
**DETECÇÃO DO CÂNCER DE MAMA UTILIZANDO DIAGNÓSTICO AUXILIADO
POR COMPUTADOR: UMA REVISÃO DE MODELOS DE DEEP LEARNING PARA
A ANÁLISE DE MAMOGRAFIAS**

**BREAST CANCER DETECTION USING COMPUTER AIDED DIAGNOSIS:
A REVIEW OF DEEP LEARNING MODELS FOR MAMMOGRAM ANALYSIS**

Erika Barros de Alcântara¹

Ricardo Petri Silva²

RESUMO

O câncer de mama é um dos diagnósticos mais comuns de câncer e também uma das maiores causas de morte entre mulheres pelo mundo. Por essa razão, a detecção precoce é primordial para o aumento das chances de sobrevivência. Nos últimos anos, o Diagnóstico Auxiliado por Computador (Computer aided Diagnosis - CAD) utilizando a Inteligência Artificial e técnicas de Machine Learning tem se mostrado ferramenta importante na tomada de decisão de especialistas através da análise, classificação e detecção de padrões em exames de imagem, como a mamografia. Com o emprego de técnicas de Deep Learning, um ramo de Machine Learning e Redes Neurais Artificiais mais Profundas, os sistemas CAD demonstraram performance ainda mais elevada. Este estudo apresenta uma revisão de modelos recentes de CAD que utilizaram as técnicas e arquiteturas mais empregadas atualmente para o campo de análise de imagens médicas com o fim de detectar e diagnosticar o câncer de mama, fazendo um comparativo entre eles e mostrando como cada um pode contribuir para esse propósito. Dentre os achados deste trabalho, destacaram-se os modelos Você Só Olha Uma Vez (You Only Look Once - YOLO) combinados com Redes Neurais Convolucionais, pela sua viabilidade de aplicação clínica. Por fim, são extraídas conclusões e suscitados alguns pontos interessantes para pesquisas futuras.

263

Palavras-chave: câncer de mama; diagnóstico auxiliado por computador; inteligência artificial; deep learning; redes neurais convolucionais.

ABSTRACT

Breast cancer is one of the most common diagnoses of cancer and also one of the leading causes of death among women worldwide. For this reason, early detection is crucial to increase survival rates. In recent years, Computer-aided Diagnosis (CAD) assisted by Artificial Intelligence and Machine Learning techniques has proven to be an important tool in decision making by experts through analysis, classification and detection of patterns in imaging exams, such as mammography. With the use of Deep

¹ Centro Universitário Filadélfia de Londrina – UniFil.

² Centro Universitário Filadélfia de Londrina - UniFil

Learning techniques, a branch of Machine Learning, and Deeper Artificial Neural Networks, CAD systems showed even higher performance. This study presents a review of recent CAD models that used the techniques and architectures most currently used in the field of medical image analysis in order to detect and diagnose breast cancer, making a comparison between them and showing how each one can contribute to this purpose. Among the findings of this work, the You Only Look Once (YOLO) models combined with Convolutional Neural Networks stood out, due to their feasibility of clinical application. Finally, conclusions are drawn and some interesting points for future research are raised.

Keywords: breast cancer; computer-aided diagnosis; artificial intelligence; deep learning; convolutional neural network.

1 INTRODUÇÃO

O câncer de mama é o tipo de câncer com maior incidência e também uma das maiores causas de morte entre as mulheres. Segundo dados do Global Cancer Observatory - GCO, estima-se que até 2040 o número de casos irá aumentar em 80.2%, considerando mulheres de 0 a 85 anos, de todos os continentes (International Agency for Research of Cancer, 2023). A detecção precoce da doença é de fundamental importância para aumentar a taxa de sobrevivência das mulheres acometidas e os exames de rastreios anuais são considerados uma das melhores estratégias para tanto (Hassan; Hamad; Mahar, 2022). Dentre estes exames, estão os de imagem, como a mamografia, a ultrassonografia, a tomossíntese digital de mama e a ressonância magnética. A mamografia é considerada a modalidade de rastreio mais eficaz e sensível para detectar a doença em seus primeiros estágios (Hassan; Hamad; Mahar, 2022).

No entanto, a interpretação dessas imagens é considerada um verdadeiro desafio para médicos radiologistas, exigindo muito tempo e esforço, uma vez que as anormalidades são diversas e muitas vezes se confundem com alguns tecidos da mama ou são escondidas por tecidos sobrepostos. Além disso, mesmo com muito treinamento e expertise dos médicos radiologistas, a análise pode levar a resultados de falsos positivos, ou falsos negativos (Hamed *et al.*, 2020). Para enfrentar esses desafios e assistir os radiologistas no processo de interpretação de exames de imagens, sistemas de Diagnóstico Auxiliado por Computador (Computer Aided Diagnosis - CAD) com Inteligência Artificial têm sido desenvolvidos e empregados no

uso clínico ao longo dos anos, contribuindo para o tempo e precisão dos diagnósticos (Alju Aid *et al.*, 2022; Balkenende; Teuwen; Mann, 2022; Chugh; Kumar; Singh, 2021; Dogra; Goyal; Kaushik, 2019; Hamed *et al.*, 2020; Kim *et al.*, 2020; Shen *et al.*, 2019).

Tradicionalmente, esses sistemas utilizavam algoritmos de Aprendizado de Máquina (Machine Learning) para interpretação de imagens e dependiam da extração manual de características estruturais pelos especialistas. Mas, devido à grande variação dos dados entre pacientes, esses métodos tradicionais não se mostraram tão confiáveis (Razzak; Naz; Zaib, 2018), nem demonstraram melhorar a precisão dos diagnósticos (Balkenende; Teuwen; Mann, 2022). Dessa forma, tiveram progresso estagnado por mais de uma década desde que foram apresentados (Shen *et al.*, 2019).

Com o avanço do Aprendizado Profundo (Deep Learning), um ramo de Machine Learning, os CADs ganharam uma grande vantagem ao aplicar suas técnicas e arquiteturas, a de gerar características representativas de alto nível diretamente das imagens brutas (Dogra; Goyal; Kaushik, 2019), elevando suas performances e melhorando a qualidade dos diagnósticos (Balkenende; Teuwen; Mann, 2022). Embora os modelos de Deep Learning funcionem melhor com grandes conjuntos de dados e tenham grande custo computacional, esses desafios vêm sendo contornados com a introdução de conceitos como a Transferência de Aprendizado (Transfer Learning - TL), por exemplo, melhorando seus funcionamentos (Khan *et al.*, 2019).

Este artigo tem como objetivo fazer uma revisão sobre as técnicas de Deep Learning e arquiteturas mais utilizadas atualmente nos sistemas CAD para o diagnóstico do câncer de mama, que são os modelos baseados nas Redes Neurais Convolucionais (Convolutional Neural Network - CNN) (Fathy; Ghoneim, 2019), apresentando suas características, resultados e um comparativo entre eles, a fim de encontrar técnicas promissoras e viáveis para a aplicação clínica, contribuindo para a evolução no diagnóstico do câncer de mama e para a pesquisa científica nesse ramo. Os achados desse trabalho apontaram um destaque para os modelos YOLO combinado com CNNs, como modelos que possuem viabilidade de aplicação em cenários reais, por sua performance rápida e precisa, com capacidade de detecção de massas mamárias até mesmo em situações mais complexas como tecidos densos ou sobre o músculo peitoral. Porém existem ressalvas, uma vez que os resultados

desses sistemas são influenciados por inúmeros fatores e desafios como a escassez de dados ainda persistem para os modelos de Deep Learning.

Este trabalho está estruturado da seguinte forma: a segunda seção apresenta cinco modelos recentes desenvolvidos para o fim de detecção e classificação de lesões mamárias em imagens de mamografia, demonstrando os resultados obtidos. A terceira seção expõe conceitos importantes sobre tema e explica as métricas mais utilizadas para medir o desempenho dos sistemas CAD. A quarta seção demonstra a metodologia utilizada. Por fim, a quinta e sexta seções trazem, respectivamente, uma discussão sobre a análise feita nos estudos e a conclusão de como esses modelos se comportam no momento atual, mostrando as vantagens e os desafios que persistem no campo de análise de imagens médicas, além de entendimentos importantes que podem ser aprofundados em pesquisas futuras.

2 TRABALHOS RELACIONADOS

Muitos estudos e modelos já foram desenvolvidos baseados em técnicas e arquiteturas de Deep Learning, sendo as baseadas nas CNNs as mais utilizadas no momento (Boumaraf *et al.*, 2021) para a análise de imagens. Uma das grandes contribuições desses modelos foi a possibilidade extrair características relevantes automaticamente a partir de imagens de entrada.

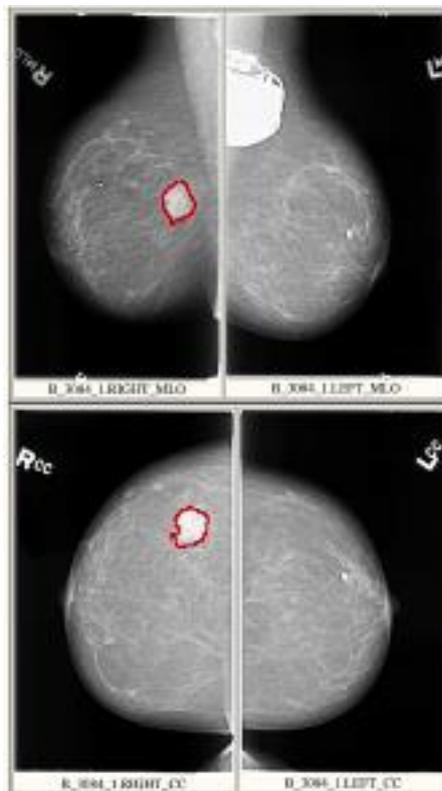
Para ilustrar e melhor compreender tais afirmações, foram reunidos cinco sistemas recentes que utilizaram modelos de Deep Learning baseados na arquitetura CNN para a análise de imagens de mamografia. Todos aplicaram o Banco de Dados Digital para Exames de Mamografia de Rastreamento (Digital Database for Screening Mammography - DDSM), sozinho ou, na maioria dos casos, em conjunto com o seu aumentado.

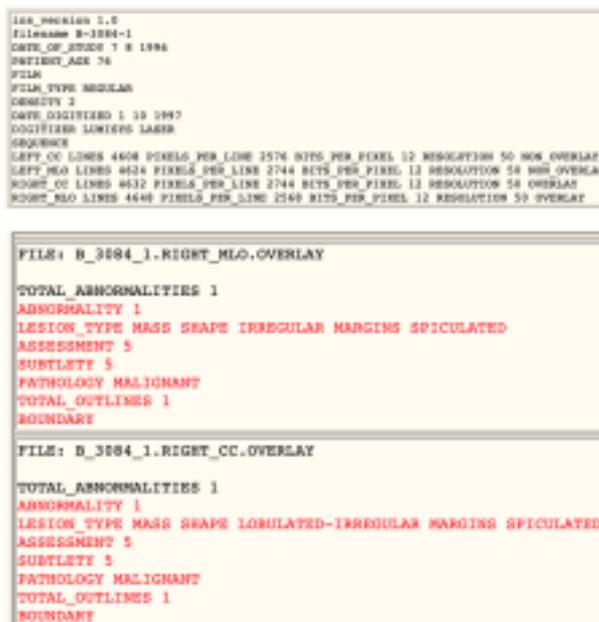
O DDSM é um projeto colaborativo de diversas instituições, mantido pela Universidade do Sul da Flórida e disponível publicamente para a comunidade de pesquisa na área de análise de imagens para desenvolvimento de algoritmos e sistemas de detecção e classificação do câncer de mama. O banco de dados é composto de 2620 imagens digitalizadas de mamografias realizadas por diferentes escâneres, resoluções e organizadas em 43 volumes. Os volumes são divididos de

acordo com a severidade dos casos e mantém um equilíbrio entre casos normais, benignos e câncer. Cada volume contém quatro imagens com incidências médio lateral oblíqua (MLO) e craniocaudal (CC), associadas a arquivos com anotações de lesões e outras informações dos pacientes (USF, 2023; Heath *et al.*, 2001), conforme a Figura 1.

Um dos modelos a utilizar esse banco de dados foi o proposto por Al-Masni *et al.* (2018). Desenvolveram um CAD com a técnica YOLO, um modelo de CNN baseada na Região de Interesse (Region of Interest - ROI) que realiza a detecção e classificação simultaneamente na imagem inteira. O estudo foi concentrado em imagens de mamografia e o treinamento do sistema foi realizado com o DDSM e o seu conjunto aumentado, sendo metade dos casos malignos e metade benignos. O aumento do conjunto de dados foi feito selecionando casos originais e utilizando diferentes técnicas de transformação de imagens, como rotação em diferentes ângulos.

Figura 1 – Volume: cancer_01 Case: B30841





Fonte: Digital Database for Screening Mammography (2023).

Após, o estudo consistiu em 4 etapas, o pré-processamento, a extração de características utilizando Redes Convolucionais Profundas (Deep Convolutional Networks - DCNN), a detecção de massas e classificação, empregando Redes Neurais Totalmente Conectadas (Fully Connected Neural Networks - FCNNs). Esse sistema foi capaz de detectar as massas e classificá-las em benignas ou malignas. Sua performance foi testada em uma validação cruzada de cinco partes ($k = 5$), obtendo acurácia geral de 99,7% para a detecção e 97% para a classificação, conforme a Tabela 1. Além disso, conseguiu detectar e classificar massas existentes sobre o músculo peitoral ou regiões mais densas. A detecção para cada mamograma, levou menos de 3s (Al-Masni *et al.*, 2018).

268

Em mais uma pesquisa envolvendo detecção e classificação simultânea com o intuito de melhorar a performance diagnóstica de lesões mamárias, Al-Antari *et al.* (2020) desenvolveram um modelo CAD integrado, de ponta-a ponta, utilizando YOLO (YOLO 9000) para detecção em mamografias e três classificadores de Deep Learning baseados em CNN para a classificação: uma CNN direcionada regular, ResNet 50 e InceptionResNet-V2 modificados. O sistema foi avaliado em testes de validação cruzada de 5 vezes usando dois conjuntos de dados diferentes DDSM e INBreast, separando pontuações de desempenho de cada. Os resultados da avaliação da detecção de lesões mamárias mostram que o detector YOLO é capaz de alcançar

precisões gerais de detecção de 99,17% e 97,27%, com pontuação F1 de 99,28% e 98,02% para os conjuntos de dados DDSM e INBreast, respectivamente. Usando as lesões mamárias detectadas, os modelos de classificação da CNN direcionada, ResNet-50 e InceptionResNet-V2 alcançaram médias de precisão geral promissoras de 94,50%, 95,83% e 97,50%, respectivamente, para o conjunto de dados DDSM, e 88,74%, 92,55% e 95,32%, respectivamente, para o conjunto de dados INBreast, de acordo com a Tabela 1. O estudo testou os três modelos de classificadores baseados em CNN em conjunto com o detector YOLO, concluindo que a melhor combinação foi YOLO + InceptionResNet- V2 as quais obtiveram melhores resultados e melhor tempo, detectando e classificando lesões mamárias em uma única mamografia em menos de 0.025s. Durante o teste em ambos os conjuntos de dados DDSM e INBreast, tal sistema pode prever 71 quadros por segundo.

Já Fathy *et al.* (2019) desenvolveram uma abordagem de Deep Learning diferente para classificação e localização do câncer de mama. O estudo utilizou um subconjunto de dados do banco DDSM e o seu aumentado para treinamento e teste e consistiu em dois passos, o primeiro foi o emprego de uma arquitetura ResNet-50 pré-treinada para a extração de características relevantes das imagens de mamografia e classificação entre classe normal e classe massa. O segundo passo foi usar a técnica de Mapa de Ativação de Classe (Class Map Activation - CAM) para mapear as regiões de interesse na mamografia, com base nas características extraídas, possibilitando a localização de anormalidades na imagem. O CAM explora classificadores de uma CNN com a camada de Agrupamento da Média Global (Global Average Pooling - GAP) para a localização de objetos sem supervisão. Ele é gerado ponderando a soma dos mapas de ativação na última camada de convolução da ResNet-50 com os pesos associados à classe de interesse (como a classe "massa"). Os resultados experimentais desse CAD foi de 96% da Curva da Área Sob o Receptor de Características Operacionais (Area Under The Receiver Operating Characteristics - ROC), com 99.8% de sensibilidade e 82.1% de especificidade para a classificação e 93.67% para a localização de massas, com uma média de 12.2% de falsos positivos por imagem, conforme Tabela 1.

Rahman *et al.* (2020) propuseram uma abordagem de Deep Learning para classificação de tumores de mama a partir de mamografias pré-segmentadas,

aplicando várias técnicas de pré-processamento de imagens e aumento de dados, bem como modelos de arquiteturas CNN pré-treinadas e modificadas, a InceptionV3 e Res Net50. Adotou o conjunto de dados DDSM, mas apenas com mamografias do modelo de escâner LUMISYS que fornece informações sobre massas tumorais por meio de máscaras. Foram coletadas um total de 1167 imagens de diferentes pacientes, as quais foram divididas aleatoriamente em três conjuntos: 80% para treinamento, 10% para teste e 10% para validação. O conjunto de treinamento tinha um equilíbrio de classes, com 491 imagens de câncer e 465 imagens de casos benignos. Os conjuntos de validação e teste continham 106 e 105 imagens, respectivamente, com aproximadamente o mesmo número de casos de câncer e benignos. Para Inception V3, utilizaram um exemplo de treinamento de tamanho 64 e 20 épocas. Já para ResNet50 o tamanho do exemplo de treinamento foi 50 e de épocas, 10. Obtiveram resultados de 79.6% de acurácia, 75.4% de precisão, 89.1% de sensibilidade e 86% de especificidade para o primeiro modelo, bem como 85.71% de acurácia, 85.7% de precisão, 87.3% de sensibilidade e 84% de especificidade para o segundo, consoante a Tabela 1. Concluíram que menor número de épocas produz melhores resultados e que o número utilizado nos modelos foi suficiente. Também descobriram que a inserção de 2 de camadas totalmente conectadas foi o número ideal para ambos os modelos, levando aos melhores resultados.

Salama *et al.* (2021) desenvolveram um *framework* para classificação de imagens entre benigno e maligno aplicando vários modelos ponta-a-ponta de CNN pré treinados, como InceptionV3, DenseNet121, ResNet50, VGG16 e MobileNetV2, bem como um *framework* modificado de U-Net para segmentar a área da mama a partir da imagem de mamografia. Utilizaram três conjuntos de dados diferentes e separadamente, o DDSM, o MIAS e o CBIS-DDSM. Aplicaram também o aumento de dados com rotação de imagens em 4 ângulos diferentes. Os melhores resultados foram atingidos utilizando os recortes de imagem médio lateral oblíquo (MLO) do conjunto DDSM e a técnica de aumento de dados com a junção dos modelos da U-Net modificada e InceptionV3, chegando aos 98.87% de acurácia, 98.88% de AUC, 98.98% de sensibilidade, 98.79% de precisão e 97.99% de pontuação F1, no tempo computacional de 1.2134s. Tais métricas melhoraram quando os autores testaram os mesmos modelos com uma combinação dos recortes de imagem MLO e craniocaudal

(CC), registrando 99.43% de acurácia, 99.22% de AUC, 99.12% de sensibilidade, 98.99% de precisão e 98.98% de pontuação F1, conforme a Tabela 1.

Tabela 1 - Modelos de Deep Learning baseados em CNNs para detecção e classificação do câncer de mama em mamografias

Artigo	Estrutura da CNN	Conjunto de Treinamento	Conjunto de Teste	TL	PP	Ac	Pr	Sen	Es	F1	AUC
Al-Masni <i>et al.</i> , 2018	YOLO (Detecção) DCNN (Extração de características) FCNN (Detecção e classificação)	DDSM	DDSM Com k-fold cross validation (k=5)	Sim	Sim	99.7% (detecção) 97% (classificação)		100%	94%		96.45%
Al-Antari <i>et al.</i> , 2020	YOLO (Detecção) CNN direcionada regular (Classificação) ResNet-50 modificada (Classificação) InceptionResNet-V2 modificada (Classificação)	DDSM + INBreast, Separadamente	DDSM + INBreast, Separadamente	Sim	Sim	99.17% (detecção) 94.50%, 95.83% e 97.50%, respectivamente (classificação)		ResNet50 - 96.33% (classificação) InceptionResNet-V2 - 97.67% (classificação)		99.28% (detecção)	95%, 96/67% e 97.50%, respectivamente (classificação)
Fathy <i>et al.</i> , 2019	ResNet-50 (Extração de características e classificação) + CAM (localização)	DDSM + Aumentado	DDSM + Aumentado	Sim	Sim			99.8% (classificação) 93.67% (localização)	82.1% (classificação) 93.67% (localização de massas)		96%
Rahman <i>et al.</i> , 2020	InceptionV3 e ResNet50	DDSM (LUMISYS)	DDSM (LUMISYS)	Sim	Sim	79.6% e 85.71%, respectivamente	75.4% e 85.7%, respectivamente	89.1% e 87.3%, respectivamente	86% e 84%, respectivamente		
Salama <i>et al.</i> , 2021	InceptionV3, DenseNet121, ResNet50, VGG16 e MobileNetV2 + U-Net modificado (segmentação)	DDSM + seu aumentado MIAS CBIS-DDSM Separadamente	DDSM + seu aumentado MIAS CBIS-DDSM Separadamente	Sim	Sim	U-net + InceptionV3 (melhor combinação) 98.87% (incidência MLO), 99.43% (MLO + CC)	U-net + InceptionV3 (melhor combinação) 98.87% (incidência MLO), 99.43% (MLO + CC)	U-net + InceptionV3 (melhor combinação) 98.98% (incidência MLO), 99.12% (MLO + CC)		U-net + InceptionV3 (melhor combinação) 97.99% (incidência MLO), 98.98% (MLO + CC)	U-net + InceptionV3 (melhor combinação) 98.88% (incidência MLO), 99.22% (MLO + CC)

Como se pode observar, todos os modelos tiveram boa performance, com métricas elevadas, alguns se destacando um pouco mais que os demais. Tais resultados constituem base fundamental para a discussão que seguirá em seções subsequentes, explorando as implicações e relevância desses achados.

3 FUNDAMENTAÇÃO TEÓRICA

Esta seção visa construir um alicerce teórico para enriquecer o entendimento da pesquisa, trazendo um breve histórico dos CADs, bem como conceitos fundamentais a respeito de Deep Learning e Redes Neurais. Por fim, são explicadas as métricas mais utilizadas para medir o desempenho dos modelos.

3.1 CAD em Detecção do Câncer de Mama

Os sistemas CAD têm suas origens nas décadas de 50, com a primeira ideia concebida. Na década de 60, tiveram início as pesquisas em torno de métodos automáticos de detecção e classificação de anormalidades nas imagens médicas, incluindo imagens de mama. Um dos primeiros sistemas automatizados para assistir radiologistas na análise de mamografias foi desenvolvido por um time da Universidade de Chicago em 1987. E em 1998, ocorreu a primeira aprovação de um CAD pela U.S.

Food and Drug Administration (FDA). A partir de então, os CAD foram passando por evoluções importantes, sendo separados em duas categorias, os CADs convencionais, desenvolvidos por volta de 2009 a 2017, com técnicas de Machine Learning e os CADs desenvolvidos a partir de 2017, adotando técnicas de Deep Learning (Hassan; Hamad; Mahar, 2022).

Ambos os sistemas possuem como etapas, o processamento de imagens, o qual engloba o pré-processamento, etapa responsável por preparar as imagens para as fases subsequentes, eliminando ruídos, corrigindo contraste, redimensionando, etc. e a segmentação que consiste na delimitação das ROIs. A diferença entre eles reside nas outras fases, a extração e seleção de características e a classificação, sendo que nos CADs com Deep Learning, essas duas últimas se tornam uma única fase (Hassan; Hamad; Mahar, 2022).

Os sistemas CAD que utilizam técnicas de Machine Learning envolvem a atividade manual de extração de um conjunto de características das imagens médicas, como morfologia, textura, coloração ou intensidade. Em seguida, a esse conjunto é aplicado algum algoritmo, como Máquinas de Vetores de Suporte (Support Vector Machine - SVM), Florestas Aleatórias (Random Forest - RF) (Boumaraf *et al.*, 2021) para que seja feita a classificação. No entanto, a extração manual de características é considerada extremamente demorada e geralmente requer um conhecimento amplo da doença para definir características que sejam realmente representativas. Além disso, essas características podem ser indistinguíveis, o que prejudica muito a interpretação e compreensão do que foi aprendido a partir dos dados de entrada.

Já os CADs baseados em técnicas de Deep Learning foram desenvolvidos para superar as limitações dos sistemas convencionais, uma vez que as características não são ajustadas manualmente, mas extraídas de um conjunto de dados, a partir de imagens brutas, em um processo de aprendizado (Khan *et al.*, 2019).

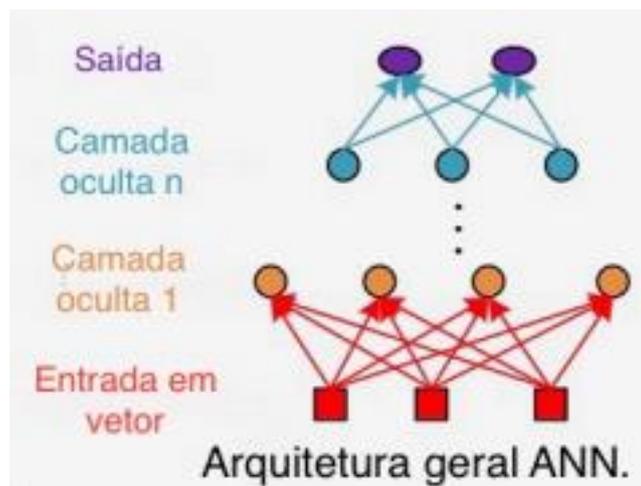
Com o desenvolvimento das CNNs, uma das arquiteturas de Deep Learning mais utilizadas para análise de imagens, os CADs alcançaram um desempenho sem precedentes no campo de análise de imagens de mamografia, principalmente devido a sua capacidade de aprender automaticamente as características dos dados de imagem (Boumaraf *et al.*, 2021). Porém, as CNNs funcionam melhor com grandes conjuntos de dados e também têm grande custo computacional. Para contornar esses problemas e melhorar o desempenho das CNNs, alguns outros conceitos têm sido explorados, como TL (Khan *et al.*, 2019), que consiste na utilização de um modelo CNN pré-treinada a partir de um conjunto de dados de imagens genéricas, como o Image Net e ajustado para tarefas relacionadas que serão efetuadas em um outro conjunto de imagens.

3.2 Redes Neurais Artificiais e Deep Learning

As Redes Neurais Artificiais (Artificial Neural Networks - ANNs) foram inspiradas no funcionamento do sistema nervoso humano e consistem basicamente em uma estrutura de camadas que operam por meio de neurônios interconectados os

quais processam individualmente dados de entrada, aplicando funções de ativação e encaminham a saída para a próxima camada, consoante se observa na Figura 2. Inicialmente, a estrutura possuía apenas uma camada de entrada e outra de saída e performava bem para classificar padrões linearmente separáveis, como por exemplo, a Perceptron, umas das primeiras ANNs (Razzak; Naz; Zaib, 2018). Mas, com a necessidade de lidar com padrões mais complexos, foram introduzidas uma ou mais camadas ocultas ao meio e, conforme foram ganhando profundidade, resultaram nas Redes Neurais Profundas (Deep Neural Networks - DNNs), capazes de classificar padrões não lineares. Estas, por conterem múltiplas camadas hierárquicas de neurônios, permitem a criação de representações em múltiplos níveis de características.

Figura 2 – Estrutura básica de uma ANN



Fonte: Hassan (2022). Traduzido para o português

Com camadas adicionais nas DNNs, as características extraídas de camadas mais baixas compõem as características das camadas mais altas, tornando-as ideais para lidar com dados complexos e realizar previsões inteligentes, como por exemplo a identificação de tumores em exames de imagens médicas (Razzak; Naz; Zaib, 2018).

O Deep Learning é uma subcategoria do campo de Machine Learning e inteligência artificial que utiliza DNNs (Razzak; Naz; Zaib, 2018). Suas redes neurais multiníveis derivam uma estrutura hierárquica de características a partir de dados

brutos de entrada, como imagens, possibilitando o reconhecimento de padrões e extração de informações para o aprendizado. Essa abordagem é distinta dos métodos tradicionais de Machine Learning, que dependem da extração manual de características (Houssein *et al.*, 2021). O desenvolvimento de algoritmos de Deep Learning ganhou impulso com o aumento na capacidade de processamento de Unidades de Processamento Gráfico (Graphics Processing Unit - GPU), permitindo a aprendizagem mais rápida dos pesos das conexões entre os neurônios (Razzak; Naz; Zaib, 2018), bem como o treinamento em grandes conjuntos de dados (Houssein *et al.*, 2021).

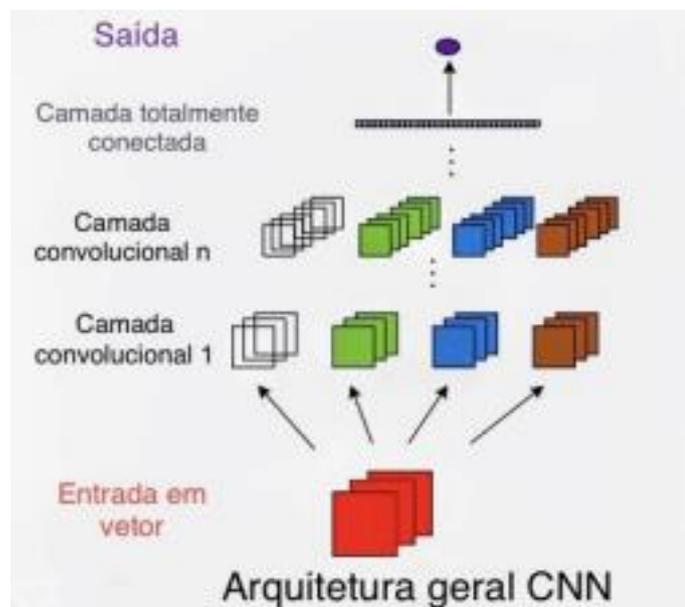
3.3 Redes Neurais Convolucionais

As CNNs são um tipo especial de DNN e foram especificamente desenvolvidas para encontrar características relevantes em imagens (Balkenende; Teuwen; Mann, 2022). São uma tecnologia bem-sucedida para resolver previsões nos casos em que os dados de entrada tem uma topologia em forma de grade conhecida, como é o caso das imagens, formadas por uma grade 2-D (Ketkar *et al.*, 2021). Tornaram-se o modelo de Deep Learning mais utilizado para o campo de análise de imagens médicas, pelo fato de eliminar a necessidade da extração manual de características e pela sua performance para as tarefas de reconhecimento, que superou as redes neurais convencionais, principalmente com o advento das GPUs (Yu *et al.*, 2021), que permitiram maior capacidade de processamento.

Uma CNN, assim como uma ANN, também é composta de três camadas, uma camada de entrada, uma oculta e outra de saída. Porém, a diferença entre elas reside na camada oculta. A camada oculta de uma CNN geralmente consiste em três subcamadas, a camada de convolução, a camada de subamostragem, também conhecida como camada de agrupamento e a camada totalmente conectada, conforme Figura 3. A primeira utiliza filtros para analisar partes da imagem e extrair características de alto nível, como bordas horizontais ou verticais. A camada de subamostragem é responsável por garantir as características mais importantes da imagem através de um processo de agrupamento máximo que reduz a quantidade de informações do mapa de características, dividindo a imagem de entrada em pequenos

quadrados não sobrepostos e escolhendo o valor mais alto presente em cada quadrado como sendo o seu valor representativo. Esse processo é útil para extração de características de baixo nível ou mais nítidas, como bordas e pontos. Outro processo utilizado por essa camada é o agrupamento comum, mais rápido e útil para características suaves. Por fim, a camada totalmente conectada tem como função, mapear as características extraídas das camadas anteriores e linearizadas em um vetor de forma que a rede possa realizar a tarefa de classificação ou reconhecimento. Após, geralmente se aplica uma função de ativação para essas tarefas, por exemplo, a "softmax", a fim de se obter uma distribuição de probabilidade (Yu *et al.*, 2021).

Figura 3 – Estrutura básica de uma CNN



Fonte: Hassan (2022). Traduzida para o português

São tipos de arquiteturas CNN a AlexNet, de 2012, uma das primeiras arquiteturas CNN que superou a acurácia das abordagens tradicionais de Machine Learning em reconhecimento de imagens, com 3 camadas convolucionais e 2 totalmente conectadas, bem como a Visual Geometry Group (VGG), de 2014, cuja contribuição foi demonstrar que a profundidade da rede é fator determinante para a precisão do reconhecimento e classificação. Os nomes de suas variações vieram seguidos dos números de camadas convolucionais, como a VGG-16. Também são exemplos a GoogleLe Net, de 2014, que incorporou o conceito de camadas Inception,

caracterizado pelo uso de múltiplos tamanhos de filtros de convolução em paralelo em várias camadas, permitindo a extração de padrões de correlação esparsos nos mapas de características resultantes, abordagem que ajuda a extrair informações relevantes de maneira eficaz e a Residual Network (ResNet), de 2015, desenvolvida com a intenção de projetar redes ultraprofundas que não sofressem com o problema do gradiente desvanecente presentes em modelos anteriores. A mais popular é a ResNet-50, com 50 camadas de profundidade, 49 convolucionais (Alom *et al.*, 2018).

Esses são apenas alguns dos diversos modelos de CNNs existentes, sendo que arquitetura por trás de uma CNN, a forma como suas camadas são organizadas e quais elementos essas últimas utilizam são fatores que influenciam na velocidade e precisão das tarefas realizadas (Houssein *et al.*, 2021).

3.4 Métricas

Diversas métricas são utilizadas para avaliar os resultados dos trabalhos e a performance dos algoritmos na detecção e classificação. Estas são baseadas em verdadeiros positivos (VP) e falsos positivos (FP), bem como verdadeiros negativos (VN) e falsos negativos (FN) (Hamed *et al.*, 2020). Tais termos dizem respeito aos acertos e erros das previsões do modelo em relação às classes reais dos dados. Por exemplo, VP é a situação em que a classe real tinha o valor 1 (Verdadeiro) e o modelo previu 1. As métricas mais encontradas nos estudos analisados foram melhor detalhadas a seguir.

Acurácia é a medida que indica o número de previsões corretas feitas pelo modelo em relação ao total de previsões realizadas (Hamed *et al.*, 2020).

$$\text{Acurácia} = \frac{VP+VN}{VP+VN+FP+FN}$$

Precisão é a medida que demonstra a proporção de casos que foram diagnosticados como câncer em relação aos que foram previstos como verdadeiros positivos e falsos positivos. (Hamed *et al.*, 2020). Os valores devem ser altos para prevenir o diagnóstico incorreto (Houssein *et al.*, 2021).

$$\text{Precisão} = \frac{VP}{VP+FP}$$

Sensibilidade, recall ou taxa de verdadeiros positivos mede a probabilidade de que um verdadeiro positivo é identificado corretamente pelo modelo (Hassan; Hamad; Mahar, 2022).

$$\text{Sensibilidade} = \frac{VP}{VP+FN}$$

Especificidade, também chamada de taxa de verdadeiros negativos (Hassan; Hamad; Mahar, 2022) mede a proporção de casos que não possuem câncer foram corretamente previstos pelo modelo como verdadeiramente negativos (Hamed *et al.*, 2020).

$$\text{Especificidade} = \frac{VN}{VN+FP}$$

278

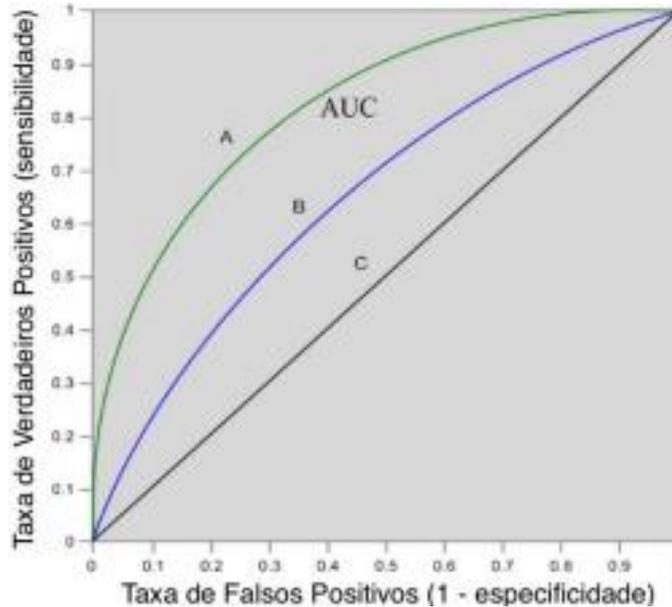
Pontuação F1 (F1-score) é uma medida utilizada para classificações binárias como classificação das massas entre malignas e benignas (Hassan; Hamad; Mahar, 2022). Representa uma média harmônica das medidas de precisão e sensibilidade, medindo a taxa de similaridade entre o que foi previsto e as regiões de referência verdadeira (Hamed *et al.*, 2020).

$$F1 - score = \frac{2 \times \text{Precisão} \times \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$$

Além dessas medidas, existem a Curva de Características de Operação do Receptor (curva ROC) e a Área Sob a Curva ROC (Area Under the Curve - AUC). A curva ROC tem grande importância na avaliação de performance dos sistemas CAD, uma vez que ilustra o equilíbrio entre a taxa de verdadeiros positivos e a taxa de falsos positivos considerando diferentes limiares de classificação (Hassan; Hamad; Mahar, 2022). Cada eixo representa uma dessas taxas. Já a AUC indica a capacidade do sistema de distinguir entre classes positivas e negativas, sendo que quanto maior o seu valor, melhor pode ser considerado o desempenho do modelo. A Figura 4

demonstra um exemplo de ROC com diferentes classificadores (A, B, C).

Figura 4 – Curva ROC e AUC



Fonte: Hassan; Hamad; Mahar (2022). Traduzido para o português.

279

Nesse exemplo, a curva ROC de A possui a maior AUC, portanto, tem desempenho superior (Hassan; Hamad; Mahar, 2022).

4 METODOLOGIA

Esta pesquisa teve como intuito realizar uma revisão dos modelos que utilizam técnicas de Deep Learning aplicadas à detecção e classificação de imagens de câncer de mama usando sistemas de CAD. Foi desenvolvida em quatro etapas, melhor descritas a seguir.

4.1 Coleta de Artigos

Para conduzir este estudo, foi feita uma extensa busca em bases de dados acadêmicas, como Google Scholar, IEEE, ScienceDirect e Springer. Foram utilizadas palavras-chave relevantes, incluindo "câncer de mama", "CAD", "deep learning", "reconhecimento de imagens", "redes neurais" e "CNN". A pesquisa se concentrou, em

sua maior parte, em artigos publicados nos últimos cinco anos para garantir a relevância e atualidade dos modelos.

4.2 Análise e Seleção de Artigos

Após a coleta inicial de artigos, foi realizada uma triagem com base em critérios de inclusão e exclusão. Os critérios de inclusão abrangeram estudos que abordam o uso de técnicas de Deep Learning, especialmente com modelos baseados em CNNs para a detecção e classificação de câncer de mama por meio de sistemas CAD e que utilizaram o mesmo banco de dados público, o DDSM. Além disso, também foram incluídos artigos que forneceram conceitos fundamentais sobre Deep Learning e Redes Neurais. Artigos que não atendiam a esses critérios ou estavam desatualizados foram excluídos.

4.3 Extração de Informações

280

Foi conduzida uma análise dos artigos selecionados para extrair informações relevantes, como:

- Conceitos fundamentais.
- Arquiteturas de CNNs mais utilizadas nos sistemas CAD.
- Conjuntos de dados utilizados para treinamento e validação.
- Métricas de avaliação de desempenho relatadas.
- Resultados.

4.4 Revisão e Discussão

Com base nas informações extraídas, foi feita uma revisão detalhada dos modelos de CNNs aplicados aos sistemas CAD para câncer de mama, bem como um demonstrativo das arquiteturas, desempenho e contribuições de cada modelo. Os resultados foram discutidos, assim como os desafios enfrentados pelos modelos de Deep Learning e as perspectivas futuras dessa área de pesquisa.

5 DISCUSSÃO

A Inteligência Artificial, assim como vem sendo empregada na resolução de inúmeros outros problemas nas mais diversas áreas do conhecimento, teve muita contribuição na área médica ao longo dos anos. Especificamente no campo de diagnósticos do câncer de mama, vem aperfeiçoando sistemas preocupados em analisar imagens médicas, detectar e classificar lesões suspeitas, os chamados CAD. O emprego de Deep Learning trouxe um impacto significativo na área. Suas técnicas e arquiteturas tornaram-se o estado da arte em sistemas CADs para a detecção do câncer de mama (Chan; Hadjiiski; Samala, 2020), superando a geração de técnicas tradicionais de Machine Learning, principalmente pelo fato de ter eliminado a necessidade de extração manual de características dos dados de entrada, como também pelos excelentes resultados, já que os modelos de Deep Learning alcançaram altas taxa de acerto em detecção e classificação.

Nesta pesquisa foram reunidos e revisados cinco modelos que se serviram das técnicas e arquiteturas de Deep Learning mais utilizadas no momento para esses CADs. Houve a preocupação de que as referidas pesquisas tratassem de analisar as imagens de um exame relativamente comum e acessível, a mamografia e também de que pudessem partir de um cenário semelhante, um mesmo banco de dados público, o DDSM, para que um comparativo entre elas pudesse ser elaborado. Todos os modelos apresentaram resultados excelentes. A maior parte, com métricas acima dos 95%.

O modelo de Al-masni (2018) demonstrou ser muito promissor para aplicação real, uma vez que demonstrou taxas de precisão expressivas e economia de tempo no trabalho de detecção e classificação, por realizar essas tarefas de modo simultâneo numa única imagem, em aproximadamente 3s. Além disso, a capacidade de identificar massas até sobre o músculo peitoral ou em mamas mais densas o torna um modelo ainda mais eficiente para a análise de imagens de mamografias. Isso, pois a visualização de tecidos mais densos nesse tipo de exame tem certa limitação, que é o problema da superposição dos tecidos, ou seja, o tecido fibroglandular pode "mascarar" lesões, tornando mais difícil a detecção (Balkenende; Teuwen; Mann, 2022).

O modelo de Al-antari (2020), comparado com o estudo acima descrito no qual teve participação, não observou mudanças significativas em métricas, mas o tempo para detecção e classificação diminuiu expressivamente, passando para a casa dos milissegundos. Ele também realiza a detecção e classificação simultânea de lesões a partir de uma imagem inteira e possui a capacidade de detectá-las sobre o músculo peitoral e tecidos densos.

Já o modelo de Fathy *et al.* (2019), teve uma abordagem bem diferente dos demais, utilizando um classificador CNN, a ResNet-50, para extrair da imagem classes normal e massa. Após, empregou o CAM para localizar a massa na imagem. Sua abordagem também obteve excelentes resultados.

O modelo de Rahman *et al.* (2020), não performou tão bem como os demais, mas os resultados ainda sim foram bons. Como pode se observar, o conjunto de dados utilizado para treinamento e teste, apesar de bem balanceado e processado, foi baseado apenas em uma máquina de escâner, fato que pode ter influenciado no resultado final. O estudo trouxe contribuições importantes como a constatação de que técnicas de pré-processamento, aumento de dados e TL são ferramentas com capacidade de superar a limitação de tamanho do conjunto de dados.

Por fim, o modelo de Salama *et al.* (2021) obteve um dos melhores resultados e também mostrou viabilidade de aplicação em cenários reais. Sua abordagem foi abrangente, explorando vários modelos de Deep Learning e comprovando suas eficácias e adaptabilidades em cenários diversificados, uma vez que os testaram em três bancos de dados diferentes. Uma característica marcante desse estudo foi introduzir uma fase de segmentação antes da fase de classificação, por meio da arquitetura CNN U-Net, a qual teve papel fundamental no desempenho do sistema. Em adição, o tempo computacional foi de um pouco mais que um segundo.

Embora numa análise superficial seja possível perceber que alguns dos modelos obtiveram melhores resultados, não foi possível concluir pelo melhor modelo, pois existem inúmeras variáveis que influenciam no resultado final e, apesar da tentativa, as condições em que cada modelo partiu não foram as mesmas. Por exemplo, todos eles foram pré-treinados (TL), não sendo possível verificar com quais bancos de dados. Também houve casos em que foram empregados conjuntos aumentados ou um conjunto de dados de outro banco além do DDSM. Ou ainda, um

conjunto de dados de apenas um tipo de escâner dentro do próprio DDSM. Além disso, o pré processamento de imagens foi diferente entre os modelos, o que também influencia nos resultados finais.

Considerando a possibilidade de aplicação clínica os modelos YOLO combinados com Redes Neurais Convolucionais demonstraram maior viabilidade, uma vez que, após treinados, a execução do sistema para detecção e classificação é muito rápida e simultânea, obtendo o resultado em poucos segundos ou milissegundos a depender do modelo, além de detectar lesões em tecidos mais densos e sobre o músculo peitoral. Entretanto, ainda não se pode firmar conclusão de que um sistema desses possa eliminar seguramente a expertise de um médico radiologista, sendo necessária ainda a revisão dos resultados e acompanhamento clínico dos casos. Ainda existem inúmeros desafios quando se trata da detecção do câncer de mama por sistemas CAD.

Um deles é a questão da interpretabilidade (Razzak; Naz; Zaib, 2018). Apesar de esses algoritmos de Deep Learning resolverem muito bem as tarefas de detecção e classificação de massas mamárias e de se poder compreender a matemática da construção das redes neurais, o processo pelo qual estes chegam a um resultado ainda é complexo e obscuro. E, quando se trata de diagnóstico médico, esse conhecimento é essencial.

Outra questão é a escassez de dados para treinamento, que ainda persiste apesar de ter sido amenizada com abordagens de pré-treinamento e técnicas de aumento de dados (Al-Masni *et al.*, 2018). Os modelos de Deep Learning precisam ser alimentados com uma grande quantidade de dados anotados e com a maior variedade possível, tanto na qualidade de imagens quanto nas características das mamas e das lesões, devendo ser considerada também a diversidade da população (Chan; HAD Jiiski; Samala, 2020), fatos que se transformam em grandes barreiras, já que existe um número relativamente pequeno de bancos de dados públicos disponíveis e a obtenção de imagens e dados de pacientes depende de diversas questões legais, já que são considerados dados sensíveis (Razzak; Naz; Zaib, 2018). Uma aposta moderna para contornar esse desafio é o emprego das Redes Adversárias Generativas (Generative Adversarial Networks - GAN) (Wu *et al.*, 2018) para a criação de imagens sintéticas de lesões de câncer de mama de maneira mais

realista que simples técnicas de aumento de dados (Hassan; Hamad; Mahar, 2022). Tal modelo é capaz de criar uma grande quantidade de imagens, gerando inúmeras variações que posteriormente treinarão e testarão modelos como as CNNs, por exemplo. É um caminho promissor para pesquisas futuras, juntamente com o aprimoramento de técnicas e arquiteturas.

6 CONCLUSÃO

Este estudo surgiu com a identificação de um problema global, principalmente para a população feminina, que é o câncer de mama, bem como da urgência do diagnóstico como forma de aumentar a taxa de sucesso no tratamento das pacientes e redução de mortes. Com isso, nasceu o propósito de contribuir com a pesquisa dos sistemas CAD com aplicação de Deep Learning para a evolução dos diagnósticos. Após detalhada revisão de modelos, concluiu-se que os modelos atuais que utilizam abordagens e arquiteturas de Deep Learning, principalmente as baseadas em CNN são muito promissoras para aplicação clínica, apresentando resultados precisos de detecção e classificação, muitos deles, superando a precisão humana, com destaque para os modelos YOLO combinados com arquiteturas CNN, que performaram muito bem e rápido, realizando a detecção e classificação simultaneamente em uma imagem inteira, detectando massas suspeitas até mesmo em situações mais complexas. No entanto, ainda não se pode firmar conclusões de que é possível eliminar a atuação humana para essas tarefas. Tais resultados, apesar de surpreendentes, nem sempre revelam a realidade, pois existem inúmeros fatores que podem alterá-los, principalmente o conjunto de dados utilizados para treinamento dos sistemas. Conforme já discutido, os modelos de Deep Learning necessitam de uma grande quantidade de dados para performar bem, sendo que, quanto maior a quantidade e a variação das imagens de entrada, mais condizentes com a realidade se tornam a detecção e a classificação, considerando que existem diversos tipos de massas e lesões de câncer. Esse problema já vem sendo amenizado com o uso de técnicas de transferência de aprendizado e aumento de dados, mas ainda não é o suficiente. Uma outra solução que vem sendo desenvolvida para vencer esses desafios é a utilização de GANs para a produção de um conjunto de dados sintético extenso, simulando as

mais diversas variações de imagens e anormalidades. É uma grande aposta para a evolução dos sistemas CADs atuais e um campo vasto para exploração e aprofundamento em pesquisas futuras.

REFERÊNCIAS

AL-ANTARI, M. A.; HAN, S.-M.; KIM, T.-S. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital x-ray mammograms. *Computer methods and programs in biomedicine*, Elsevier, v. 196, p. 105584, 2020.

AL-MASNI, M. A. *et al.* Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. *Computer methods and programs in biomedicine*, Elsevier, v. 157, p. 85–94, 2018.

ALJUAID, H. *et al.* Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 223, p. 106951, 2022.

ALOM, M. Z. *et al.* The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018.

285

BALKENENDE, L.; TEUWEN, J.; MANN, R. M. Application of deep learning in breast cancer imaging. In: ELSEVIER. *Seminars in Nuclear Medicine*. [S.l.], 2022.

BOUMARAF, S. *et al.* Conventional machine learning versus deep learning for magnification dependent histopathological breast cancer image classification: A comparative study with visual explanation. *Diagnostics*, MDPI, v. 11, n. 3, p. 528, 2021.

CHAN, H.-P.; HADJIISKI, L. M.; SAMALA, R. K. Computer-aided diagnosis in the era of deep learning. *Medical physics*, Wiley Online Library, v. 47, n. 5, p. e218–e227, 2020.

CHUGH, G.; KUMAR, S.; SINGH, N. Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cognitive Computation*, Springer, p. 1–20, 2021.

DOGRA, A.; GOYAL, B.; KAUSHIK, K. A brief review of breast cancer detection via computer aided deep learning methods. *International Journal of Engineering Research & Technology*, v. 8, n. 12, p. 326–331, 2019.

FATHY, W. E.; GHONEIM, A. S. A deep learning approach for breast cancer mass detection. *International Journal of Advanced Computer Science and Applications*, Science and Information (SAI) Organization Limited, v. 10, n. 1, 2019.

HAMED, G. *et al.* Deep learning in breast cancer detection and classification. In: SPRINGER. *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*. [S.l.], 2020. p. 322–333.

HASSAN, N. M.; HAMAD, S.; MAHAR, K. Mammogram breast cancer cad systems for mass detection and classification: A review. *Multimedia Tools and Applications*, Springer, v. 81, n. 14, p. 20043–20075, 2022.

HEATH, M. *et al.* The digital database for screening mammography. In: MEDICAL PHYSICS PUBLISHING. *Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe*. [S.l.], 2001. p. 212–218.

HOUSSEIN, E. H. *et al.* Deep and machine learning techniques for medical imaging based breast cancer: A comprehensive review. *Expert Systems with Applications*, Elsevier, v. 167, p. 114161, 2021.

International Agency for Research of Cancer. *Estimated number of new cases from 2020 to 2040, Females, age [0-85+] Africa + Latin America and Caribbean + Northern America + Europe + Oceania + Asia*. 2023. Disponível em: https://gco.iarc.fr/tomorrow/en/dataviz/bubbles?types=0&sexes=1_2&mode=cancer&group_populations=1&multiple_populations=1&multiple_cancers=1&cancers=20&populations=903_904_905_908_909_935&apc=cat_ca20v1.5_ca23v-1.5&group_cancers=0. Acesso em: 10 abr. 2023.

286

KETKAR, N. *et al.* Convolutional neural networks. *Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch*, Springer, p. 197–242, 2021.

KHAN, S. *et al.* A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, Elsevier, v. 125, p. 1–6, 2019.

KIM, M. *et al.* Deep learning in biomedical image analysis. In: *Biomedical information technology*. [S.l.]: Elsevier, 2020. p. 239–263.

RAHMAN, A. S. A. *et al.* Breast mass tumor classification using deep learning. In: IEEE. *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*. [S.l.], 2020. p. 271–276.

RAZZAK, M. I.; NAZ, S.; ZAIB, A. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps: Automation of Decision Making*, Springer, p. 323–350, 2018.

SALAMA, W. M.; ALY, M. H. Deep learning in mammography images segmentation and classification: Automated cnn approach. *Alexandria Engineering Journal*, Elsevier, v. 60, n. 5, p. 4701–4709, 2021.

SHEN, L. *et al.* Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, Nature Publishing Group UK London, v. 9, n. 1, p.

12495, 2019.

USF. *USF Digital Mammography Home Page*. 2023. Disponível em: <http://www.eng.usf.edu/cvprg/mammography/database.html>. Acesso em: 19 out. 2023.

WU, E. *et al.* Conditional infilling gans for data augmentation in mammogram classification. In: SPRINGER. *Image Analysis for Moving Organ, Breast, and Thoracic Images: Third International Workshop, RAMBO 2018, Fourth International Workshop, BIA 2018, and First International Workshop, TIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16 and 20, 2018, Proceedings 3*. [S.l.], 2018. p. 98–106.

YU, H. *et al.* Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing*, Elsevier, v. 444, p. 92–110, 2021. 12, 13