetivando-se o aprimoramento da precisão dos classificadores em relação às classes BI-RADS.

As Redes Neurais permitem a classificação multiclasse, e recorrentemente demonstram resultados ótimos na resolução desse tipo de problema. No entanto para alguns problemas com mais de duas classes, torna-se difícil atingir uma precisão elevada na classificação de imagens, seja pela natureza do problema ou por limitações no ajuste da Rede.

Nesse trabalho, os experimentos correspondentes à abordagem DCTr foram definidos pelo treinamento e teste dos modelos de classificação binária para cada uma das cinco classes BI-RADS (BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4 e BI-RADS 5) que estruturam o conjunto de imagens de mamografias. No treinamento de cada modelo, uma classe BI-RADS é ajustada em relação a todas as outras classes. Portanto, em cada arquitetura CNN, inicialmente foram desenvolvidos cinco modelos de classificação para imagens completas, e posteriormente mais cinco modelos de cada arquitetura foram treinados para a classificação das imagens correspondentes às Regiões de Interesse (ROI) das mamografias.

A aplicação da abordagem DCTr nos experimentos executados neste trabalho funcionou como alternativa para o processo de análise das imagens, contribuindo para a

ampliação do desempenho dos modelos e favorecendo a classificação das mamografias para a identificação das lesões, de acordo a avaliação BI-RADS.

Os experimentos correspondentes à abordagem DCTr foram definidos pelo desenvolvimento e avaliação dos modelos CNN que possuem apenas duas saídas. A primeira equivale à classe positiva para um BI-RADS específico que será diferenciado dos outros quatro BI-RADS mesclados na segunda classe, que é ajustada na segunda saída da rede.

A Tabela 9 apresenta o desempenho dos modelos, conforme as arquiteturas AlexNet, ResNet-18 e ResNet-34 para a classificação dos pares (BI-RADS 1 / NÃO BI-RADS 1, BI-RADS 2 / NÃO BI-RADS 2, BI-RADS 3 / NÃO BI-RADS 3, BI-RADS 4 / NÃO BI-RADS 4 e BI-RADS 5 / NÃO BI-RADS 5), com uso da imagem completa (*img full*), ressaltando a média e o desvio padrão (STD) em cada métrica de avaliação. Todos os modelos foram treinados por 50 épocas e com taxa de aprendizado de 0,001.

Tabela 9 – Classificação BI-RADS com DCTr (*img full*)

		AlexNet				ResNet-18				ResNet-34			
Modelo	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	
B1_NB1	0,532	0,540	0,530	0,490	0,920	0,920	0,920	0,920	0,907	0,920	0,910	0,910	
B2_NB2	0,509	0,510	0,510	0,510	0,626	0,660	0,630	0,610	0,626	0,640	0,630	0,620	
B3_NB3	0,531	0,530	0,530	0,530	0,540	0,550	0,540	0,540	0,577	0,590	0,580	0,580	
B4_NB4	0,543	0,540	0,540	0,530	0,625	0,640	0,620	0,600	0,662	0,660	0,660	0,660	
B5_NB5	0,463	0,450	0,460	0,440	0,541	0,540	0,540	0,540	0,592	0,600	0,590	0,580	
Média	0,516	0,514	0,514	0,500	0,650	0,662	0,650	0,642	0,673	0,682	0,650	0,670	
STD	0,031	0,037	0,032	0,037	0,156	0,153	0,156	0,158	0,135	0,136	0,156	0,138	

Observando o desempenho dos modelos na classificação com DCTr (Tabela 9), e traçando um comparativo com os resultados da classificação Multiclasse (Tabela 8), é possível afirmar que houve um significativo aprimoramento na predição das mamografias para as classes BI-RADS, evidenciando a arquitetura ResNet-34, na qual os modelos DCTr obtiveram uma precisão média de $0,682 \pm 0,136$, enquanto na abordagem Multiclasse, a média da precisão para os modelos dessa arquitetura ficou limitada em $0,498 \pm 0,022$. Em relação à acurácia, na arquitetura ResNet-34 com abordagem Multiclasse foi obtida uma média de $0,485 \pm 0,011$, enquanto na abordagem DCTr observou-se uma média de $0,673 \pm 0,135$, o que significa que houve um aumento de 38,76% no desempenho dos modelos dessa arquitetura, com treinamento de imagens completas pertencentes aos dois Bancos, CBIS-DDSM e INbreast.

Seguindo a mesma abordagem de classificação BI-RADS com DCTr para imagens completas, foi decidido observar o desempenho dos modelos com as imagens ROI das

mamografias, visando alcançar um desempenho melhor dos classificadores com a análise de imagens em campo reduzido. Para esses novos testes foram adotadas 954 ROI's das mamografias pertencentes ao Banco CBIS-DDSM, com balanceamento e sem aplicação de aumento de dados. Todos os modelos binários foram retreinados por 50 épocas e com taxa de aprendizado de 0,001.

Na Tabela 10 destacam-se os novos resultados a partir das métricas de desempenho, e através das médias calculadas é possível afirmar que os modelos da arquitetura ResNet-18 obtiveram o melhor desempenho para a classificação BI-RADS com DCTr de imagens em ROI, evidenciando-se a precisão média de $0,752 \pm 0,162$. Na arquitetura ResNet-18, também observa-se que a média da acurácia foi elevada de $0,650 \pm 0,156$ para $0,745 \pm 0,171$, resultante do treinamento com as imagens ROI.

Tabela 10 – Classificação BI-RADS com DCTr (*img ROI*)

AlexNet					ResNet-18				ResNet-34			
Modelo	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
B1_NB1	0,515	0,520	0,510	0,440	0,989	0,990	0,990	0,990	0,973	0,970	0,970	0,970
B2_NB2	0,522	0,590	0,520	0,410	0,843	0,840	0,840	0,840	0,812	0,810	0,810	0,810
B3_NB3	0,516	0,520	0,520	0,510	0,666	0,670	0,670	0,670	0,666	0,670	0,670	0,670
B4_NB4	0,474	0,470	0,470	0,470	0,552	0,580	0,550	0,510	0,526	0,540	0,530	0,490
B5_NB5	0,472	0,470	0,470	0,470	0,677	0,680	0,680	0,680	0,682	0,690	0,680	0,680
Média	0,500	0,514	0,498	0,460	0,745	0,752	0,746	0,738	0,732	0,736	0,732	0,724
STD	0,024	0,049	0,025	0,037	0,171	0,162	0,170	0,182	0,168	0,162	0,165	0,178

Na Tabela 11 destaca-se especificamente a média da acurácia correspondente a cada arquitetura CNN no processo de classificação BI-RADS na abordagem Multiclasse e na abordagem DCTr (*full* e ROI). Na classificação com DCTr de imagens ROI, as arquiteturas ResNet-18 e ResNet-34 obtiveram uma acurácia média bastante aproximadas. Então, para avaliar a diferença de desempenho em nível estatístico de significância, foi aplicado o Teste de Friedman, que é um teste estatístico não-paramétrico desenvolvido por Milton Friedman, utilizado para avaliar se existem diferenças estatisticamente significativas entre as distribuições de três ou mais grupos pareados (ZAIONTZ, 2022).

Tabela 11 – Acurácia dos modelos - Classificação BI-RADS

Classificação	AlexNet	ResNet-18	ResNet-34
Acurácia			
Multiclasse (<i>img full</i>)	$0,207 \pm 0,014$	$0,489 \pm 0,022$	$0,485 \pm 0,011$
DCTr (<i>img full</i>)	$0,516 \pm 0,031$	$0,650 \pm 0,156$	$0,673 \pm 0,135$
DCTr (<i>img ROI</i>)	$0,500 \pm 0,024$	$0,745 \pm 0,171$	$0,732 \pm 0,168$

5.2.1 Teste de Friedman - Análise dos Classificadores

Diante dos resultados obtidos com os modelos DCTr (img ROI), descritos na Tabela 11, buscou-se avaliar a partir do Teste de Friedman se existe diferença estatisticamente significativa entre os classificadores, principalmente entre os modelos ResNets que apresentaram um desempenho muito aproximado. O Teste de Friedman é a alternativa não paramétrica à Análise de Variância (ANOVA) de uma via com medidas repetidas. Esse teste pode ser aplicado quando se pretende analisar três ou mais grupos pareados que são avaliados por meio de uma variável quantitativa, que não obedeça a parâmetros normais de distribuição ou por meio de uma variável qualitativa ordinal (ZAIONTZ, 2022).

A propriedade que formaliza o teste é definida pela Equação Q (Equação 5.1), onde k é o número de grupos, n é o número de sujeitos, R_j é a soma das classificações para o grupo j (ZAIONTZ, 2022). Essa propriedade caracteriza a diferença geral estatisticamente significativa entre as classificações médias de seus grupos relacionados.

$$Q = \frac{12}{nk(k+1)} \sum_{j=1}^k R_j^2 - 3n(k+1) \quad (5.1)$$

Neste trabalho, o teste de Friedman foi executado com o auxílio do software Real Statistics Using Excel (ZAIONTZ, 2022), considerando os resultados obtidos nas métricas de desempenho (Acurácia, Precisão, *Recall* e Pontuação F1) dos modelos treinados com as ROI's na abordagem de Classificação com DCTr, e um nível de significância (α) = 0,05 para o cálculo da distribuição qui-quadrado (Q) do p-valor. Além disso, duas hipóteses exclusivas foram definidas para formalizar o procedimento comparativo dos modelos:

- Hipótese nula (H_0): "Os classificadores são iguais".
- Hipótese alternativa (H_1) "Os classificadores são diferentes".

Os resultados iniciais do teste indicaram um valor $Q = 8$ e um p-valor = 0,018345639 para o nível de significância de 0,05. Tomando como referência o nível de significância adotado para a comparação dos classificadores, percebe-se que o p-valor obtido é menor que o (α) (0,05). Assim, é possível afirmar que existe diferença estatisticamente significativa entre os modelos das três arquiteturas avaliadas e a Hipótese nula (H_0) é rejeitada. Consequentemente a Hipótese alternativa (H_1) foi aceita, e portanto, tornou-se relevante aplicar um teste *Post-hoc* para examinar onde as diferenças realmente ocorrem no grupo de classificadores.

O Teste *Post-hoc* selecionado foi o teste de Nemenyi (Wilcoxon-Nemenyi-McDonald-Thompson). O Teste de Nemenyi é uma adaptação do Teste Tukey HSD, sendo um teste de comparação múltipla que é usado após a aplicação de teste não paramétricos com três ou mais fatores (LIU; CHEN, 2012). No procedimento comparativo, a diferença entre as somas de classificação é verificada através do seguinte erro padrão:

$$q = \frac{|\bar{R}_i - \bar{R}_j|}{s.e.}; \quad s.e. = \sqrt{\frac{k(n+1)}{12}} \quad (5.2)$$

Onde k é igual ao número de grupos e n é igual ao tamanho de cada uma das amostras do grupo. O tamanho das amostras do grupo devem ser iguais.

Na tabela de classificações do Teste de Nemenyi, a arquitetura AlexNet pontuou 4 pontos, a ResNet-18 marcou 12 e a ResNet-34 acumulou 8 pontos. Os modelos da arquitetura ResNet-18 obtiveram a melhor classificação (R soma = 12), no entanto, conforme descrito na Tabela 12, somente foi possível identificar uma diferença estatisticamente significativa na comparação entre os modelos AlexNet e ResNet-18, pois o p-valor é igual 0,013092 (menor que $\alpha = 0,05$). Enquanto na comparação entre a ResNet-18 e a ResNet-34 é possível afirmar que não existe diferença estatisticamente significativa em relação ao desempenho dos modelos, pois no teste *Post-hoc* o p-valor (0,333674) obtido foi maior que o nível de significância (α).

Tabela 12 – Teste Friedman-Nemenyi

a) Friedman-Nemenyi (alfa = 0,05)					
Grupo	R soma	tamanho	std err	q-crit	R-crit
AlexNet	4	4			
ResNet-18	12	4			
Resnet-34	8	4			
			2	3,314	6,628

b) Teste Q				
Grupo 1	Grupo 2	R soma	q-stat	p-valor
AlexNet	ResNet-18	8	4	0,013092
AlexNet	ResNet-34	4	2	0,333674
ResNet-18	ResNet-34	4	2	0,333674

Fonte: Elaborada pela autora (2022)

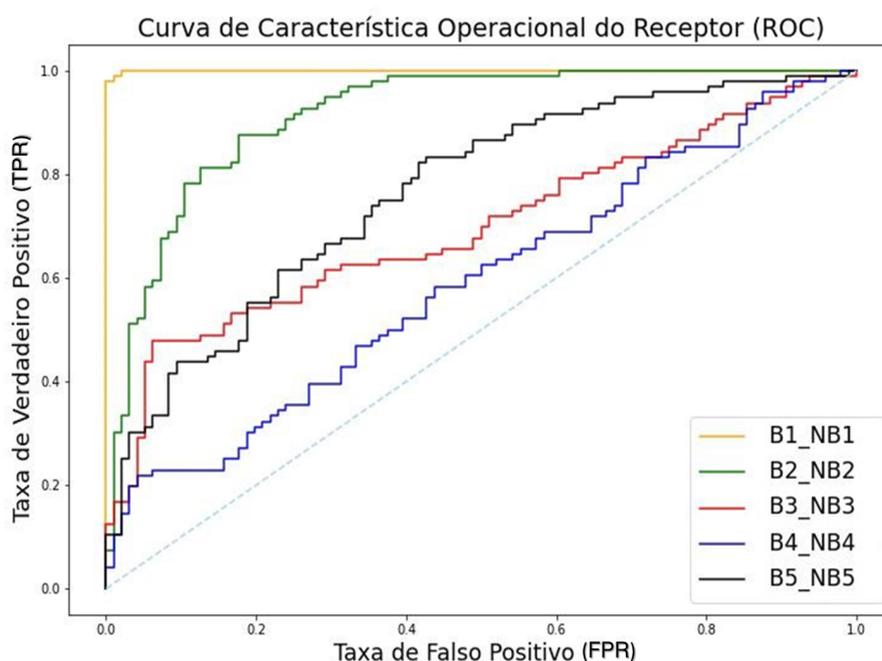
Levando em consideração que o teste estatístico de Friedman-Nemenyi não revelou diferença estatisticamente significativa entre os modelos ResNet, foi selecionada por

critério de média de desempenho, a arquitetura ResNet-18 para se observar em cada modelo binário a relação entre a taxa de verdadeiro positivo (TPR) e taxa de falso positivo (FPR) na classificação de ROI's das mamografias.

5.2.2 ResNet-18 - Análise em Curva ROC

O desempenho dos modelos ResNet-18 na classificação com DCTr (img ROI) das categorias BI-RADS também foi avaliado por meio da Curva de Característica Operacional do Receptor (*Receiver Operating Characteristic - ROC*), que representa a relação entre a taxa de verdadeiro positivo e a taxa de falso positivo. Na visualização gráfica da Figura 17, quanto mais próxima a curva estiver do canto superior esquerdo, maior será a Área Sob a Curva (AUC), ou seja, melhor o desempenho do classificador para o conjunto teste. A elevação da AUC representa uma maximização da taxa de verdadeiros positivos e a minimização da taxa de falsos positivos.

Figura 17 – Curva ROC - Classificação BI-RADS com DCTr (ROI) [ResNet-18]



Fonte: Elaborada pela autora (2022)

De acordo com o comportamento da curva ROC, os melhores resultados em AUC foram alcançados pelos modelos B1_NB1 (AUC: 0,99967) e B2_NB2 (AUC: 0,91666), que fazem a classificação positiva dos BI-RADS 1 e BI-RADS 2 respectivamente. O

terceiro melhor resultado foi identificado no modelo B5_NB5 (AUC: 0,76399), que realiza a classificação do BI-RADS 5 em relação aos demais. O modelo B3_NB3 (AUC: 0,69661) treinado para a classificação positiva do BI-RADS 3 e o modelo B4_NB4 (AUC: 0,59027), que classifica as imagens de acordo com o BI-RADS 4, foram os classificadores que originaram as menores áreas de precisão.

A dificuldade do modelo B4_NB4 obter uma precisão elevada na classificação do BI-RADS 4 pode ser interpretada como uma limitação do classificador para distinguir as anormalidades dentro de um intervalo muito longo de variações de malignidade para essa categoria. Relembrando que pelo protocolo de avaliação BI-RADS, a categoria 4 pode variar de 2% a 95% de chance de malignidade, e mesmo que nesse trabalho as subdivisões do BI-RADS 4 (4A, 4B, e 4C) não tenham sido consideradas, percebe-se que a predição para essa classe é mais complexa de ser predita que as demais classes do conjunto.

A Tabela 13 contém a especificação do desempenho dos modelos de acordo com a pontuação F1 que representa a média harmônica entre a precisão e *recall* conforme o BI-RADS classificado, sendo possível ratificar a limitação dos classificadores na predição da categoria BI-RADS 4, na qual o melhor resultado alcançado na métrica em observação foi 0,510 no modelo de arquitetura ResNet-18, considerando a classificação com DCTr usando imagens ROI.

Tabela 13 – Pontuação F1 da Classificação BI-RADS com DCTr (img ROI)

	Pontuação F1				
	BI-RADS 1	BI-RADS 2	BI-RADS 3	BI-RADS 4	BI-RADS 5
AlexNet	0,440	0,410	0,510	0,470	0,470
ResNet-18	0,990	0,840	0,670	0,510	0,680
ResNet-34	0,970	0,810	0,670	0,490	0,680

O desempenho dos modelos DCTr baseados em ResNet-18 foi reavaliado a partir de uma análise comparativa individual entre a classificação de imagens completas e a classificação de imagens ROI para o conjunto CBIS-DDSM. O treinamento dos modelos binários com as imagens completas do CBIS-DDSM foi executado também em 50 épocas, com taxa de aprendizado 0,001. No desenvolvimento desses modelos foram aplicadas 490 imagens para o treinamento, 68 imagens para validação e 140 para teste.

Na Tabela 14 são apresentados os resultados correspondentes ao desempenho dos modelos para a classificação das imagens completas (IMG FULL) e para as imagens em ROI (IMG ROI). Com base nas métricas de desempenho, é possível identificar que a

redução do campo de análise das imagens possibilitou que a acurácia média se elevasse de $0,621 \pm 0,161$ (IMG FULL) para $0,745 \pm 0,171$ (IMG ROI), um aumento de 19,96%.

Em experimentos futuros será avaliado o aprimoramento de desempenho dos modelos Multiclasse para os modelos DCTr com testes individuais em cada conjunto de imagens, sendo assim possível definir uma comparação específica e identificar melhorias significativas à respeito da classificação das categorias BI-RADS.

Tabela 14 – Classificação BI-RADS com DCTr - ResNet-18

ResNet-18								
Dataset CBIS-DDSM								
Modelo	IMG FULL				IMG ROI			
	Acc.	Prec.	Rec.	F1	Acc	Prec	Rec	F1
B1_NB1	0,979	0,980	0,980	0,980	0,989	0,990	0,990	0,990
B2_NB2	0,700	0,740	0,700	0,690	0,843	0,840	0,840	0,840
B3_NB3	0,621	0,660	0,620	0,600	0,666	0,670	0,670	0,670
B4_NB4	0,586	0,680	0,590	0,530	0,552	0,580	0,550	0,510
B5_NB5	0,614	0,620	0,610	0,610	0,677	0,680	0,680	0,680
Média	0,621	0,680	0,620	0,610	0,745	0,752	0,746	0,738
STD	0,161	0,143	0,162	0,175	0,171	0,162	0,170	0,182

Fonte: Elaborada pela autora (2022)

Realizando um procedimento comparativo parcial dos resultados apresentados neste trabalho, com os resultados de estudos similares (Ver Tabela 15), é possível destacar como principal limitação do estudo, a dificuldade de alcançar resultados semelhantes aos divulgados por Domingues *et al.* (2018) e Siddeeq *et al.* (2021) para a classificação BI-RADS em 5 classes, principalmente na primeira abordagem de classificação citada no Seção 5.1. A comparação entre os trabalhos é parcial em virtude da variação dos conjuntos de imagens manipuladas nos experimentos deste trabalho em relação aos trabalhos citados.

Tabela 15 – Comparação com estudos semelhantes

Referência	Objetivo	Dados	Arquitetura	Resultados
Domingues <i>et al.</i> (2018)	BI-RADS 1 a 6 (FULL)	INbreast	AlexNet	Acc: 0,834
Siddeeq <i>et al.</i> (2021)	BI-RADS 1 a 6 (FULL)	INbreast	RN-BCNN	Acc: 0,859
Este Trabalho	BI-RADS 1 a 5 (FULL)	CBIS-DDSM e INbreast	ResNet-18	Acc: $0,489 \pm 0,021$
Este Trabalho	BI-RADS 1 a 5 com DCTr (FULL)	CBIS-DDSM e INbreast	ResNet-34	Acc: $0,673 \pm 0,138$
Este Trabalho	BI-RADS 1 a 5 com DCTr (ROI)	CBIS-DDSM	ResNet-18	Acc: $0,745 \pm 0,182$

É possível ressaltar também que a abordagem DCTr para a classificação das categorias BI-RADS em pares associados permitiu uma elevação significativa no desempenho dos modelos para a classificação das mamografias, viabilizando que em trabalhos futu-

ros resultados melhores possam ser alcançados, com a aplicação de outras técnicas de pré-processamento, aprimoramento do *dataset* de treinamento, além da implementação de outras arquiteturas de Redes Neurais Convolucionais.

5.3 Considerações Finais

Neste capítulo foram apresentados os resultados do trabalho, constituídos da descrição e avaliação do desempenho dos classificadores baseados em três arquiteturas de Rede Neural Convolucional, AlexNet, ResNet-18 e ResNet-34.

Nos experimentos relatados, duas abordagens de classificação das mamografias foram executadas, a Classificação Multiclasse e a Classificação com DCTr, caracterizando estratégias de análise para predição das categorias de avaliação BI-RADS.

6 CONCLUSÕES

A avaliação médica de mamografias para identificar o câncer de mama é um processo complexo, com limitações técnicas e biológicas. A presente pesquisa teve como objetivo desenvolver uma metodologia e um modelo de Aprendizado Profundo para a classificação de lesões mamárias, com base no protocolo de avaliação BI-RADS, visando auxiliar médicos radiologistas no processo de tomada de decisão acerca do diagnóstico do câncer de mama.

A Revisão Sistemática da Literatura permitiu ampliar o conhecimento sobre o desenvolvimento de modelos de Aprendizado Profundo, baseado em Redes Neurais Convolucionais, além de retornar trabalhos relevantes para o contexto desta pesquisa. As principais pesquisas que foram avaliadas na revisão bibliográfica serviram como referencial técnico para os procedimentos práticos dos experimentos, constituídos por etapas de pré-processamento dos dados, treinamento das Redes Neurais, e avaliação dos modelos de classificação a partir de métricas de desempenho.

O trabalho foi desenvolvido em duas abordagens de classificação das imagens de mamografias. A primeira abordagem refere-se à Classificação Multiclasse, constituída pelo processo de implementação e avaliação dos modelos com cinco saídas da rede, responsáveis pela classificação das imagens em cinco categorias da avaliação BI-RADS. A segunda abordagem, a Classificação com *Deep Cross-Training* (DCTr) corresponde à estratégia de treinamento de modelos binários específicos para cada classe BI-RADS. A estratégia DCTr é uma contribuição desta pesquisa para o processo de classificação das imagens mamográficas, a qual foi desenvolvida para ampliar o desempenho dos classificadores e obter uma predição mais precisa para as classes de análise.

A abordagem DCTr para a classificação das mamografias gerou os melhores resultados de predição, na qual foi alcançada uma acurácia média de $0,745 \pm 0,171$, precisão de $0,752 \pm 0,162$, *recall* de $0,746 \pm 0,170$ e pontuação F1 de $0,738 \pm 0,182$ nos modelos da arquitetura ResNet-18, com uso das imagens em ROI do Banco CBIS-DDSM. Na análise exclusiva das imagens dessa Base por meio da estratégia DCTr é possível destacar que o treinamento com as ROI ampliou a acurácia média em 19,96% em comparação a análise realizada com as imagens completas.

No comparativo entre o desempenho dos modelos de classificação Multiclasse e os modelos de classificação com DCTr, também ressalta-se que os modelos DCTr baseados

em ResNet-34 obtiveram uma precisão média de $0,682 \pm 0,136$, enquanto na abordagem Multiclasse, a média da precisão para os modelos dessa arquitetura ficou limitada em $0,498 \pm 0,022$. Desse modo, afirma-se que a abordagem DCTr possibilitou um aprimoramento de 36,94% na precisão dos modelos ResNet-34 para a predição das classes BI-RADS em imagens completas.

Dentre as principais limitações do trabalho, é relevante enfatizar a dificuldade para elevar a precisão dos modelos de classificação, principalmente em relação à categoria BI-RADS 4 aos níveis de precisão relatados em trabalhos relacionados. Tornando-se relevante desenvolver em futuros experimentos novos procedimentos de pré-processamento e de ajustes das Redes Profundas.

Destaca-se também como limitação da pesquisa a aquisição de um grande conjunto de imagens de mamografias para treinar uma Rede Neural Convolucional. Um modelo de Aprendizado Profundo requer uma grande quantidade de dados para se desenvolver uma aprendizagem eficiente e a obtenção de imagens médicas em grande volume é um processo desafiador.

Em trabalhos futuros pretende-se realizar o aprimoramento da aprendizagem dos modelos preditivos com a aquisição de novas imagens de mamografias, visando o desenvolvimento de classificadores precisos que sejam capazes de auxiliar o processo de reconhecimento de lesões malignas em imagens de mamografias. Afirma-se também a pretensão de realizar a avaliação de outras arquiteturas de Redes Neurais Convolucionais para a classificação das categorias de avaliação BI-RADS, incluindo as subcategorias 4A, 4B e 4C para o diagnóstico das mamografias.

Ademais, ressalta-se que alguns dos objetivos de pesquisa definidos neste trabalho serão continuados em um projeto de pesquisa também com foco no estudo das mamografias, atualmente vinculado à Liga Mossoroense de Estudos e Combate ao Câncer.

REFERÊNCIAS

- ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z. *et al.* **TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems**. 2015. Disponível em: <<http://download.tensorflow.org/paper/whitepaper2015.pdf>>.
- ABDELHAFIZ, D.; YANG, C.; AMMAR, R.; NABAVI, S. Deep convolutional neural networks for mammography: advances, challenges and applications. **BMC Bioinformatics**, Springer Science and Business Media LLC, v. 20, n. S11, jun. 2019. Disponível em: <<https://doi.org/10.1186/s12859-019-2823-4>>.
- ACS. **Breast Cancer Early Detection and Diagnosis**. American Cancer Society, 2019. Disponível em: <<https://www.cancer.org/content/dam/CRC/PDF/Public/8579.00.pdf>>.
- AGARWAL, R.; DÍAZ, O.; YAP, M. H.; LLADÓ, X.; MARTÍ, R. Deep learning for mass detection in full field digital mammograms. **Computers in Biology and Medicine**, Elsevier BV, v. 121, p. 103774, jun. 2020. Disponível em: <<https://doi.org/10.1016/j.combiomed.2020.103774>>.
- AGGARWAL, C. C. Data classification. In: **Data Mining**. Springer International Publishing, 2015. p. 285–344. Disponível em: <https://doi.org/10.1007/978-3-319-14142-8_10>.
- AGGARWAL, C. C. **Neural Networks and Deep Learning**. [S.l.]: Springer, 2018.
- AYDINER, A.; IGCI, A.; SORAN, A. (Ed.). **Breast Cancer - A Guide to Clinical Practice**. Springer International Publishing, 2019. Disponível em: <<https://doi.org/10.1007/978-3-319-96947-3>>.
- BAKALO, R.; GOLDBERGER, J.; BEN-ARI, R. Weakly and semi supervised detection in medical imaging via deep dual branch net. **Neurocomputing**, Elsevier BV, v. 421, p. 15–25, jan. 2021. Disponível em: <<https://doi.org/10.1016/j.neucom.2020.09.037>>.
- BALLEYGUIER, C.; AYADI, S.; NGUYEN, K. V.; VANEL, D.; DROMAIN, C.; SIGAL, R. BIRADS™ classification in mammography. **European Journal of Radiology**, Elsevier BV, v. 61, n. 2, p. 192–194, fev. 2007. Disponível em: <<https://doi.org/10.1016/j.ejrad.2006.08.033>>.
- BANSAL, G. J.; THOMAS, K.; LIM, K. Imaging techniques in breast cancer. **Surgery (Oxford)**, Elsevier BV, v. 31, n. 1, p. 15–21, jan. 2013. Disponível em: <<https://doi.org/10.1016/j.mpsur.2012.10.011>>.
- CALAS, M. J. G.; GUTFILEN, B.; PEREIRA, W. C. de A. CAD e mamografia: por que usar esta ferramenta? **Radiologia Brasileira**, FapUNIFESP (SciELO), v. 45, n. 1, p. 46–52, fev. 2012. Disponível em: <<https://doi.org/10.1590/s0100-39842012000100011>>.
- CAMPOS, M.; MARTINEZ, J. M. S.; CARRON, C. B.; GUIROLA, J. A.; GOMEZ, J. A. F.; PEREZ, J. S. Tracks to face a breast imaging and succeed. in european society of radiology (ecr). 2013.
- CARNEIRO, G.; NASCIMENTO, J.; BRADLEY, A. P. Deep learning models for classifying mammogram exams containing unregistered multi-view images and

segmentation maps of lesions. this work is an extension of the paper published by the same authors at the medical image computing and computer-assisted intervention (MICCAI 2015) [1]. In: **Deep Learning for Medical Image Analysis**. Elsevier, 2017. p. 321–339. Disponível em: <<https://doi.org/10.1016/b978-0-12-810408-8.00019-5>>.

CBR. **Atlas BI-RADS do ACR: Sistema de laudos e registro de dados de imagem da mama**. 2. ed. São Paulo: Colégio Brasileiro de Radiologia, 2016. (Trad. ACR BI-RADS atlas: Breast Imaging Reporting and Data System (5th ed.), American College of Radiology).

CDC. **What Is a Mammogram?** Centers for Disease Control and Prevention, 2020. Disponível em: <https://www.cdc.gov/cancer/breast/basic_info/mammograms.htm>.

CHEDDAD, A. **The Complete Mini-DDSM**. 2021. <<https://www.kaggle.com/datasets/cheddad/miniddsm2>>. Acessado: 2021-10-12.

CHEN, H.; DOU, Q.; YU, L.; QIN, J.; HENG, P.-A. Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. **NeuroImage**, v. 170, p. 446–455, 2018. ISSN 1053-8119. Segmenting the Brain. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1053811917303348>>.

CHOLLET, F. *et al.* **Keras**. 2015. <<https://keras.io>>.

CHOUGRAD, H.; ZOUAKI, H.; ALHEYANE, O. Multi-label transfer learning for the early diagnosis of breast cancer. **Neurocomputing**, Elsevier BV, v. 392, p. 168–180, jun. 2020. Disponível em: <<https://doi.org/10.1016/j.neucom.2019.01.112>>.

CLARO, M.; VOGADO, L.; SANTOS, J.; VERAS, R. Utilização de técnicas de data augmentation em imagens: Teoria e prática. In: **Minicursos da ERCEMAPI 2020**. SBC, 2020. p. 47–71. Disponível em: <<https://doi.org/10.5753/sbc.11.5.3>>.

COGAN, T.; COGAN, M.; TAMIL, L. RAMS: Remote and automatic mammogram screening. **Computers in Biology and Medicine**, Elsevier BV, v. 107, p. 18–29, abr. 2019. Disponível em: <<https://doi.org/10.1016/j.compbiomed.2019.01.024>>.

DECHTER, R. Learning while searching in constraint-satisfaction-problems. In: . [S.l.: s.n.], 1986. p. 178–185.

Dicom Standard. **About DICOM: Overview**. 2022. <<https://www.dicomstandard.org/about-home>>. Acessado: 2022-05-02.

DOI, K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. **Computerized Medical Imaging and Graphics**, Elsevier BV, v. 31, n. 4-5, p. 198–211, jun. 2007. Disponível em: <<https://doi.org/10.1016/j.compmedimag.2007.02.002>>.

DOMINGUES, I.; ABREU, P. H.; SANTOS, J. Bi-rads classification of breast cancer: A new pre-processing pipeline for deep models training. In: **2018 25th IEEE International Conference on Image Processing (ICIP)**. [S.l.: s.n.], 2018. p. 1378–1382.

DSA. **Deep Learning Book**. Data Science Academy, 2021. Disponível em: <<https://www.deeplearningbook.com.br/>>.

EDH. **Breast Cancer Stages: What do They Mean?** Every Day Health, 2018. Disponível em: <<https://www.everydayhealth.com/breast-cancer/stages-what-they-mean/>>.

EKPO, E. U.; ALAKHRAS, M.; BRENNAN, P. Errors in mammography cannot be solved through technology alone. **Asian Pacific Journal of Cancer Prevention**, West Asia Organization for Cancer Prevention (WAOCP), v. 19, n. 2, fev. 2018. Disponível em: <<https://doi.org/10.22034/APJCP.2018.19.2.291>>.

ELLIS, H.; MAHADEVAN, V. Anatomy and physiology of the breast. **Surgery (Oxford)**, Elsevier BV, v. 31, n. 1, p. 11–14, jan. 2013. Disponível em: <<https://doi.org/10.1016/j.mpsur.2012.10.018>>.

FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P.; UTHURUSAMY, R. (Ed.). **Advances in Knowledge Discovery and Data Mining**. USA: American Association for Artificial Intelligence, 1996. ISBN 0262560976.

GIESS, C. S.; FROST, E. P.; BIRDWELL, R. L. Difficulties and errors in diagnosis of breast neoplasms. **Seminars in Ultrasound, CT and MRI**, Elsevier BV, v. 33, n. 4, p. 288–299, ago. 2012. Disponível em: <<https://doi.org/10.1053/j.sult.2012.01.007>>.

GOLD, R. H.; BASSETT, L. W.; WIDOFF, B. E. Highlights from the history of mammography. **RadioGraphics**, Radiological Society of North America (RSNA), v. 10, n. 6, p. 1111–1131, nov. 1990. Disponível em: <<https://doi.org/10.1148/radiographics.10.6.2259767>>.

GOLDEN, R. Artificial neural networks: Neurocomputation. In: **International Encyclopedia of the Social & Behavioral Sciences**. Elsevier, 2001. p. 806–811. Disponível em: <<https://doi.org/10.1016/b0-08-043076-7/00563-5>>.

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016.

HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. 3. ed. [S.l.]: Elsevier, 2011.

HAYWARD, J. H.; RAY, K. M.; WISNER, D. J.; KORNAK, J.; LIN, W.; JOE, B. N.; SICKLES, E. A. Improving screening mammography outcomes through comparison with multiple prior mammograms. **American Journal of Roentgenology**, American Roentgen Ray Society, v. 207, n. 4, p. 918–924, out. 2016. Disponível em: <<https://doi.org/10.2214/ajr.15.15917>>.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. **CoRR**, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>.

HEATH, M.; BOWYER, K.; KOPANS, D.; KEGELMEYER, P.; MOORE, R.; CHANG, K.; MUNISHKUMARAN, S. Current status of the digital database for screening mammography. In: **Computational Imaging and Vision**. Springer Netherlands, 1998. p. 457–460. Disponível em: <https://doi.org/10.1007/978-94-011-5318-8_75>.

HERNÁNDEZ-LÓPEZ, J.; GÓMEZ-FLORES, W. Predicting the bi-rads lexicon for mammographie masses using hybrid neural models. **2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)**, p. 1–6, 2020.

IARC. **Globocan 2020: Estimated number of new cases in 2020, worldwide, females, ages 50+**. International Agency for Research on Cancer, 2021. Disponível em: <<https://gco.iarc.fr/today/online-analysis-pie>>.

IMNRC. **Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis**. Institute of Medicine and National Research Council, National Academies Press, 2005. Disponível em: <<https://doi.org/10.17226/11016>>.

INCA. **Estimativa 2020: Incidência de câncer no Brasil**. Rio de Janeiro: Instituto Nacional do Câncer, 2019. ISBN 9788573183894.

INCA. **A situação do câncer de mama no Brasil: síntese de dados dos sistemas de informação**. [S.l.]: Instituto Nacional do Câncer, 2019. ISBN 9788573183771.

INCA. **Detecção Precoce do Câncer de Mama**. Rio de Janeiro: Divisão de Detecção Precoce e Apoio à Organização de Rede, Secretaria de Atenção Especializada à Saúde, 2020.

JALES, R. M. **Bases físicas da mamografia**. 2015. Disponível em: <<https://drpixel.fcm.unicamp.br/conteudo/bases-fisicas-da-mamografia>>.

JAMESON, J. L.; FAUCI, A. S.; KASPER, D. L.; HAUSER, S. L.; LONGO, D. L.; LOSCALZO, J. **Harrison's Principles of Internal Medicine**. 20. ed. [S.l.]: McGraw-Hill Education / Medical, 2020. ISBN 1259643999.

JATOI, I.; KAUFMANN, M.; PETIT, J.-Y. **ATLAS OF BREAST SURGERY**. 2006. ed. [S.l.]: Springer, 2005. ISBN 9783540243519.

KAUR, P.; SINGH, G.; KAUR, P. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification. **Informatics in Medicine Unlocked**, Elsevier BV, v. 16, p. 100151, 2019. Disponível em: <<https://doi.org/10.1016/j.imu.2019.01.001>>.

KÖRNER, M.; WEBER, C. H.; WIRTH, S.; PFEIFER, K.-J.; REISER, M. F.; TREITL, M. Advances in digital radiography: Physical principles and system overview. **RadioGraphics**, Radiological Society of North America (RSNA), v. 27, n. 3, p. 675–686, maio 2007. Disponível em: <<https://doi.org/10.1148/rg.273065075>>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. Imagenet classification with deep convolutional neural networks. **Neural Information Processing Systems**, v. 25, 01 2012.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1**. Red Hook, NY, USA: Curran Associates Inc., 2012. (NIPS'12), p. 1097–1105.

LBN Medical. **Analog and Digital Mammography Machines – Advantages and Disadvantages**. 2020. Disponível em: <<https://lbnmedical.com/digital-and-analog-mammography-machines-advantages-and-disadvantages/>>.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Springer Science and Business Media LLC, v. 521, n. 7553, p. 436–444, maio 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>.

LECUN, Y.; HAFFNER, P.; BOTTOU, L.; BENGIO, Y. Object recognition with gradient-based learning. In: _____. **Shape, Contour and Grouping in Computer Vision**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999. p. 319–345. ISBN 978-3-540-46805-9. Disponível em: <https://doi.org/10.1007/3-540-46805-6_19>.

LEE, R. S.; GIMENEZ, F.; HOOGI, A.; MIYAKE, K. K.; GOROVOY, M.; RUBIN, D. L. A curated mammography data set for use in computer-aided detection and diagnosis research. **Scientific Data**, Springer Science and Business Media LLC, v. 4, n. 1, dez. 2017. Disponível em: <<https://doi.org/10.1038/sdata.2017.177>>.

LIU, Y.; CHEN, W. A SAS macro for testing differences among three or more independent groups using kruskal-wallis and nemenyi tests. **Journal of Huazhong University of Science and Technology [Medical Sciences]**, Springer Science and Business Media LLC, v. 32, n. 1, p. 130–134, jan. 2012. Disponível em: <<https://doi.org/10.1007/s11596-012-0023-9>>.

LODISH, H.; BERK, A.; ZIPURSKY, S. L.; MATSUDAIRA, P.; BALTIMORE, D.; DARNELL, J. Overview of neuron structure and function. In: **Molecular Cell Biology**. fourth. New York: W. H. Freeman: [s.n.], 2000. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK21535/>>.

LOPEZ, M. A. G.; POSADA, N.; MOURA, D.; POLLÁN, R.; FRANCO-VALIENTE, J.; ORTEGA, C.; SOLAR, M. D.; DÍAZ-HERRERO, G.; RAMOS, I.; LOUREIRO, J.; FERNANDES, T.; ARAÚJO, B. Bcdr: A breast cancer digital repository. In: . [S.l.: s.n.], 2012. p. 1065–1066.

MADHAVAN, S.; JONES, M. T. **Deep learning architectures**. IBM, 2021. Disponível em: <<https://developer.ibm.com/technologies/artificial-intelligence/articles/cc-machine-learning-deep-learning-architectures>>.

MOORE-HIGGS, G. J.; LIGIBEL, J. A. Psychosocial consequences and lifestyle interventions. In: **The Breast**. Elsevier, 2009. p. 1443–1454. Disponível em: <<https://doi.org/10.1016/b978-1-4160-5221-0.00091-7>>.

MOREIRA, I. C.; AMARAL, I.; DOMINGUES, I.; CARDOSO, A.; CARDOSO, M. J.; CARDOSO, J. S. Inbreast: Toward a full-field digital mammographic database. **Academic Radiology**, v. 19, n. 2, p. 236–248, 2012. ISSN 1076-6332. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S107663321100451X>>.

NASCIMENTO, R. G. do; OTONI, K. M. Histological and molecular classification of breast cancer: what do we know? **Mastology**, Mastology, v. 30, 2020. Disponível em: <<https://doi.org/10.29289/25945394202020200024>>.

NAZÁRIO, A. C. P.; FACINA, G.; FILASSI, J. R. Breast cancer: news in diagnosis and treatment. **Revista da Associação Médica Brasileira**, FapUNIFESP (SciELO), v. 61, n. 6, p. 543–552, dez. 2015. Disponível em: <<https://doi.org/10.1590/1806-9282.61.06.543>>.

PALUCH-SHIMON, S.; CARDOSO, F.; SESSA, C.; BALMANA, J.; CARDOSO, M.; GILBERT, F.; SENKUS, E. Prevention and screening in BRCA mutation carriers and other breast/ovarian hereditary cancer syndromes: ESMO clinical practice guidelines for cancer prevention and screening. **Annals of Oncology**, Elsevier BV, v. 27, p. v103–v110, set. 2016. Disponível em: <<https://doi.org/10.1093/annonc/mdw327>>.

PASQUALINI, R. **Breast cancer: prognosis, treatment, and prevention**. [S.l.]: Springer International Publishing, 2008. ISBN 9781420058727.

PATTERSON, J.; GIBSON, A. **Deep Learning: A Practitioner's Approach, Sebastopol**. 1. ed. [S.l.]: O'Reilly Media, 2017.

Pytorch Community. **END-TO-END MACHINE LEARNING FRAMEWORK**. 2022. <<https://pytorch.org/features/>>. Acessado: 2022-01-12.

RADASS-UZ. **BIRADS: Mammographic Mass Lexicon**. Radiology UZ GENT, 2015. Disponível em: <<https://radassuz.wixsite.com/radass/birads>>.

SCIKIT-LEARN. **One Vs Rest Classifier**. 2022. <<https://scikit-learn.org/stable/modules/generated/sklearn.multiclass.OneVsRestClassifier.html>>. Acessado: 2022-05-15.

SENKUS-KONEFKA, E.; CARDOSO, F.; DOUILLARD, J.-Y.; BRAMLEY, C.; LONGO, F.; JEZDIC, S. **Patient Guide Series**. [S.l.]: European Society for Medical Oncology, 2018.

SHAO, K.; TANG, Z.; ZHU, Y.; LI, N.; ZHAO, D. A survey of deep reinforcement learning in video games. 2019.

SHEN, L.; MARGOLIES, L. R.; ROTHSTEIN, J. H.; FLUDER, E.; MCBRIDE, R.; SIEH, W. Deep learning to improve breast cancer detection on screening mammography. **Scientific Reports**, Springer Science and Business Media LLC, v. 9, n. 1, ago. 2019. Disponível em: <<https://doi.org/10.1038/s41598-019-48995-4>>.

SIDDEEQ, S.; LI, J.; BHATTI, H. M. A.; MANZOOR, A.; MALHI, U. S. Deep learning rn-benn model for breast cancer bi-rads classification. In: **2021 The 4th International Conference on Image and Graphics Processing**. New York, NY, USA: Association for Computing Machinery, 2021. (ICIGP 2021), p. 219–225. ISBN 9781450389105. Disponível em: <<https://doi.org/10.1145/3447587.3447620>>.

SILVA, V. J. S. **MAMOCAD-BR: Mineração de Imagens para Classificação Inteligente de Tumores de Mama Usando o Protocolo BI-RADS**. Dissertação (Mestrado em Ciência da Computação) — Universidade do Estado do Rio Grande do Norte / Universidade Federal Rural do Semi-Árido, 2020.

SIM. **Mortalidade - Neoplasias malignas da mama**. Sistema de Informações sobre Mortalidade, 2019. Disponível em: <<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/obt10uf.def>>.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv 1409.1556**, 09 2014.

SOUTO, L. P. M. **Mineração de Imagens para a Classificação de Tumores de Mama**. Dissertação (Mestrado em Ciência da Computação) — Universidade do Estado do Rio Grande do Norte / Universidade Federal Rural do Semi-Árido, 2014.

SPAK, D.; PLAXCO, J.; SANTIAGO, L.; DRYDEN, M.; DOGAN, B. BI-RADS ® fifth edition: A summary of changes. **Diagnostic and Interventional Imaging**, Elsevier BV, v. 98, n. 3, p. 179–190, mar. 2017. Disponível em: <<https://doi.org/10.1016/j.diii.2017.01.001>>.

SUJATHA, K.; DURGADEVI, G.; KUMAR, K. S.; KARTHIKEYAN, V.; PONMAGAL, R.; HARI, R.; BHAVANI, N.; SRIVIDHYA, V.; CAO, S.-Q. Screening and early identification of microcalcifications in breast using texture-based ANFIS classification. In: **Wearable and Implantable Medical Devices**. Elsevier, 2020. p. 115–140. Disponível em: <<https://doi.org/10.1016/b978-0-12-815369-7.00005-7>>.

SUNG, H.; FERLAY, J.; SIEGEL, R. L.; LAVERSANNE, M.; SOERJOMATARAM, I.; JEMAL, A.; BRAY, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinicians**, Wiley, v. 71, n. 3, p. 209–249, fev. 2021. Disponível em: <<https://doi.org/10.3322/caac.21660>>.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. In: **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. p. 1–9.

TANG, R. ying; GAO, W.; MA, L. hua; LIN, B. yao; XU, G. wei. Bi-rads categorization and positive predictive value of mammographic features. **Chinese Journal of Cancer Research**, Chinese Journal of Cancer Research, v. 13, n. 3, p. 202–205, set. 2001. Disponível em: <<https://doi.org/10.1007/bf02983885>>.

TDS. **Demystifying Convolutional Neural Networks**. Towards Data Science, 2021. Disponível em: <<https://towardsdatascience.com/demystifying-convolutional-neural-networks-384785791596>>.

TDS. **Segmenting Abnormalities in Mammograms (Part 1 of 3)**. Towards Data Science, 2021. Disponível em: <<https://towardsdatascience.com/can-you-find-the-breast-tumours-part-1-of-3-1473ba685036>>.

WANG, L.; HE, Q.; WANG, X.; SONG, T.; LI, X.; ZHANG, S.; QIN, G.; CHEN, W.; ZHOU, L.; ZHEN, X. Multi-criterion decision making-based multi-channel hierarchical fusion of digital breast tomosynthesis and digital mammography for breast mass discrimination. **Knowledge-Based Systems**, Elsevier BV, v. 228, p. 107303, set. 2021. Disponível em: <<https://doi.org/10.1016/j.knosys.2021.107303>>.

WHO. **Breast Cancer**. World Health Organization, 2021. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/breast-cancer>>.

WINKEL, R. R.; EULER-CHELPIN, M. von; NIELSEN, M.; DIAO, P.; NIELSEN, M. B.; ULDALL, W. Y.; VEJBORG, I. Inter-observer agreement according to three methods of evaluating mammographic density and parenchymal pattern in a case control study: impact on relative risk of breast cancer. **BMC Cancer**, Springer Science and Business Media LLC, v. 15, n. 1, abr. 2015. Disponível em: <<https://doi.org/10.1186/s12885-015-1256-3>>.

XIN, Y.; KONG, L.; LIU, Z.; CHEN, Y.; LI, Y.; ZHU, H.; GAO, M.; HOU, H.; WANG, C. Machine learning and deep learning methods for cybersecurity. **IEEE Access**, Institute of Electrical and Electronics Engineers (IEEE), v. 6, p. 35365–35381, 2018. Disponível em: <<https://doi.org/10.1109/access.2018.2836950>>.

YALA, A.; MIKHAEL, P. G.; STRAND, F.; LIN, G.; SMITH, K.; WAN, Y.-L.; LAMB, L.; HUGHES, K.; LEHMAN, C.; BARZILAY, R. Toward robust mammography-based

models for breast cancer risk. **Science Translational Medicine**, American Association for the Advancement of Science, v. 13, n. 578, 2021. ISSN 1946-6234. Disponível em: <<https://stm.sciencemag.org/content/13/578/eaba4373>>.

YOUNG, T.; HAZARIKA, D.; PORIA, S.; CAMBRIA, E. Recent trends in deep learning based natural language processing [review article]. **IEEE Computational Intelligence Magazine**, v. 13, p. 55–75, 08 2018.

ZAIONTZ, C. **Friedman Test | Real Statistics Using Excel**. 2022. <<https://www.real-statistics.com/anova-repeated-measures/friedman-test/>>. Acessado: 2022-04-18.

ZHANG, H.; ZHANG, L.; JIANG, Y. Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In: **2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)**. IEEE, 2019. Disponível em: <<https://doi.org/10.1109/wcsp.2019.8927876>>.

APÊNDICE A – RESULTADOS DOS TESTES COMPLEMENTARES

Tabela 16 – AlexNet - Teste Complementar 1

AlexNet	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,200	0,200	0,200	0,190	30 épocas (05h10m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 17 – AlexNet - Teste Complementar 2

AlexNet	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,230	0,240	0,230	0,220	50 épocas (08h14m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 18 – ResNet-18 - Teste Complementar 1

ResNet-18	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,480	0,500	0,480	0,480	100 épocas (43h54m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 19 – ResNet-18 - Teste Complementar 2

ResNet-18	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,480	0,480	0,490	0,474	50 épocas (22h17m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 20 – ResNet-18 - Teste Complementar 3

ResNet-18	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,470	0,450	0,470	0,450	50 épocas (24h02m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 21 – ResNet-18 - Teste Complementar 4

ResNet-18	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,480	0,490	0,480	0,470	20 épocas (08h35m) lr = 0,0001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 22 – ResNet-18 - Teste Complementar 5

ResNet-18	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,460	0,480	0,460	0,450	30 épocas (13h) lr = 0,0001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 23 – ResNet-34 - Teste Complementar 1

ResNet-34	Multiclasse			
Acurácia	Precisão	Recall	F1	Observação
0,500	0,520	0,500	0,500	30 épocas (24h) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Com aumento de dados

Tabela 24 – AlexNet - Teste Complementar 1 (DCTr)

AlexNet	DCTr (B1_NB1)			
Acurácia	Precisão	Recall	F1	Observação
0,490	0,490	0,490	0,460	50 épocas (30m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 25 – AlexNet - Teste Complementar 2 (DCTr)

AlexNet	DCTr (B2_NB2)			
Acurácia	Precisão	Recall	F1	Observação
0,510	0,510	0,510	0,510	50 épocas (19m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 26 – AlexNet - Teste Complementar 3 (DCTr)

AlexNet	DCTr (B3_NB3)			
Acurácia	Precisão	Recall	F1	Observação
0,490	0,490	0,490	0,490	50 épocas (28m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 27 – AlexNet - Teste Complementar 4 (DCTr)

AlexNet	DCTr (B4_NB4)			
Acurácia	Precisão	Recall	F1	Observação
0,540	0,540	0,540	0,530	50 épocas (29m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 28 – AlexNet - Teste Complementar 5 (DCTr)

AlexNet	DCTr (B5_NB5)			
Acurácia	Precisão	Recall	F1	Observação
0,430	0,410	0,430	0,400	50 épocas (29m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 29 – ResNet-18 - Teste Complementar 1 (DCTr)

ResNet-18	DCTr (B1_NB1)			
Acurácia	Precisão	Recall	F1	Observação
0,920	0,920	0,920	0,920	50 épocas (01h29m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 30 – ResNet-18 - Teste Complementar 2 (DCTr)

ResNet-18	DCTr (B2_NB2)			
Acurácia	Precisão	Recall	F1	Observação
0,630	0,660	0,630	0,610	50 épocas (01h33m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 31 – ResNet-18 - Teste Complementar 3 (DCTr)

ResNet-18	DCTr (B3_NB3)			
Acurácia	Precisão	Recall	F1	Observação
0,540	0,550	0,540	0,540	50 épocas (01h35m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 32 – ResNet-18 - Teste Complementar 4 (DCTr)

ResNet-18	DCTr (B4_NB4)			
Acurácia	Precisão	Recall	F1	Observação
0,620	0,640	0,620	0,600	50 épocas (01h31m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 33 – ResNet-18 - Teste Complementar 5 (DCTr)

ResNet-18	DCTr (B5_NB5)			
Acurácia	Precisão	Recall	F1	Observação
0,540	0,540	0,540	0,540	50 épocas (01h30m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 34 – ResNet-18 - Teste Complementar 6 (DCTr)

ResNet-18	DCTr (B1_NB1)			
Acurácia	Precisão	Recall	F1	Observação
0,970	0,970	0,970	0,970	30 épocas (01h39m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 35 – ResNet-18 - Teste Complementar 7 (DCTr)

ResNet-18	DCTr (B2_NB2)			
Acurácia	Precisão	Recall	F1	Observação
0,830	0,830	0,830	0,830	30 épocas (51m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 36 – ResNet-18 - Teste Complementar 8 (DCTr)

ResNet-18	DCTr (B3_NB3)			
Acurácia	Precisão	Recall	F1	Observação
0,590	0,590	0,590	0,590	30 épocas (53m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 37 – ResNet-18 - Teste Complementar 9 (DCTr)

ResNet-18	DCTr (B4_NB4)			
Acurácia	Precisão	Recall	F1	Observação
0,570	0,590	0,570	0,540	30 épocas (55m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 38 – ResNet-18 - Teste Complementar 10 (DCTr)

ResNet-18	DCTr (B5_NB5)			
Acurácia	Precisão	Recall	F1	Observação
0,660	0,660	0,660	0,660	30 épocas (52m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 39 – ResNet-34 - Teste Complementar 1 (DCTr)

ResNet-34	DCTr (B1_NB1)			
Acurácia	Precisão	Recall	F1	Observação
0,910	0,920	0,910	0,910	50 épocas (02h54m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 40 – ResNet-34 - Teste Complementar 2 (DCTr)

ResNet-34	DCTr (B2_NB2)			
Acurácia	Precisão	Recall	F1	Observação
0,630	0,640	0,630	0,620	50 épocas (03h) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 41 – ResNet-34 - Teste Complementar 2.2 (DCTr)

ResNet-34	DCTr (B2_NB2)			
Acurácia	Precisão	Recall	F1	Observação
0,620	0,650	0,620	0,590	100 épocas (05h34m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 42 – ResNet-34 - Teste Complementar 3 (DCTr)

ResNet-34	DCTr (B3_NB3)			
Acurácia	Precisão	Recall	F1	Observação
0,580	0,590	0,580	0,580	50 épocas (03h03m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 43 – ResNet-34 - Teste Complementar 4 (DCTr)

ResNet-34	DCTr (B4_NB4)			
Acurácia	Precisão	Recall	F1	Observação
0,660	0,660	0,660	0,660	50 épocas (02h40m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 44 – ResNet-34 - Teste Complementar 5 (DCTr)

ResNet-34	DCTr (B5_NB5)			
Acurácia	Precisão	Recall	F1	Observação
0,590	0,600	0,590	0,580	50 épocas (02h45m) lr = 0,001 IMG FULL (CBIS-DDSM e INbreast) Sem aumento de dados

Tabela 45 – ResNet-34 - Teste Complementar 6 (DCTr)

ResNet-34	DCTr (B1_NB1)			
Acurácia	Precisão	Recall	F1	Observação
0,940	0,940	0,940	0,940	30 épocas (01h36m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 46 – ResNet-34 - Teste Complementar 7 (DCTr)

ResNet-34	DCTr (B2_NB2)			
Acurácia	Precisão	Recall	F1	Observação
0,770	0,780	0,770	0,760	30 épocas (01h24m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 47 – ResNet-34 - Teste Complementar 8 (DCTr)

ResNet-34	DCTr (B3_NB3)			
Acurácia	Precisão	Recall	F1	Observação
0,610	0,620	0,610	0,610	30 épocas (01h35m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 48 – ResNet-34 - Teste Complementar 9 (DCTr)

ResNet-34	DCTr (B4_NB4)			
Acurácia	Precisão	Recall	F1	Observação
0,580	0,610	0,580	0,540	30 épocas (01h39m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados

Tabela 49 – ResNet-34 - Teste Complementar 10 (DCTr)

ResNet-34	DCTr (B5_NB45)			
Acurácia	Precisão	Recall	F1	Observação
0,700	0,730	0,700	0,690	30 épocas (01h38m) lr = 0,001 IMG ROI (CBIS-DDSM) Sem aumento de dados