



UNIVERSIDADE
ESTADUAL DE LONDRINA

RAFAEL PALHETA TOKAIRIN

USO DE MODELOS GENERATIVOS PARA CRIAÇÃO DE
DATASET SINTÉTICO NO CONTEXTO DO CÂNCER DE
MAMA

LONDRINA
2025

RAFAEL PALHETA TOKAIRIN

**USO DE MODELOS GENERATIVOS PARA CRIAÇÃO DE
DATASET SINTÉTICO NO CONTEXTO DO CÂNCER DE
MAMA**

Trabalho de Conclusão de Curso apresentado
ao curso de Bacharelado em Ciência da Com-
putação da Universidade Estadual de Lon-
drina para obtenção do título de Bacharel em
Ciência da Computação.

Orientador: Prof. Dra. Helen C. de Mat-
tos Senefonte

Coorientador: Prof. Dra. Neyva Maria Lo-
pes Romeiro

**LONDRINA
2025**

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

T646	<p>Tokairin, Rafael Palheta. Uso de Modelos Generativos para Criação de Dataset Sintético no Contexto do Câncer de Mama / Rafael Palheta Tokairin. - Londrina, 2025. 66 f. : il.</p> <p>Orientador: Helen Cristina de Mattos Senefonte. Coorientador: Neyva Maria Lopes Romeiro. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Graduação em Ciência da Computação, 2025. Inclui bibliografia.</p> <p>1. Câncer de Mama - TCC. 2. Mamografia - TCC. 3. Redes Adversárias Generativas - TCC. 4. Redes Neurais Convolucionais - TCC. I. Senefonte, Helen Cristina de Mattos. II. Romeiro, Neyva Maria Lopes. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Graduação em Ciência da Computação. IV. Título.</p>
	CDU 519

RAFAEL PALHETA TOKAIRIN

**USO DE MODELOS GENERATIVOS PARA CRIAÇÃO DE
DATASET SINTÉTICO NO CONTEXTO DO CÂNCER DE
MAMA**

Trabalho de Conclusão de Curso apresentado
ao curso de Bacharelado em Ciência da Com-
putação da Universidade Estadual de Lon-
drina para obtenção do título de Bacharel em
Ciência da Computação.

BANCA EXAMINADORA

Orientador: Prof. Dra. Helen C. de Mattos
Senefonte
Universidade Estadual de Londrina

Prof. Dra. Neyva Maria Lopes Romeiro
Universidade Estadual de Londrina

Prof. Dr. Bruno Squizato Faiçal
Universidade Estadual de Londrina

Londrina, 24 de novembro de 2025.

*Este trabalho é dedicado às crianças adultas
que, quando pequenas, sonharam em se
tornar cientistas.*

AGRADECIMENTOS

Agradeço a Deus pela minha vida, por toda a minha trajetória até aqui e por todas as pessoas que me ajudaram nesse caminho.

Aos meus pais, expresso minha profunda gratidão por terem dedicado suas vidas à minha educação e por todo apoio que sempre me ofereceram. Agradeço também aos meus avôs e à minha avó, que sempre cuidaram de mim e me inspiram a ser uma pessoa melhor a cada dia. À minha irmã, sou grato pelo incentivo constante e por confiar em mim ao longo dessa jornada.

À minha orientadora, Prof. Dra. Helen C. de Mattos Senefonte, agradeço sinceramente pela paciência, dedicação e auxílio na realização deste trabalho. À minha coorientadora, Prof. Dra. Neyva Maria Lopes Romeiro, agradeço não apenas pelas orientações durante o Trabalho de Conclusão de Curso, mas também pelo apoio no projeto de iniciação científica. Sou eternamente grato pelo conhecimento transmitido, pelos aprendizados proporcionados e por ter acreditado em mim nos momentos em que mais precisei.

Agradeço ainda aos colegas de curso, que estiveram ao meu lado nos momentos de alegria e dificuldade, pela amizade e pelas trocas de conhecimento que somente a vivência universitária pode proporcionar. Por fim, agradeço ao corpo docente pelo conhecimento transmitido, que contribuiu significativamente para minha formação acadêmica.

*“E Deus disse: ‘Haja luz!’ E houve luz.
(Bíblia Sagrada, Gênesis 1,3)*

TOKAIRIN, R. P.. **Uso de modelos generativos para criação de dataset sintético no contexto do câncer de mama.** 2025. 65f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Estadual de Londrina, Londrina, 2025.

RESUMO

Esta pesquisa tem como objetivo contribuir com os médicos no diagnóstico do câncer de mama por meio da aplicação de dados sintéticos e técnicas de aprendizado de máquina. Com o avanço das abordagens de aprendizado de máquina, especialmente redes neurais convolucionais, busca-se explorar o potencial das tecnologias na identificação de características em imagens de mamografias que não são detectáveis visualmente. Diante da escassez de dados rotulados e de alta qualidade, foi desenvolvido um conjunto de dados sintéticos utilizando o modelo gerativo *StyleGAN2-ADA condicional*, com filtragem baseada na métrica *Learned Perceptual Image Patch Similarity* para garantir alta similaridade perceptual em relação aos dados reais. O conjunto final de dados reais, composto por imagens balanceadas entre benignas e malignas, foi avaliado utilizando a rede *EfficientNet-B0* em validação cruzada estratificada de cinco dobras. Diferentes proporções de imagens sintéticas foram incorporadas ao treino (1:1, 2:1, 3:1 e 4:1), mantendo a avaliação restrita a imagens reais. A inclusão de dados sintéticos mostrou melhoria consistente nas métricas de desempenho, alcançando acurácia média de 0.651, F1-score de 0.658 e AUC de 0.703 na proporção 2:1, em comparação com 0.606, 0.573 e 0.665 obtidos apenas com dados reais. Estes resultados indicam que imagens sintéticas de alta qualidade podem ampliar conjuntos de dados médicos limitados, promovendo classificadores mais robustos e com melhor capacidade de generalização, contribuindo para avanços no diagnóstico do câncer de mama.

Palavras-chave: Câncer de mama. Redes gerativas adversárias. Aprendizado profundo. Aumento de dados. Redes Neurais Convolucionais.

TOKAIRIN, R. P.. **Use of generative models for synthetic dataset creation in the context of breast cancer.** 2025. 65p. Final Project (Bachelor of Science in Computer Science) – State University of Londrina, Londrina, 2025.

ABSTRACT

This research aims to support physicians in the diagnosis of breast cancer through the application of synthetic data and machine learning techniques. With the advancement of machine learning approaches, especially convolutional neural networks, the goal is to explore the potential of these technologies in identifying features in mammography images that are not visually detectable. Given the scarcity of labeled and high-quality data, a synthetic dataset was developed using the conditional StyleGAN2-ADA generative model, with filtering based on the Learned Perceptual Image Patch Similarity metric to ensure high perceptual similarity to real data. The final real dataset, composed of images balanced between benign and malignant cases, was evaluated using the EfficientNet-B0 network in five-fold stratified cross-validation. Different proportions of synthetic images were incorporated into training (1:1, 2:1, 3:1, and 4:1), while evaluation remained restricted to real images. The inclusion of synthetic data consistently improved performance metrics, achieving an average accuracy of 0.651, F1-score of 0.658, and AUC of 0.703 in the 2:1 proportion, compared to 0.606, 0.573, and 0.665 obtained using only real data. These results indicate that high-quality synthetic images can expand limited medical datasets, enabling more robust classifiers with better generalization capabilities, ultimately contributing to advancements in breast cancer diagnosis.

Keywords: Breast cancer. Generative adversarial networks. Deep learning. Data augmentation. Convolutional neural networks.

LISTA DE ILUSTRAÇÕES

Figura 1 – Processo de aprendizagem supervisionada.	20
Figura 2 – Representação visual do classificador [1].	20
Figura 3 – Representação visual da regressão linear. Adaptada de [1].	20
Figura 4 – Representação visual de agrupamento [1].	21
Figura 5 – Representação visual de associação [1].	21
Figura 6 – Camadas de uma rede neural artificial <i>feed-forward</i> . Adaptada de [2].	22
Figura 7 – Diagrama esquemático de um neurônio artificial [3].	23
Figura 8 – Arquitetura de redes multicamadas. Adaptada de [4].	24
Figura 9 – Arquitetura de uma rede neural convolucional. Adaptada de [5].	25
Figura 10 – Operação de convolução. Adaptada de [6].	25
Figura 11 – Função de ativação ReLU.	26
Figura 12 – Comparação entre o <i>Max Pooling</i> e <i>Average Pooling</i> [7].	27
Figura 13 – Arquitetura do EfficientNet-B0 [8]	27
Figura 14 – Esquema de treinamento da GAN.	29
Figura 15 – Esquema de treinamento da WGAN-GP. Adaptada de [9].	32
Figura 16 – Esquema de treinamento da cGAN. Adaptada de [10].	33
Figura 17 – Imagens do conjunto de dados antes da filtragem.	41
Figura 18 – Arquitetura da cWGAN-GP	43
Figura 19 – Mamografias sintéticas geradas pela cWGAN-GP	43
Figura 20 – Gráficos dos valores do LPIPS na geração de 1000 imagens para cada classe usando WGAN-GP. Em laranja estão representadas as classes malignas e em azul as benignas.	44
Figura 21 – Arquitetura do gerador	45
Figura 22 – Arquitetura do discriminador	46
Figura 23 – Mamografias benignas: (a) real e (b) sintética.	49
Figura 24 – Mamografias malignas: (a) real e (b) sintética.	49
Figura 25 – Esquema de validação cruzada estratificada de 5 dobras.	52
Figura 26 – Gráficos dos valores do LPIPS na geração de 1000 imagens para cada classe usando StyleGAN2-ADA. Em laranja estão representadas as classes malignas e em azul as benignas.	56

LISTA DE TABELAS

Tabela 1 – Comparativo das abordagens para geração de mamografias sintéticas. . .	38
Tabela 2 – Resultados da validação cruzada (EfficientNet-B0).	57

LISTA DE ABREVIATURAS E SIGLAS

ML	Machine Learning (Aprendizado de Máquina)
DL	Deep Learning (Aprendizado Profundo)
CNN	Convolutional Neural Network (Rede Neural Convolucional)
SIFT	Scale-Invariant Feature Transform (Transformada de Características Invariantes à Escala)
SURF	Speeded-Up Robust Features (Características Robustas Aceleradas)
CAD	Computer-Aided Diagnosis (Diagnóstico Auxiliado por Computador)
ReLU	Rectified Linear Unit (Unidade Linear Retificada)
MLP	Multilayer Perceptron (Perceptron Multicamadas)
GAN	Generative Adversarial Network (Rede Generativa Adversária)
DCGAN	Deep Convolutional GAN (Rede Generativa Adversária Convolucional Profunda)
WGAN-GP	Wasserstein GAN with Gradient Penalty (GAN de Wasserstein com Penalidade de Gradiente)
cGAN	Conditional GAN (GAN Condicional)
PSNR	Peak Signal-to-Noise Ratio (Relação Pico Sinal-Ruído)
SSIM	Structural Similarity Index Measure (Índice de Similaridade Estrutural)
MSE	Mean Squared Error (Erro Quadrático Médio)
LPIPS	Learned Perceptual Image Patch Similarity (Similaridade Perceptual Aprendida de Pedaços de Imagem)
DDSM	Digital Database for Screening Mammography (Base de Dados Digital para Rastreio Mamográfico)
MIAS	Mammographic Image Analysis Society (Sociedade de Análise de Imagens Mamográficas)
FID	Fréchet Inception Distance (Distância de Fréchet na Inception)

U-Net	U-Net Convolutional Network (Rede Convolucional em U)
BUSI	Breast Ultrasound Images Dataset (Conjunto de Dados de Imagens de Ultrassom Mamário)
CRN	Contextual Residual Network (Rede Residual Contextual)
PatchGAN	Patch-based GAN Discriminator (Discriminador GAN baseado em Pedaços)
CycleGAN	Cycle-Consistent Generative Adversarial Network (Rede GAN com Consistência Cíclica)
SNGAN	Spectral Normalization GAN (GAN com Normalização Espectral)
ciGAN	Context-aware Image GAN (GAN de Imagens com Consciência de Contexto)

SUMÁRIO

1	INTRODUÇÃO	15
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Visão computacional	17
2.1.1	Visão computacional na área médica	18
2.2	Aprendizagem de máquina	18
2.2.1	Aprendizagem supervisionada	19
2.2.2	Aprendizagem não supervisionada	20
2.3	Redes neurais	21
2.3.1	Neurônio artificial (Perceptron)	22
2.3.2	Redes multicamadas	23
2.4	Redes neurais convolucionais	24
2.4.1	Detectores de características (Kernels)	25
2.4.2	Função de ativação ReLU	26
2.4.3	<i>Pooling</i> (Subamostragem)	26
2.4.4	<i>Flattening</i> e camadas densas	27
2.4.5	EfficientNet-B0	27
2.5	Inteligência artificial gerativa	28
2.5.1	DCGAN	31
2.5.2	WGAN-GP	31
2.5.3	<i>Conditional GAN</i>	32
2.5.4	StyleGAN2-ADA	33
2.6	Métricas de avaliação de qualidade de imagem	34
2.6.1	PSNR	35
2.6.2	SSIM	35
2.6.3	LPIPS	35
2.6.3.1	Funcionamento	36
2.6.3.2	Vantagens e limitações	36
2.7	Trabalhos correlatos: dados mamográficos rotulados	37
3	DESENVOLVIMENTO	40
3.1	Ambiente Computacional	40
3.2	Conjunto de dados	40
3.3	Geração de imagens sintéticas	42
3.3.1	Arquitetura do modelo de geração	42
3.3.2	Pré-processamento de imagens	46

3.3.3	Treinamento do modelo de geração	47
3.3.4	Uso do LPIPS para avaliação de qualidade das imagens geradas	48
3.4	Modelo de classificação	50
3.4.1	Arquitetura do modelo	50
3.4.2	Seleção das imagens sintéticas para o modelo de classificação	50
3.4.3	Divisão entre dados reais e sintéticos	51
3.4.4	Pré-processamento de imagens reais e sintéticas	51
3.4.5	Treinamento do modelo de classificação	52
3.4.6	Métricas de classificação	53
4	RESULTADOS	55
4.1	Geração de imagens de mamografia	55
4.2	Classificação de mamografias	56
5	CONCLUSÃO	59
	REFERÊNCIAS	60

1 INTRODUÇÃO

O processamento de imagens médicas tem se destacado como uma abordagem promissora no diagnóstico precoce de lesões e no desenvolvimento de ferramentas de apoio à decisão clínica, especialmente em doenças de alta incidência como o câncer. O câncer é caracterizado por alterações na estrutura genética das células, que resultam em proliferação celular descontrolada e formação de tumores [11]. Dentre os diversos tipos, o câncer de mama é o mais comum entre as mulheres ao longo da vida [12]. A mamografia é o exame mais utilizado e confiável para a detecção precoce de alterações nas mamas, permitindo a identificação de anomalias antes mesmo que possam ser percebidas por meio de exame clínico [13]. Entretanto, a interpretação manual dessas imagens depende de profissionais experientes e está sujeita à variabilidade entre avaliadores, além de ser um processo repetitivo que pode levar a erros decorrentes de cansaço. Esses fatores tornam a análise mais lenta e menos padronizada, reforçando a necessidade de métodos automatizados capazes de agilizar o processo e aumentar a precisão dos diagnósticos. Estudos demonstram que a avaliação de imagens médicas pode variar significativamente entre especialistas devido a fatores subjetivos e ao estado físico e psicológico do observador [14].

Diante desse desafio, o aprendizado de máquina (*machine learning* – ML) oferece um potencial para a identificação de padrões complexos em imagens, sendo capaz de detectar características sutis que muitas vezes escapam à percepção humana [15]. Em particular, o aprendizado profundo (*deep learning* - DL), uma subárea do ML, tem se destacado em tarefas de reconhecimento de imagens, devido à sua capacidade de aprender representações hierárquicas dos dados a partir de grandes volumes de informação [16]. Entre as arquiteturas mais eficazes, destacam-se as redes neurais convolucionais (*convolutional neural networks* - CNN), especialmente desenvolvidas para o processamento de dados em formato de grade, como imagens [17].

Contudo, o treinamento eficaz de redes neurais profundas exige grandes bases de dados rotulados e de alta qualidade, que representa um desafio em contextos médicos, onde o acesso a imagens é restrito, os dados nem sempre são padronizados ou rotulados adequadamente. Para superar essa limitação, este trabalho de conclusão de curso propõe a geração de um *dataset* sintético por meio de modelos gerativos, com o objetivo de enriquecer o processo de treinamento das CNNs e aprimorar o desempenho na detecção de lesões mamárias.

Assim, este trabalho busca integrar técnicas de geração de dados sintéticos com algoritmos de ML aplicados à análise de imagens de mamografias, comparando os métodos disponíveis para selecionar o de maior eficácia diagnóstica. A Seção 2 apresenta a fundamentação teórica sobre os conceitos utilizados, abordando temas como visão com-

putacional, aprendizado de máquina, redes neurais convolucionais, inteligência artificial generativa e métricas de avaliação de qualidade de imagem, além de trabalhos correlatos. A Seção 3 descreve o desenvolvimento do estudo, detalhando o ambiente computacional, os conjuntos de dados empregados, os processos de geração e seleção de imagens sintéticas, bem como o modelo de classificação utilizado. A Seção 4 apresenta e discute os resultados obtidos a partir dos experimentos de geração e classificação de mamografias. Por fim, a Seção 5 traz as conclusões do trabalho, destacando as contribuições, limitações e possíveis direções para pesquisas futuras.

2 FUNDAMENTAÇÃO TEÓRICA

O Capítulo 2 apresenta a fundamentação teórica necessária para compreender os conceitos e técnicas utilizados ao longo deste trabalho. Inicia-se com uma introdução à visão computacional, abordando seus princípios básicos e aplicações, com destaque para sua utilização na área médica. Em seguida, são explorados os principais paradigmas de aprendizagem de máquina, incluindo as abordagens supervisionada, não supervisionada e por reforço, fundamentais para o desenvolvimento de sistemas inteligentes. O capítulo segue com a apresentação das redes neurais, desde o funcionamento de um neurônio artificial e suas funções internas, até a estrutura de redes multicamadas. A partir disso, são discutidas as redes neurais convolucionais, amplamente utilizadas em tarefas de análise de imagens, detalhando seus componentes como kernels, funções de ativação, *pooling* e camadas densas. Em um nível mais avançado, dentro das inteligências artificiais gerativas, são introduzidas as Redes Adversárias Generativas, destacando sua arquitetura, funcionamento interno (com redes geradora e discriminadora), processo de treinamento e variantes relevantes como DCGAN, WGAN-GP e Conditional GAN. Também são discutidos os desafios inerentes ao treinamento dessas redes, bem como suas aplicações práticas. Para avaliar a qualidade das imagens geradas, são apresentadas métricas específicas, como PSNR, SSIM e LPIPS, com explicações sobre seu funcionamento, vantagens e limitações. Por fim, a seção encerra com a apresentação de trabalhos correlatos, que contextualizam a pesquisa no estado da arte da área.

2.1 Visão computacional

A visão computacional é um ramo da ciência da computação que busca desenvolver algoritmos capazes de interpretar e extrair informações úteis a partir de dados visuais, como imagens e vídeos. Essa área integra conceitos de matemática, estatística, inteligência artificial e processamento digital de imagens para permitir que sistemas computacionais simulem a percepção visual humana [18, 19].

Os primeiros esforços em visão computacional remontam aos anos 1960, com o uso de técnicas básicas de segmentação e detecção de bordas. A partir das décadas de 1980 e 1990, surgiram abordagens mais robustas, como os detectores de características SIFT e SURF, que foram fundamentais para tarefas como reconstrução 3D e reconhecimento de objetos [20, 21]. Ainda nesse período, pesquisadores desenvolveram modelos estatísticos e probabilísticos, como campos aleatórios de Markov, para melhorar a modelagem de cenas visuais [22].

O grande avanço ocorreu na década de 2010 com o surgimento de arquiteturas baseadas em redes neurais profundas, especialmente as redes neurais convolucionais (CNNs), que possibilitaram a extração automática de padrões complexos diretamente das imagens. O marco dessa transição foi a vitória do modelo AlexNet na competição ImageNet em 2012, estabelecendo um novo padrão de desempenho [23].

Esse progresso consolida o entendimento do processo de visão computacional como um fluxo estruturado em três etapas principais [19, 24], descritas por: (i) extração de características (como cor, forma e textura), (ii) interpretação por meio de reconhecimento de padrões, e (iii) tomada de decisão com base nos resultados obtidos, como segmentação ou classificação.

Hoje, a visão computacional está presente em diferentes setores, como a indústria automotiva (condução autônoma e assistência ao motorista) [25], agricultura de precisão [26], com especial destaque, na medicina [27]. Dessa forma, será explorado a seguir como esses conceitos fundamentam o desenvolvimento de modelos geratitivos para mamografias.

2.1.1 Visão computacional na área médica

Na área médica, a visão computacional se destaca pela sua aplicação em sistemas de apoio ao diagnóstico, conhecidos como *computer-aided diagnosis* (CAD). Esses sistemas utilizam algoritmos de processamento de imagens para auxiliar profissionais da saúde na detecção, segmentação e análise de anomalias em exames como radiografias, ressonâncias magnéticas e tomografias computadorizadas [27, 28].

O uso de redes neurais profundas, especialmente CNNs, tornou-se relevante na análise de imagens médicas, dada sua capacidade de aprender representações discriminativas complexas. Essas redes são capazes de classificar tecidos, identificar lesões e monitorar o progresso de doenças [29].

Esses avanços reduzem significativamente os tempos de diagnóstico, aumentam a precisão clínica e minimizam erros humanos. Com a popularização de dispositivos médicos portáteis e o uso de modelos otimizados, a visão computacional médica caminha para soluções em tempo real, mesmo em contextos com infraestrutura limitada [30].

2.2 Aprendizagem de máquina

A Aprendizagem de Máquina é uma subárea da inteligência artificial que se concentra no desenvolvimento de algoritmos capazes de aprender a partir de dados e melhorar seu desempenho com o tempo, sem necessidade de reprogramação explícita. Esses algoritmos constroem modelos preditivos ou descritivos a partir de dados históricos, sendo

amplamente utilizados em tarefas como classificação, previsão, agrupamento e tomada de decisão [31, 32].

As principais categorias da aprendizagem de máquina são: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço.

2.2.1 Aprendizagem supervisionada

A aprendizagem supervisionada envolve o uso de dados rotulados, ou seja, cada entrada (conjunto de características) está associada a uma saída esperada (rótulo). O objetivo é construir um modelo que aprenda a associar corretamente as entradas às saídas, de forma que possa generalizar para dados novos [32, 33].

- **Principais características:**
 - O modelo aprende a partir de pares entrada/saída conhecidos.
 - Após o treinamento, pode prever o rótulo de novos exemplos.

- **Exemplos típicos no contexto médico:**
 - **Classificação:** Determinar se uma lesão mamária identificada em uma mamografia é benigna ou maligna.
 - **Regressão:** Estimar a probabilidade de malignidade de um nódulo ou o tamanho esperado de um tumor com base em características radiométricas extraídas da imagem.

O processo de aprendizagem supervisionada é bastante eficaz quando se dispõe de grande volume de dados anotados, e seu desempenho pode ser medido diretamente pela precisão das previsões em dados de teste [34]. A Figura 1 ilustra os principais passos desse processo. Para complementar, a Figura 2 apresenta uma representação visual de um classificador, enquanto a Figura 3 mostra a representação visual de um modelo de regressão linear (modelo estatístico que busca encontrar a relação linear entre uma variável dependente e uma ou mais variáveis independentes).

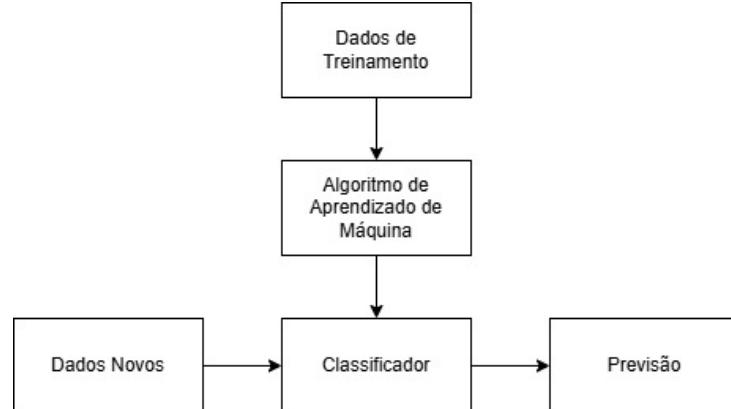


Figura 1 – Processo de aprendizagem supervisionada.

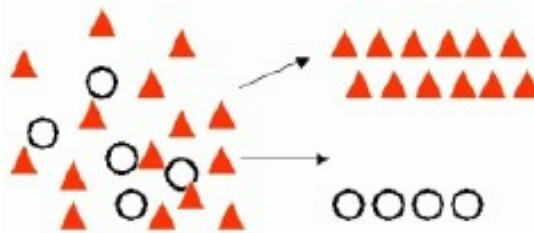


Figura 2 – Representação visual do classificador [1].

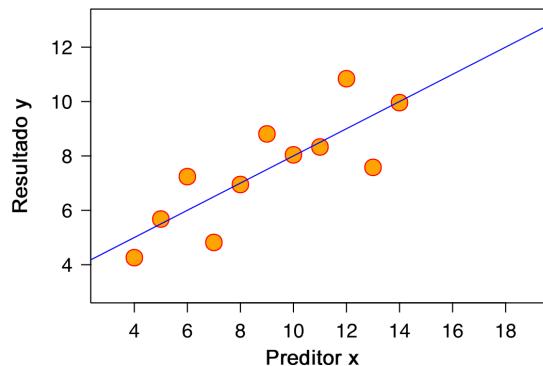


Figura 3 – Representação visual da regressão linear. Adaptada de [1].

2.2.2 Aprendizagem não supervisionada

Diferentemente da aprendizagem supervisionada, a aprendizagem não supervisionada lida com dados não rotulados. O objetivo é explorar a estrutura subjacente dos dados para identificar padrões, agrupamentos ou representações úteis [35, 36].

- **Principais características:**
 - O modelo não conhece os rótulos corretos.

- Utilizado para descoberta de conhecimento em grandes volumes de dados.
- **Exemplos típicos no contexto médico:**
 - **Clusterização:** Agrupar pacientes com base em características clínicas ou de imagem (por exemplo, mamografias com padrões de textura semelhantes), permitindo identificar subtipos de câncer de mama ou perfis de risco.
 - **Associação:** Descobrir relações entre variáveis médicas, como a ocorrência conjunta de determinadas características radiométricas e fatores genéticos ou hormonais, que podem estar associados à evolução da doença.

As técnicas empregadas na aprendizagem não supervisionada são amplamente usadas em análise exploratória de dados e em sistemas de recomendação [36, 37]. As Figuras 4 e 5 ilustram, respectivamente, a representação visual dos processos de clusterização e associação.

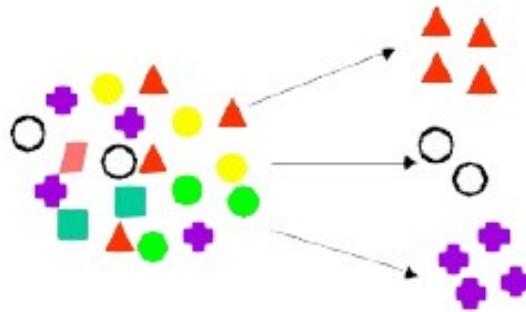


Figura 4 – Representação visual de agrupamento [1].

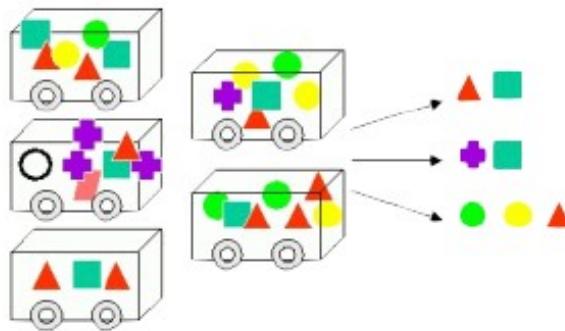


Figura 5 – Representação visual de associação [1].

2.3 Redes neurais

Dentro do campo da aprendizagem de máquina, as redes neurais artificiais se destacam como modelos computacionais inspirados na estrutura e funcionamento do cérebro

humano. Elas são compostas por unidades chamadas neurônios artificiais, organizadas em camadas interconectadas que processam informações de forma paralela, como ilustrado na Figura 6. Esses modelos são amplamente utilizados em tarefas como reconhecimento de padrões, processamento de linguagem natural e visão computacional [38, 39].

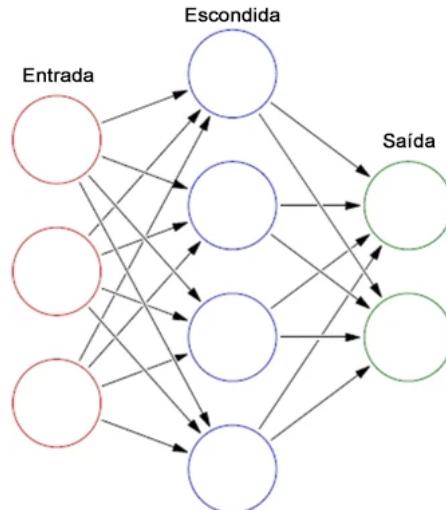


Figura 6 – Camadas de uma rede neural artificial *feed-forward*. Adaptada de [2].

2.3.1 Neurônio artificial (Perceptron)

O *Perceptron*, proposto por Rosenblatt em 1958, é o modelo mais simples de neurônio artificial, que simula o comportamento de um neurônio biológico ao combinar entradas ponderadas, aplicar uma transformação e gerar uma saída, como mostra a Figura 7. Esse modelo serviu como base para o desenvolvimento das redes neurais modernas [40, 41].

A função soma realiza a agregação ponderada das entradas recebidas. Considerando um conjunto de entradas x_1, x_2, \dots, x_n e seus respectivos pesos w_1, w_2, \dots, w_n , além de um termo de viés b . O resultado da soma é dado pela equação (2.1):

$$z = \sum_{i=1}^n w_i x_i + b, \quad (2.1)$$

onde x_i representa a entrada i , w_i o peso associado à entrada x_i e b termo de viés que permite deslocar a função de ativação.

Após o cálculo da função soma, aplica-se uma função de ativação que define a saída do neurônio. Essa função introduz não-linearidade no modelo, sendo essencial para que redes neurais possam modelar relações complexas [42, 43].

As principais funções de ativação são:

- **Degrau:** Função binária que retorna 0 ou 1 dependendo se z está abaixo ou acima de um limiar.

- **Sigmoide:** Função, apresentada na equação (2.2), que retorna valores entre 0 e 1, permitindo a interpretação probabilística da saída.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2.2)$$

- **ReLU (Rectified Linear Unit):** Introduz uma não-linearidade simples e eficiente, sendo amplamente adotada em redes neurais profundas, mostrada na equação (2.3) [44].

$$\text{ReLU}(z) = \max(0, z) \quad (2.3)$$

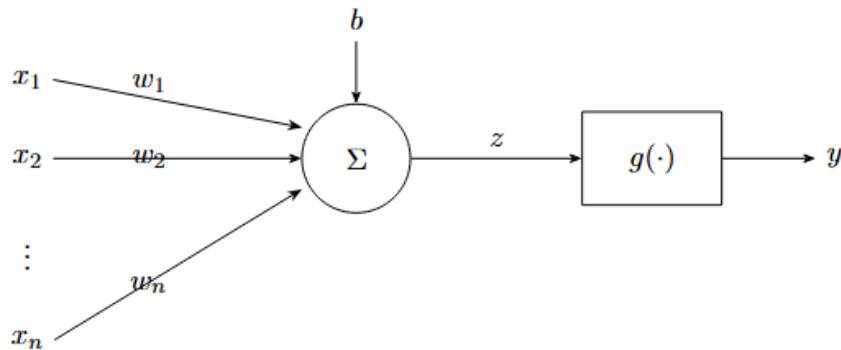


Figura 7 – Diagrama esquemático de um neurônio artificial [3].

2.3.2 Redes multicamadas

As redes do tipo *Multilayer Perceptron* (MLP) representam uma classe fundamental de redes neurais artificiais, caracterizadas por possuírem múltiplas camadas de neurônios totalmente conectadas, ilustrado na Figura 8. Essas redes são compostas por, no mínimo, três camadas: uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída [41, 38].

- **Camada de entrada:** Responsável por receber os dados iniciais e repassá-los para as camadas subsequentes.
- **Camadas ocultas:** Local onde ocorre o principal processamento dos dados. Cada neurônio de uma camada está conectado a todos os neurônios da próxima, caracterizando uma estrutura densa (*fully connected*). Quando uma rede possui muitas dessas camadas, ela é considerada uma rede profunda, formando a base do aprendizado profundo (*deep learning*) [39].
- **Camada de saída:** Responsável por gerar a saída final da rede, que pode representar uma classificação, regressão ou outro tipo de predição.

O aprendizado em redes multicamadas é viabilizado pelo algoritmo de retropropagação do erro, conhecido como *backpropagation*, que permite o ajuste eficiente dos pesos sinápticos da rede com base nos erros cometidos durante a previsão [45, 46].

O processo de treinamento com *backpropagation* pode ser descrito em três etapas principais:

1. **Forward pass:** Os dados fluem da entrada até a saída, passando por todas as camadas, e a rede gera uma predição.
2. **Cálculo do erro:** A diferença entre a saída gerada e o valor real é calculada por meio de uma função de perda, como o erro quadrático médio ou a entropia cruzada.
3. **Backward pass:** Utilizando a regra da cadeia do cálculo diferencial, o algoritmo computa os gradientes da função de perda em relação aos pesos da rede, os gradientes são utilizados para atualizar os pesos, geralmente com o algoritmo de gradiente descendente ou suas variantes [47].

O ciclo é repetido por várias iterações, conhecidas como épocas, o que permite à rede ajustar seus parâmetros internos para capturar padrões complexos nos dados.

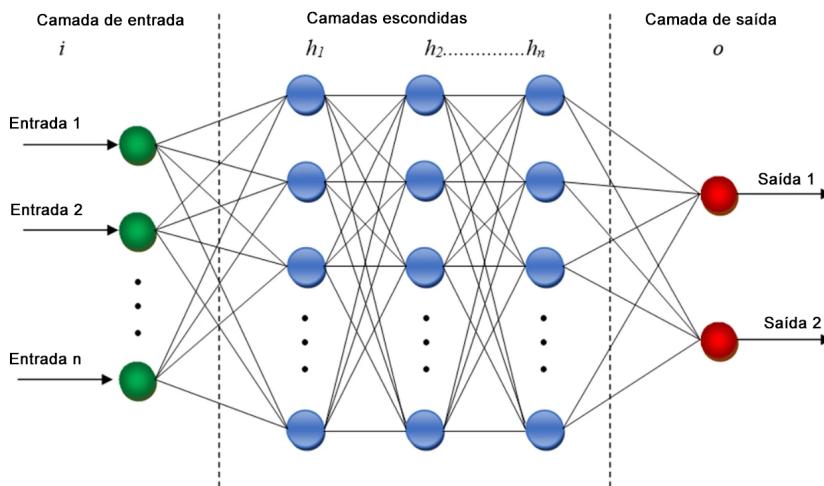


Figura 8 – Arquitetura de redes multicamadas. Adaptada de [4].

2.4 Redes neurais convolucionais

As redes neurais convolucionais (CNNs) são uma classe de modelos amplamente utilizada em tarefas de visão computacional, como reconhecimento de objetos, detecção de rostos ou classificar mamografias em benignas ou malignas. Inspiradas na organização do córtex visual animal, as CNNs utilizam camadas especializadas para extração automática de características espaciais em imagens, demonstradas na Figura 9 [16, 48].

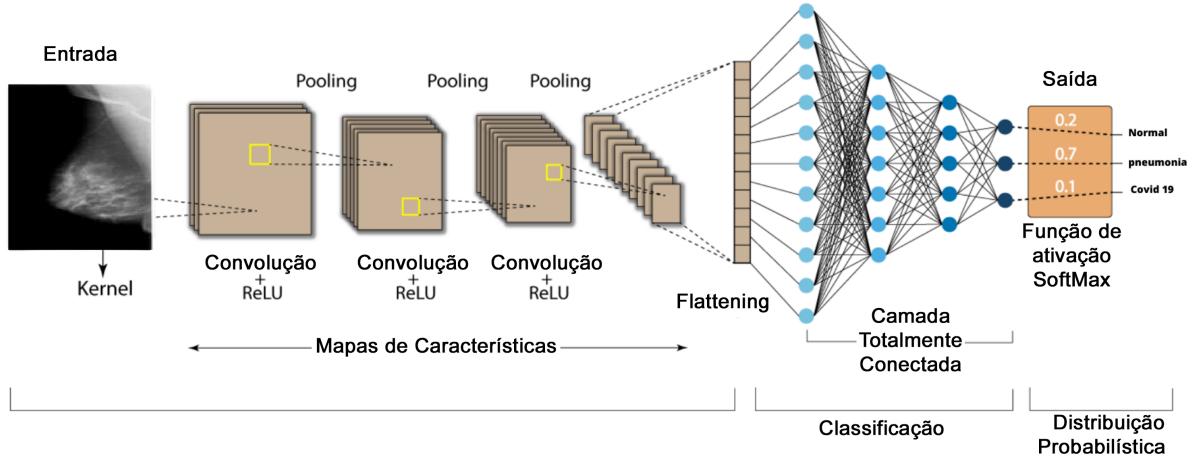


Figura 9 – Arquitetura de uma rede neural convolucional. Adaptada de [5].

2.4.1 Detectores de características (Kernels)

O núcleo funcional de uma CNN é a camada convolucional, que utiliza pequenos filtros chamados *kernels* ou *feature detectors*. Esses filtros percorrem a imagem de entrada realizando uma operação de convolução, Figura 10, responsável por extrair padrões locais, como bordas, texturas e formas simples [6].

A saída da operação de convolução $O(i, j)$ entre um kernel $F \in \mathbb{R}^{k \times k}$ e uma imagem I é dada pela equação 2.4:

$$O(i, j) = \sum_{m=1}^k \sum_{n=1}^k F(m, n) \cdot I(i + m - 1, j + n - 1) \quad (2.4)$$

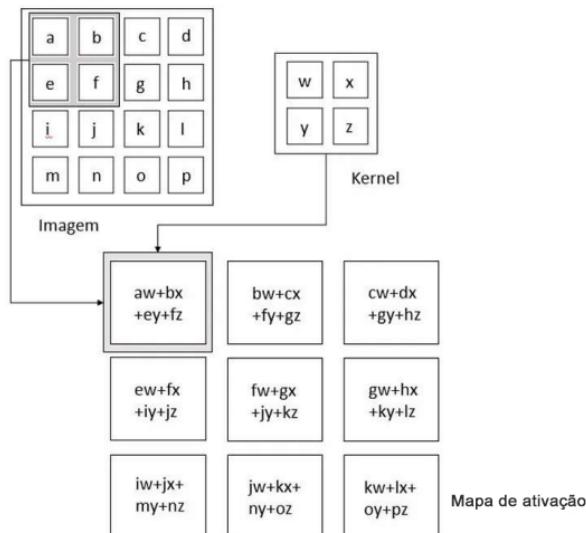


Figura 10 – Operação de convolução. Adaptada de [6].

2.4.2 Função de ativação ReLU

Após a convolução, é aplicada uma função de ativação para introduzir não-linearidade no modelo. A mais comum é a *ReLU* (*Rectified Linear Unit*), definida pela equação (2.5):

$$\text{ReLU}(x) = \max(0, x) \quad (2.5)$$

A ReLU descarta valores negativos e preserva os positivos, promovendo a esparsidade das ativações e reduzindo o risco de saturação, o que acelera o treinamento e melhora a performance de redes profundas, como representado na Figura 11 [49].

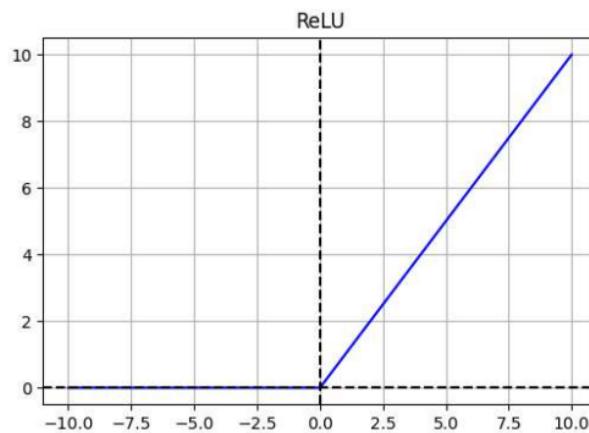


Figura 11 – Função de ativação ReLU.

2.4.3 Pooling (Subamostragem)

A operação de *pooling* tem como função reduzir a dimensionalidade das representações, Figura 12, preservar as características mais relevantes e evitar o sobreajuste. São utilizadas janelas móveis sobre a imagem convoluída para realizar uma redução da informação local.

- **Max Pooling:** Seleciona o maior valor da região analisada.
- **Average Pooling:** Calcula a média dos valores na região.

Entre os dois, o *Max Pooling* é preferido por sua capacidade de preservar contornos e elementos dominantes [50].

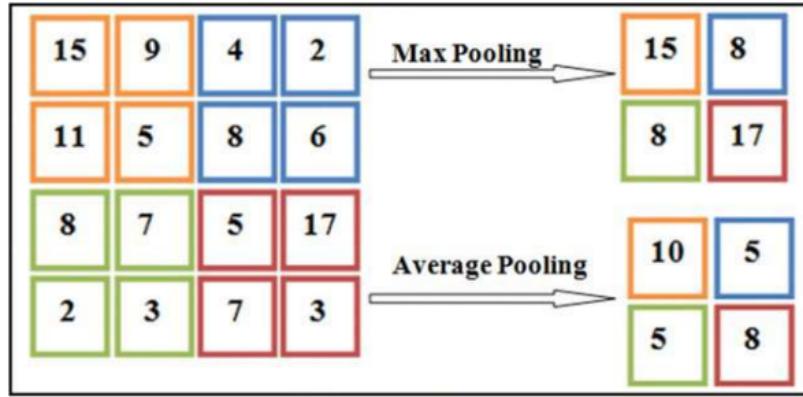


Figura 12 – Comparaçāo entre o *Max Pooling* e *Average Pooling* [7].

2.4.4 *Flattening* e camadas densas

Após várias camadas de convolução e *pooling*, a estrutura tridimensional resultante é convertida em um vetor unidimensional por meio do processo de *flattening*. Esse vetor é então passado para camadas densas (ou *fully connected*), que são responsáveis pelas decisões finais do modelo, como classificação ou regressão.

2.4.5 EfficientNet-B0

A EfficientNet-B0, representada pela Figura 13, é uma arquitetura moderna de *rede neural convolucional* projetada para equilibrar alto desempenho preditivo com eficiência computacional. Proposta por Tan & Le [51], esta arquitetura introduz o conceito de *compound scaling*, que permite dimensionar de forma balanceada a profundidade, largura e resolução da rede, ao contrário de abordagens tradicionais que aumentam apenas um desses fatores isoladamente.

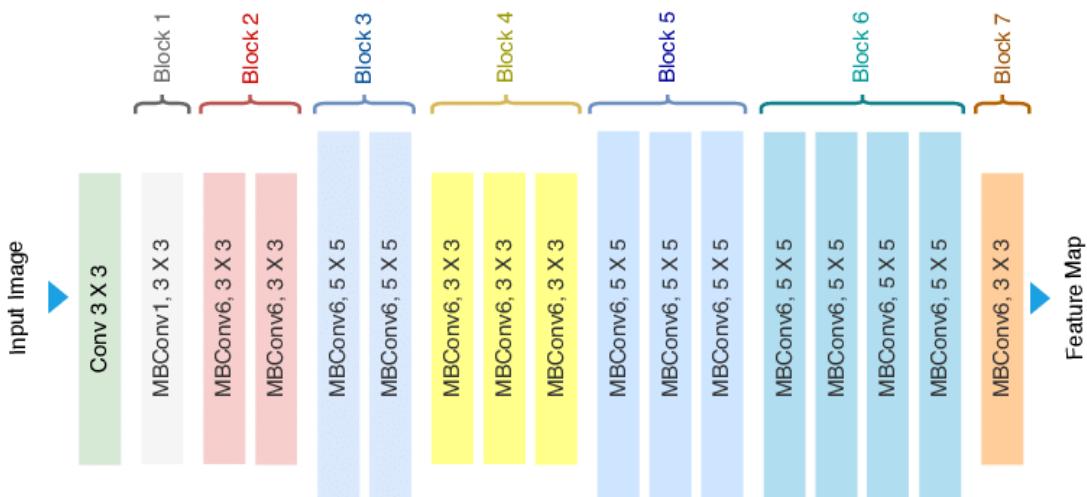


Figura 13 – Arquitetura do EfficientNet-B0 [8]

A arquitetura da EfficientNet-B0 é baseada em *Mobile Inverted Bottleneck Convolution (MBCConv)*, que combina convoluções ponto a ponto e convoluções *depthwise* para reduzir o custo computacional sem comprometer a expressividade da rede. Além disso, incorpora mecanismos de *squeeze-and-excitation*, que ajustam dinamicamente a importância de diferentes canais, melhorando a capacidade de capturar características relevantes da imagem.

Entre as principais contribuições da EfficientNet-B0 destacam-se: (i) alta eficiência computacional, permitindo desempenho competitivo mesmo em dispositivos com recursos limitados; (ii) melhor generalização em *datasets* de diferentes tamanhos e complexidades; (iii) uso otimizado de parâmetros, reduzindo o risco de sobreajuste; e (iv) capacidade de escalonamento sistemático para variantes maiores (B1–B7), mantendo consistência de desempenho em tarefas mais complexas.

A EfficientNet-B0 consolidou-se para tarefas de visão computacional, sendo amplamente utilizada em classificações de imagens médicas, reconhecimento de objetos e outros cenários que exigem redes precisas e leves. Sua combinação de precisão e eficiência estabelece novos padrões para o desenvolvimento de modelos robustos e práticos em aplicações reais.

2.5 Inteligência artificial generativa

Dentre os diversos modelos disponíveis para geração de dados sintéticos, foram utilizadas as Redes Adversárias Generativas, conhecidas como *Generative Adversarial Networks* (GANs), propostas por Goodfellow et al. em 2014 [52]. Elas representam uma das abordagens mais inovadoras para geração de dados sintéticos em diversas áreas, como visão computacional, processamento de linguagem natural e síntese de áudio.

Uma arquitetura típica de GAN é composta por dois modelos principais:

- **Gerador:** Responsável por produzir amostras sintéticas que buscam imitar os dados reais.
- **Discriminador:** Tenta distinguir entre as amostras reais, oriundas do conjunto de dados original, e as amostras falsas, geradas pelo Gerador.

Esses dois modelos são treinados de maneira simultânea em um processo de otimização adversarial, onde o gerador tenta enganar o discriminador, e este, por sua vez, tenta melhorar sua capacidade de distinção [53].

O treinamento de uma GAN é baseado em um processo competitivo, representado na Figura 14.

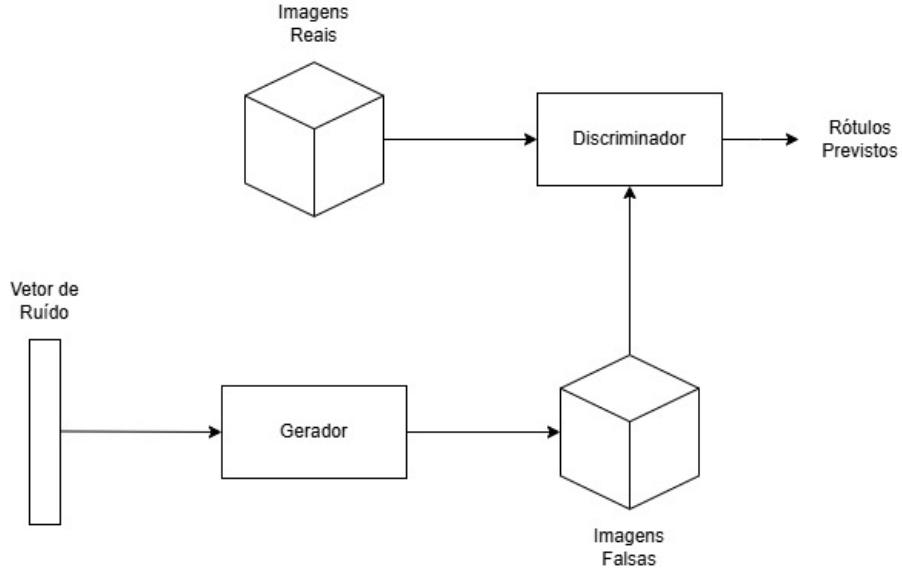


Figura 14 – Esquema de treinamento da GAN.

A seguir, detalham-se as principais funções de cada rede:

A Rede Geradora é uma rede neural que recebe como entrada um vetor de ruído aleatório, geralmente amostrado de uma distribuição uniforme ou normal, e gera uma saída $G(z)$ que busca se assemelhar aos dados reais. O objetivo é maximizar a probabilidade de que o discriminador classifique suas amostras como reais [54].

O Discriminador é uma rede neural que tem como função principal classificar as amostras como reais ou falsas. A rede é alimentada tanto com os dados reais do conjunto de treinamento quanto com os dados sintéticos produzidos pelo gerador. Sua saída é normalmente uma probabilidade, calculada via função de ativação sigmoide [55].

O treinamento de uma GAN é formalizado como um problema min-max, onde o discriminador tenta maximizar a probabilidade de classificar corretamente as amostras, enquanto o gerador tenta minimizar essa probabilidade. A função de perda é expressa pela equação (2.6):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] , \quad (2.6)$$

onde $D(x)$ descreve a probabilidade atribuída pelo discriminador para uma amostra real x , $G(z)$ a saída gerada pelo gerador a partir de um vetor de ruído z , $p_{\text{data}}(x)$ a distribuição dos dados reais e $p_z(z)$ a distribuição do ruído de entrada.

O processo de treinamento das GANs é realizado utilizando o algoritmo de retro-propagação do erro (*backpropagation*), com o método de *gradient descent* para atualização dos pesos das redes [54].

O treinamento ocorre em duas fases principais, repetidas iterativamente:

1. **Atualização do discriminador:** O discriminador é treinado para aumentar sua capacidade de classificar corretamente amostras reais e falsas.
2. **Atualização do gerador:** O gerador é treinado para melhorar a qualidade das amostras sintéticas, com o objetivo de enganar o discriminador.

Nas GANs destinadas ao processamento de imagens, como as *Deep Convolutional GANs* (DCGANs), tanto o gerador quanto o discriminador são implementados utilizando redes neurais convolucionais (CNNs) [55].

- O gerador utiliza camadas convolucionais transpostas (*Transposed Convolutional Layers*) para converter o vetor de ruído em uma imagem com estrutura espacial coerente.
- O discriminador aplica camadas convolucionais padrão para extrair características discriminativas que auxiliem na classificação entre imagens reais e falsas.

As GANs têm sido amplamente aplicadas em diversas áreas:

- **Geração de imagens realistas:** Produção de rostos humanos, objetos ou imagens médicas sintéticas [56].
- **Aprimoramento de imagens:** Técnicas de *super-resolution* para aumentar a resolução de imagens [57].
- **Transferência de estilo:** Alteração do estilo visual de imagens mantendo o conteúdo [58].
- **Geração de arte e design:** Criação automática de obras artísticas [59].
- **Geração de dados para treinamento:** Uso de dados sintéticos para平衡amento de conjuntos de dados ou aumento de amostras raras [60].

Apesar de seu potencial, o treinamento de GANs apresenta diversos desafios, entre eles:

- **Instabilidade no treinamento:** O equilíbrio entre as duas redes pode ser difícil de alcançar, resultando em flutuações ou não convergência [61].
- **Colapso de modo (*Mode collapse*):** O gerador pode aprender a produzir apenas um pequeno número de amostras, prejudicando a diversidade dos dados gerados [62].
- **Sensibilidade a hiperparâmetros:** Pequenas variações nos hiperparâmetros podem levar a grandes mudanças no desempenho do modelo.

2.5.1 DCGAN

A *Deep Convolutional Generative Adversarial Network* (DCGAN) representa um marco no desenvolvimento de redes gerativas, introduzindo pela primeira vez uma arquitetura totalmente convolucional para ambos os componentes da GAN. Proposta por Radford et al. [55], essa abordagem resolveu problemas fundamentais das GANs originais, especialmente na geração de imagens com alta fidelidade visual.

O núcleo inovador da DCGAN reside na substituição das camadas densamente conectadas por operações convolucionais especializadas. No gerador, camadas convolucionais transpostas permitem a transformação progressiva de um vetor latente de baixa dimensão em imagens com estrutura espacial coerente. Paralelamente, o discriminador utiliza uma arquitetura CNN convencional para classificação, porém otimizada para o contexto adversarial.

Entre as contribuições técnicas mais impactantes destacam-se: (i) a aplicação sistemática de *batch normalization* em todas as camadas exceto a de saída do gerador e entrada do discriminador, (ii) o uso combinado de funções de ativação ReLU (gerador) e LeakyReLU (discriminador), e (iii) a eliminação de camadas totalmente conectadas nas camadas intermediárias. Esses avanços permitiram pela primeira vez o treinamento estável de redes gerativas profundas.

A importância histórica da DCGAN se manifesta em três dimensões principais: estabelece padrões arquiteturais que permanecem relevantes, demonstra a viabilidade prática de GANs profundas e possibilita novas aplicações em síntese de imagens.

2.5.2 WGAN-GP

A WGAN-GP, ilustrada na Figura 15, surge como uma evolução crítica da arquitetura WGAN original, resolvendo limitações fundamentais no treinamento de redes gerativas. Desenvolvida por Gulrajani et al.[63] como aprimoramento do trabalho seminal de Arjovsky, Chintala, & Bottou [64], esta abordagem redefine o paradigma de otimização adversarial através da combinação inédita da métrica Wasserstein-1 com regularização explícita do gradiente.

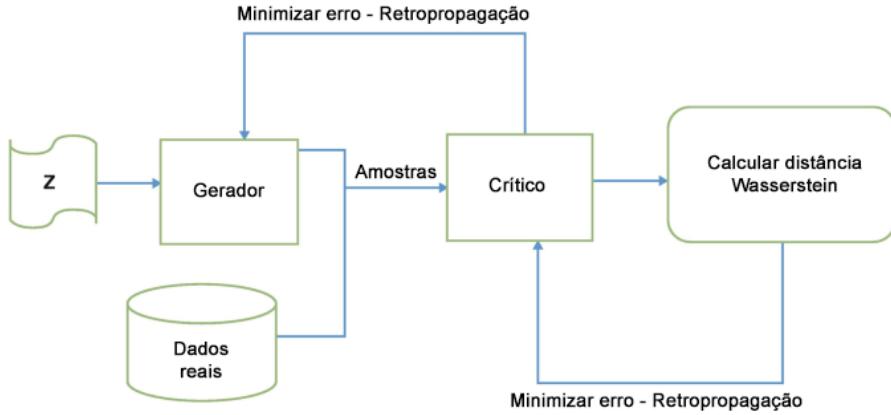


Figura 15 – Esquema de treinamento da WGAN-GP. Adaptada de [9].

O cerro teórico da WGAN-GP baseia-se em três pilares inovadores: (i) a substituição da divergência de Jensen-Shannon pela distância de Wasserstein (*Earth Mover's Distance*), que proporciona gradientes mais informativos mesmo com distribuições disjuntas; (ii) a imposição da condição de Lipschitz via penalização direta da norma do gradiente (Gradient Penalty), superando as limitações do *weight clipping* da WGAN original; e (iii) a reformulação da função custo para manter a continuidade do espaço latente durante o treinamento.

As implicações práticas dessa arquitetura manifestam-se em quatro dimensões principais. Primeiramente, a mitigação do *mode collapse* através da preservação da diversidade modal nas amostras geradas. Em segundo lugar, a estabilização notável do processo de treinamento, evidenciada pela convergência mais consistente em *benchmarks* complexos. Terceiro, a eliminação de artefatos de treinamento comuns em GANs tradicionais, como oscilações bruscas na função de perda. Por fim, a capacidade de treinar críticas (discriminadores) mais profundas sem comprometer a estabilidade numérica.

A WGAN-GP estabeleceu novos padrões para treinamento de modelos generativos, influenciando diretamente o desenvolvimento de arquiteturas subsequentes como Progressive GANs e StyleGAN. Sua principal contribuição reside na demonstração empírica de que a combinação de métricas Wasserstein com regularização geométrica permite o treinamento estável de redes profundas em espaços de alta dimensionalidade, superando limitações fundamentais das formulações adversarial clássicas.

2.5.3 *Conditional GAN*

A *Conditional Generative Adversarial Network* (cGAN), Figura 16, representa um avanço paradigmático no controle de saída de redes generativas, introduzindo o conceito de aprendizado condicional no *framework* adversarial. Proposta pioneiramente por Mirza [65], esta arquitetura estende o princípio básico das GANs ao incorporar informações

auxiliares tanto no gerador quanto no discriminador, permitindo controle preciso sobre as características das amostras geradas.

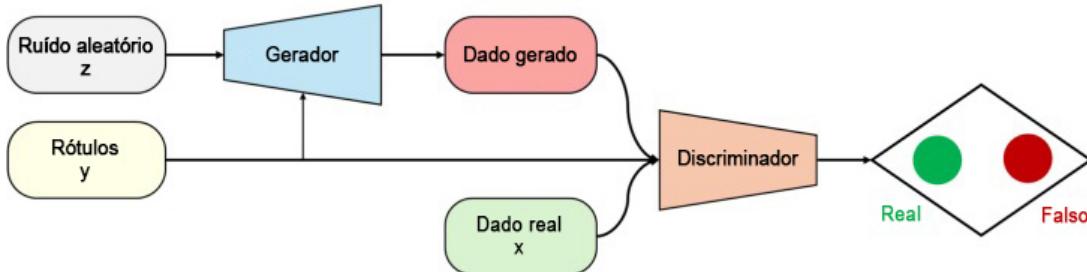


Figura 16 – Esquema de treinamento da cGAN. Adaptada de [10].

O mecanismo fundamental da cGAN baseia-se na concatenação do vetor latente tradicional z com vetores de condicionamento y , que podem representar classes discretas, atributos contínuos ou mesmo outras modalidades de dados. Esta modificação aparentemente simples produz impactos profundos: (i) o espaço latente torna-se particionado de acordo com as condições de entrada, (ii) o discriminador aprende a avaliar não apenas a realistica de mas também a consistência semântica com a condição fornecida, e (iii) o processo gerativo adquire capacidade de controle fino sobre os atributos de saída.

As aplicações da cGAN destacam-se em três domínios principais. Na síntese de imagens, permite a geração de amostras por categoria específica. Em tradução imagem-a-imagem, viabiliza mapeamentos condicionados entre domínios visuais. Por fim, em tarefas de manipulação, possibilita a edição precisa de atributos em imagens existentes (como modificação de expressões faciais ou estilos artísticos).

A cGAN estabeleceu as bases para toda uma família de modelos condicionais subsequentes, incluindo arquiteturas como AC-GAN, InfoGAN e *Projection GAN*. Sua principal contribuição reside na demonstração de que o *framework* adversarial pode ser estendido para aprendizado supervisionado e semi-supervisionado, mantendo simultaneamente as vantagens da geração não-condicional. Este avanço abriu caminho para aplicações práticas onde o controle preciso sobre as características geradas é essencial.

2.5.4 StyleGAN2-ADA

O *StyleGAN2-ADA* representa uma evolução significativa no campo de redes geradoras adversariais, incorporando avanços de estilo e regularização adaptativa para melhorar a qualidade de imagens sintéticas e a estabilidade do treinamento. Proposto por Karras et al. [66], esta arquitetura refina o *StyleGAN2* original ao introduzir *Adaptive Discriminator Augmentation (ADA)*, um mecanismo que aplica aumentos de dados de forma controlada no discriminador para combater sobreajuste, especialmente em conjuntos de dados de pequeno porte.

A arquitetura do *StyleGAN2-ADA* mantém o conceito central de *style-based generator*, em que um vetor latente z é transformado em um vetor de estilos w , que modula múltiplos *layers* de convolução através de operações de *adaptive instance normalization* (*AdaIN*). Esta abordagem permite controlar características de diferentes níveis de detalhe na imagem, desde atributos globais (como pose e estrutura) até texturas finas (como detalhes de pele ou tecidos). O mecanismo ADA, por sua vez, aplica aumentos de imagem adaptativos apenas quando o discriminador mostra sinais de sobreajuste, equilibrando a diversidade das amostras geradas sem degradar a qualidade.

Entre as principais contribuições do *StyleGAN2-ADA* destacam-se: (i) melhoria significativa na fidelidade visual das imagens sintéticas, (ii) redução de artefatos presentes nas gerações do StyleGAN original, (iii) robustez ao treinamento com datasets pequenos ou desbalanceados, e (iv) controle hierárquico dos atributos da imagem por meio do espaço de estilos.

O *StyleGAN2-ADA* consolidou-se como referência no desenvolvimento de GANs avançadas, servindo de base para variantes condicionais e projetos de geração de alta fidelidade. Sua capacidade de gerar imagens realistas com controle detalhado de atributos, mesmo com bases limitadas, destaca-se como um avanço para pesquisas que dependem de dados sintéticos de qualidade, conforme demonstrado por Karras et al. [66].

2.6 Métricas de avaliação de qualidade de imagem

A avaliação da qualidade de imagens geradas ou reconstruídas é uma etapa fundamental em tarefas de visão computacional, especialmente no contexto de redes neurais profundas, como autoencoders ou redes adversárias generativas (GANs). A escolha adequada da métrica de avaliação é importante para medir o quanto próxima a imagem produzida está da imagem original, tanto em termos objetivos quanto subjetivos.

Métricas tradicionais, como PSNR e SSIM, baseiam-se em comparações matemáticas entre pixels e propriedades estruturais da imagem, sendo amplamente utilizadas por sua simplicidade e eficiência. Contudo, com o avanço das técnicas de geração de imagens, surgiu a necessidade de métricas que reflitam melhor a percepção visual humana, pois pequenas distorções imperceptíveis podem resultar em baixas pontuações nas métricas tradicionais, mesmo quando as imagens parecem visualmente similares.

Neste contexto, surgem métricas perceptuais, como o LPIPS (*Learned Perceptual Image Patch Similarity*), que utilizam redes neurais treinadas para estimar a similaridade entre imagens de maneira mais próxima à forma como os humanos as percebem.

2.6.1 PSNR

O *Peak Signal-to-Noise Ratio* (PSNR) é uma métrica objetiva que quantifica a diferença entre a imagem original e a imagem reconstruída, sendo expressa em decibéis (dB). O PSNR é derivado do erro quadrático médio (MSE – *Mean Squared Error*), conforme definido na equação (2.7):

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \quad (2.7)$$

onde MAX_I é o valor máximo possível de pixel (por exemplo, 255 para imagens em 8 bits) e

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2. \quad (2.8)$$

Em geral, quanto maior o valor de PSNR, maior é a similaridade entre as imagens. Contudo, o PSNR não leva em consideração aspectos perceptuais ou estruturais da imagem, limitando sua efetividade em contextos onde pequenas variações semânticas são relevantes [67].

2.6.2 SSIM

O *Structural Similarity Index* (SSIM) foi proposto como uma alternativa ao PSNR, com o objetivo de quantificar a similaridade entre imagens considerando características perceptuais humanas, como luminância, contraste e estrutura. A métrica é definida pela equação (2.9):

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (2.9)$$

onde μ_x e μ_y são as médias das imagens x e y , σ_x^2 e σ_y^2 são as variâncias, σ_{xy} é a covariância, e C_1 , C_2 são constantes de estabilização [68].

O valor de SSIM varia entre -1 e 1, onde valores próximos de 1 indicam alta similaridade estrutural entre as imagens. O SSIM é mais robusto que o PSNR em diversos contextos, mas ainda apresenta limitações em avaliar imagens com alterações semânticas sutis.

2.6.3 LPIPS

O *Learned Perceptual Image Patch Similarity* (LPIPS) é uma métrica de similaridade perceptual baseada em redes neurais profundas, desenvolvida para avaliar a qualidade visual de imagens com maior correlação com a percepção humana [69]. Diferente

de PSNR e SSIM, o LPIPS opera sobre representações internas de redes convolucionais treinadas, como AlexNet, VGG ou SqueezeNet.

2.6.3.1 Funcionamento

O LPIPS funciona conforme os seguintes passos:

1. Duas imagens são passadas por uma rede neural convolucional pré-treinada.
2. Ativações (ou *feature maps*) de diferentes camadas da rede são extraídas para ambas as imagens.
3. Calcula-se a distância L2 entre as ativações correspondentes de cada camada.
4. L2 é ponderada por pesos treinados que refletem a importância perceptual das camadas.
5. A soma ponderada das distâncias resulta no valor final do LPIPS.

A equação geral do LPIPS é definida pela equação (2.10):

$$\text{LPIPS}(x, y) = \sum_l w_l \cdot \|f_l(x) - f_l(y)\|_2^2, \quad (2.10)$$

onde $f_l(\cdot)$ representa os mapas de características da camada l e w_l são os pesos aprendidos.

2.6.3.2 Vantagens e limitações

O LPIPS apresenta diversas vantagens:

- Alta correlação com avaliações humanas;
- Sensível a distorções perceptualmente relevantes;
- Ideal para tarefas envolvendo GANs ou geração de imagem.

Entretanto, a métrica também possui limitações:

- Depende da arquitetura da rede base;
- Custo computacional elevado;
- Escala não intuitiva para interpretação direta.

Valores próximos de zero, equação (2.10), indicam alta similaridade perceptual, enquanto valores maiores sugerem diferenças perceptíveis entre as imagens.

2.7 Trabalhos correlatos: dados mamográficos rotulados

Diversos estudos têm explorado o uso de Redes Geradoras Adversariais (GANs) na geração de imagens sintéticas para mitigar os desafios relacionados à escassez de dados mamográficos rotulados. A seguir, são apresentados os principais trabalhos correlatos que embasaram este TCC, com foco nas estratégias de geração de dados e nos impactos observados na performance de modelos de classificação.

Desai et al. [70] propuseram a utilização de DCGANs para gerar imagens sintéticas a partir de um conjunto limitado de mamografias do banco de dados DDSM. O objetivo do trabalho foi superar o problema da escassez de dados rotulados e melhorar a acurácia da classificação de lesões mamárias. Os autores validaram a qualidade das imagens geradas por meio de um teste de Turing com especialistas e análise via ImageJ, observando um aumento de 8,77% na precisão de um modelo CNN simples ao incluir imagens sintéticas no conjunto de treinamento.

Joseph et al. [71] avançaram ao empregar uma arquitetura de cGAN baseada em UNet e discriminador PatchGAN para a geração de imagens sintéticas a partir do banco de dados MIAS. O estudo demonstrou que o balanceamento do conjunto de dados por meio de imagens sintéticas levou a um ganho de 3,9% na acurácia da classificação de câncer de mama. A avaliação foi conduzida por meio da métrica de Distância de Início de Fréchet (FID), além de métricas tradicionais de classificação, como precisão e F1-score.

No trabalho de Guan e Loew [72], os autores compararam diretamente o impacto de imagens sintéticas geradas por GANs com técnicas tradicionais de aumento de dados (como transformações afins) no desempenho de um classificador CNN. Utilizando também o banco DDSM, os resultados indicaram que o uso de dados sintéticos levou a uma melhora de 3,6% na acurácia de validação em relação às transformações afins. No entanto, os autores ressaltam que as imagens reais ainda são insubstituíveis, pois o uso exclusivo de dados sintéticos pode acarretar *overfitting*.

Shah et al. [73] utilizaram o banco OPTIMAM para treinar uma rede DCGAN com arquitetura inspirada na U-Net, com o objetivo de gerar mamografias sintéticas que preservassem padrões e características realistas. A qualidade das imagens geradas foi avaliada por similaridade média, métricas estatísticas e validação por especialistas humanos, demonstrando que as imagens sintéticas produzidas possuíam alta fidelidade visual em relação às imagens reais.

Jiménez-Gaona et al. [74] realizaram um estudo comparativo entre diferentes variantes de GANs, incluindo WGAN-GP, CycleGAN, cGAN e SNGAN, para gerar dados sintéticos a partir de múltiplos bancos de dados (BUSI, DDSM, MIAS, InBreast). A análise revelou que o SNGAN apresentou melhor desempenho para mamografias em termos de FID, enquanto o CycleGAN obteve melhor diversidade e fidelidade visual. O estudo

também destacou que o desempenho dos modelos de classificação, como o ResNet-18, depende fortemente da escolha do modelo GAN e das características específicas do conjunto de dados.

Wu et al. [75] propuseram uma abordagem de GAN condicional com preenchimento contextual (ciGAN), com o objetivo de sintetizar lesões malignas em mamografias saudáveis. O treinamento foi feito em nível de patch (256x256) e as imagens geradas foram avaliadas quanto à capacidade de melhorar o desempenho de classificadores. O modelo ciGAN demonstrou ser capaz de gerar lesões realistas e aumentar a generalização dos modelos de classificação em relação a técnicas de aumento tradicionais.

Por fim, Shen et al. [76] propuseram uma abordagem em três estágios para gerar massas malignas realistas em mamografias normais. Inicialmente, máscaras de segmentação foram geradas por uma DCGAN, posteriormente combinadas com tecidos normais por um GAN de preenchimento baseado em CRN e U-Net. Os resultados indicaram que a abordagem aumentou a diversidade do conjunto de dados e melhorou o desempenho dos modelos de detecção em bases pequenas e desbalanceadas.

Tabela 1 – Comparativo das abordagens para geração de mamografias sintéticas.

Estudo	Abordagem	Base de Dados	Resultados
[70]	DCGAN	DDSM (287 imagens)	Acurácia +8.77%
[71]	cGAN	MIAS (322 imagens)	Melhoria de 3.9%
[72]	GAN tradicional	DDSM (2620 imagens)	+3.6% vs aumento afim
[73]	DCGAN modificada	OPTIMAM (26k imagens)	Validação especialista 87%
[74]	Multi-GANs	Vários conjuntos	FID 52,89 (SNGAN)
[75]	ciGAN	DDSM (10k imagens)	Superior a aumento tradicional
[76]	DCGAN + preenchimento	DDSM + Hospital Nanfang	Melhoria em datasets pequenos

Em síntese, os estudos analisados demonstram de forma consistente que o uso de GANs para geração de dados sintéticos contribui significativamente para a melhoria da acurácia e generalização de modelos de detecção de lesões mamárias, especialmente em contextos onde há escassez de dados rotulados. As abordagens variam desde arquiteturas mais simples como DCGANs até variantes mais sofisticadas como cGANs condicionais, ciGANs e GANs com normalização espectral, refletindo a evolução da aplicação dessas redes no domínio médico.

Diferentemente dos trabalhos analisados, esta pesquisa propõe uma abordagem que usa StyleGAN2-ADA, buscando maior estabilidade no treinamento e geração de imagens

de alta qualidade. Além disso, será utilizada a métrica LPIPS para avaliar a similaridade perceptual entre as imagens sintéticas e mamografias reais, proporcionando uma análise mais próxima da percepção humana em relação à fidelidade visual. Outro diferencial é o foco na utilização das imagens sintéticas para aprimorar a detecção de lesões benignas e malignas em classificadores, permitindo uma avaliação detalhada do impacto dos dados gerados não apenas na acurácia geral, mas também na capacidade dos modelos em diferenciar tipos de lesões.

3 DESENVOLVIMENTO

Este capítulo descreve as etapas envolvidas no desenvolvimento do trabalho, abrangendo desde a configuração do ambiente computacional até o treinamento e avaliação dos modelos propostos. Inicialmente, são apresentadas as especificações do ambiente de execução utilizado nos experimentos, incluindo hardware e software. Em seguida, descreve-se o conjunto de dados adotado e os procedimentos realizados para sua preparação.

Na sequência, detalha-se o processo de geração de imagens sintéticas, contemplando a arquitetura do modelo gerativo, as etapas de pré-processamento, os parâmetros de treinamento e o uso da métrica LPIPS para avaliação da qualidade perceptual das imagens geradas.

Por fim, são apresentados os métodos aplicados ao modelo de classificação, incluindo a definição da arquitetura, a integração de imagens reais e sintéticas, o pré-processamento dos dados e o treinamento do modelo. As métricas utilizadas para a avaliação de desempenho completam a seção, permitindo analisar o impacto do uso de imagens sintéticas na tarefa de classificação.

3.1 Ambiente Computacional

Os experimentos realizados neste trabalho foram conduzidos em uma máquina de uso pessoal equipada com *hardware* de desempenho intermediário, configurada para suportar as demandas de treinamento de modelos de aprendizado profundo. O sistema foi composto por uma GPU NVIDIA RTX 4060 com 8 GB de memória dedicada, um processador Intel Core i5-12400F de 6 núcleos e 12 *threads*, além de 16 GB de memória RAM 6000Mhz. O ambiente de desenvolvimento foi configurado em sistema operacional Debian Linux, com suporte às bibliotecas CUDA e cuDNN para aceleração de operações matriciais em GPU.

O treinamento dos modelos foi realizado utilizando a linguagem *Python*, em conjunto com o *framework* PyTorch¹, amplamente empregado em aplicações de visão computacional, mantendo tempos de processamento adequados e sem comprometimento significativo de desempenho.

3.2 Conjunto de dados

O conjunto de dados utilizado neste trabalho foi obtido a partir do Kaggle [77], contendo imagens de mamografias provenientes do *Mammographic Image Analysis So-*

¹ <<https://pytorch.org/>>

ciety (MIAS) Database, uma base amplamente utilizada para pesquisa em detecção e classificação de lesões mamárias. As imagens se encontram acompanhadas de informações estruturadas em um arquivo CSV, contendo as colunas *patient_id*, *breast_density*, *left or right breast*, *image view*, *abnormality id*, *abnormality type*, *mass shape*, *mass margins*, *assessment*, *pathology*, *subtlety*, *image file path*, *cropped image file path* e *ROI mask file path*. O dataset original possui um total de 1.898 imagens no formato JPEG, distribuídas entre imagens da mama completa (a), imagens recortadas contendo a lesão (b) e máscaras de regiões de interesse (c), mostrada na Figura 17.

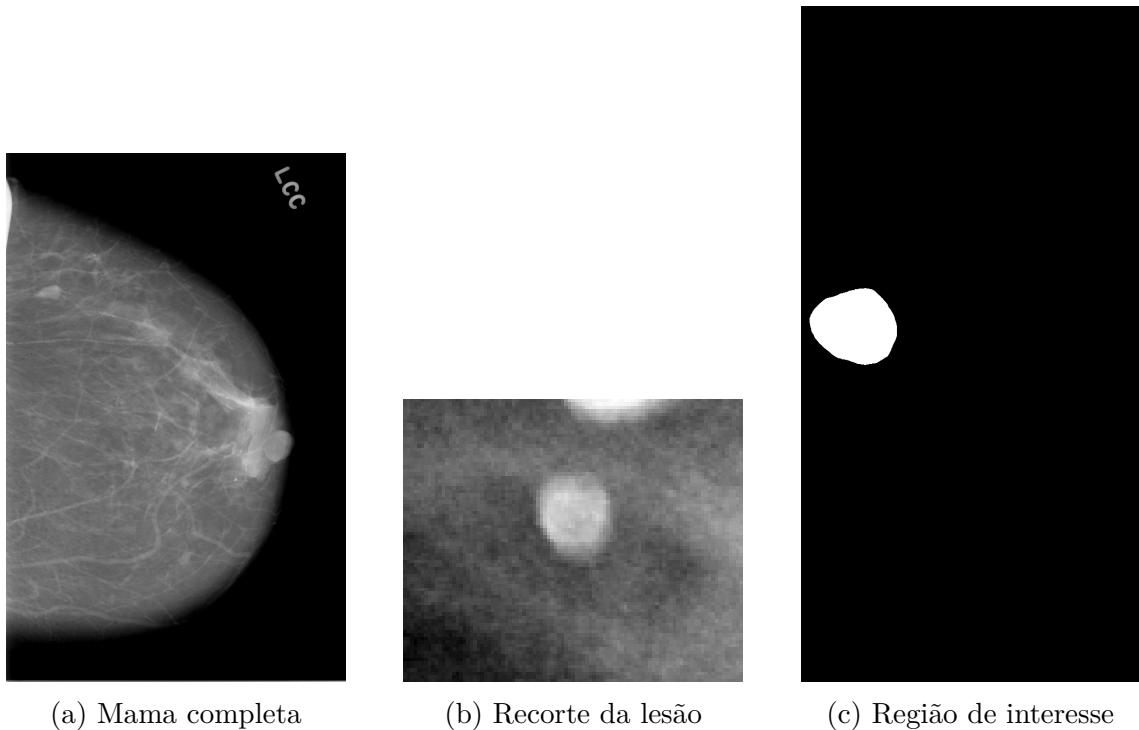


Figura 17 – Imagens do conjunto de dados antes da filtragem.

Para a utilização nesta pesquisa, foi necessário realizar um processo de filtragem e organização do conjunto de dados, de forma a selecionar apenas as imagens da mama completa ao treinamento do modelo gerador. Considerando que o objetivo da rede gerativa adversarial (GAN) é aprender padrões visuais globais das mamografias, optou-se pelo uso exclusivo das imagens de mama completa, descartando as versões recortadas e as máscaras de ROI. Após essa seleção, o conjunto final resultou em 553 imagens, sendo 245 classificadas como malignas e 308 como benignas.

Em seguida, o arquivo CSV original foi ajustado, mantendo apenas as colunas essenciais para o treinamento: *pathology* e *image file path*, correspondentes à classe da amostra e ao caminho da respectiva imagem. As amostras rotuladas como “benigno sem necessidade de retorno” foram reclassificadas como “benigno”, com o objetivo de simplificar a categorização e evitar subdivisões desnecessárias. Essa padronização possibilitou uma distinção clara entre as duas principais classes do estudo: benigno e maligno.

Por fim, todas as imagens foram orientadas na mesma direção, independentemente de representarem a mama esquerda ou direita. Essa normalização foi feita para evitar que o modelo aprendesse padrões relacionados apenas à posição anatômica, concentrando-se nas características visuais e estruturais das lesões. Dessa forma, o conjunto de dados foi devidamente preparado para as etapas subsequentes de pré-processamento, treinamento e avaliação do modelo de geração, garantindo coerência, qualidade e integridade das informações utilizadas.

Algumas limitações foram identificadas referentes ao conjunto de dados. Primeiramente, o número total de imagens é relativamente pequeno para a aplicação de modelos de classificação, o que aumenta o risco de sobreajuste (*overfitting*). Além disso, apesar do equilíbrio relativo entre classes benignas e malignas, o conjunto ainda é limitado em termos de diversidade de casos. Estratégias de aumento da variabilidade, como técnicas de *data augmentation* e geração sintética de imagens, foram adotadas e serão detalhadas nas próximas seções.

3.3 Geração de imagens sintéticas

3.3.1 Arquitetura do modelo de geração

O desenvolvimento do modelo gerativo deste trabalho envolveu a avaliação de diferentes arquiteturas de *Generative Adversarial Networks* (GANs) condicionais, incluindo DCGAN condicional e WGAN-GP condicional. Ambos os modelos foram inicialmente testados como base para a geração de mamografias sintéticas, permitindo observar a capacidade de aprendizado de padrões visuais das classes benignas e malignas. Embora tenham apresentado resultados coerentes em termos de estrutura geral das imagens, ambos demonstraram limitações na fidelidade visual e na preservação de detalhes anatômicos sutis, evidenciando a necessidade de uma arquitetura mais avançada.

Foi desenvolvida uma arquitetura híbrida que combinou elementos da DCGAN condicional e da WGAN-GP condicional, com o objetivo de unir a capacidade de representação convolucional profunda da primeira à estabilidade de treinamento e robustez da segunda. Essa fusão resultou na *Conditional Wasserstein Deep Convolutional GAN with Gradient Penalty* (cWGAN-GP), uma abordagem que emprega convoluções transpostas no gerador e convoluções diretas no discriminador, associadas à função de perda de Wasserstein com penalização de gradiente, que reduz oscilações durante o aprendizado adversarial, representada pela Figura 18.

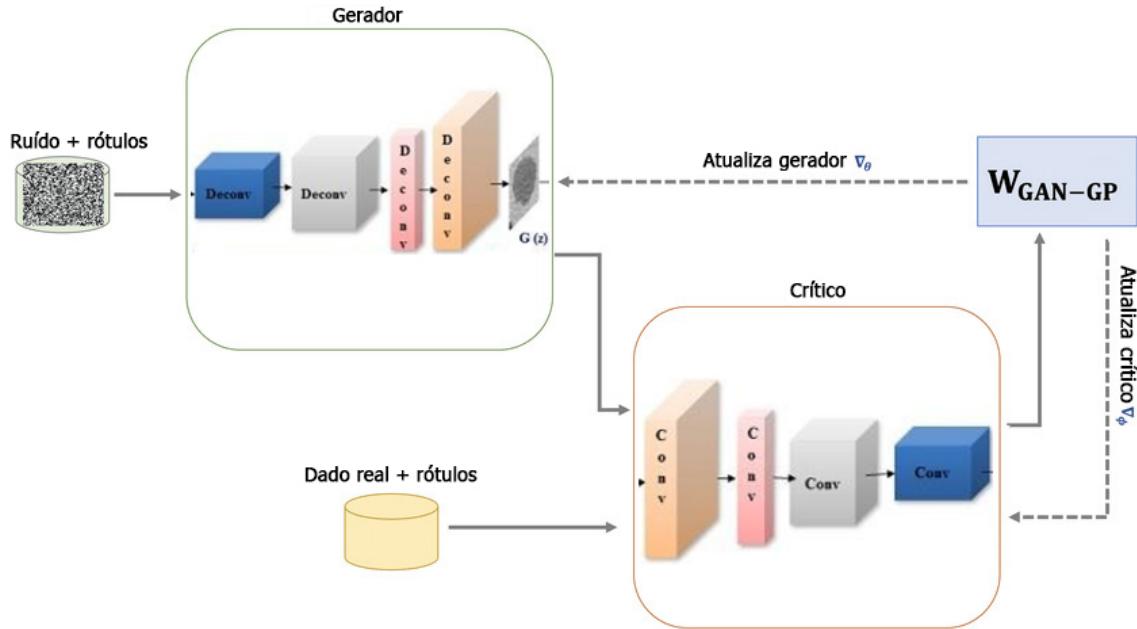


Figura 18 – Arquitetura da cWGAN-GP

O modelo recebe como entrada um vetor de ruído concatenado a um vetor de rótulo *one-hot*, permitindo controlar a classe de saída (benigno ou maligno). A implementação possibilitou gerar amostras com estrutura anatômica coerente e melhor estabilidade de convergência, embora ainda apresentasse ruídos texturais e imprecisões sutis nas bordas das imagens, como mostra a Figura 19, indicando a necessidade de arquiteturas mais avançadas.

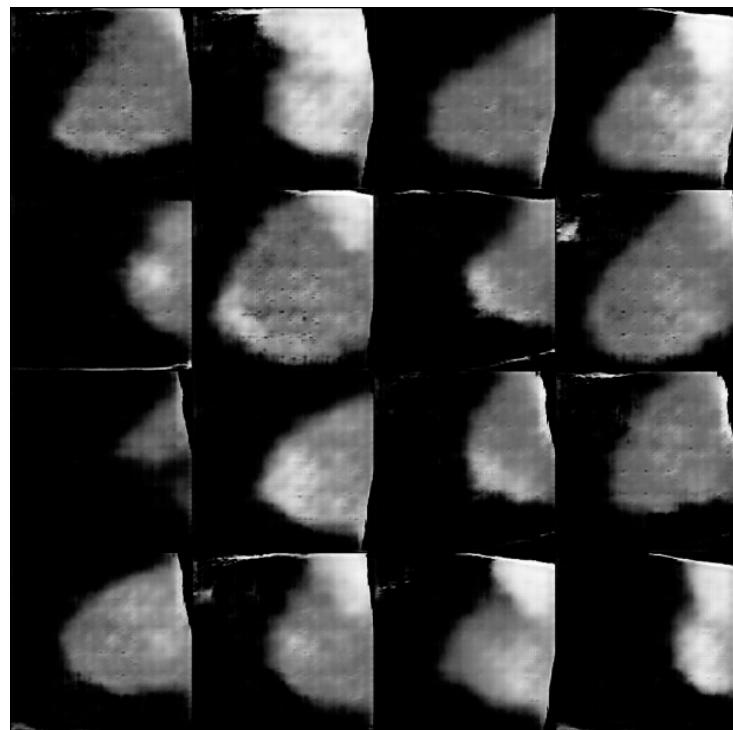


Figura 19 – Mamografias sintéticas geradas pela cWGAN-GP

Os resultados apresentados na Figura 20 mostram que a arquitetura cWGAN-GP obteve média de 0.352 e mediana de 0.347 nos valores de LPIPS, indicando uma similaridade perceptual moderada entre as imagens geradas e as reais. Apesar de demonstrar estabilidade na geração, esses valores refletem uma limitação do modelo em capturar com precisão as variações de textura e contraste características das mamografias.

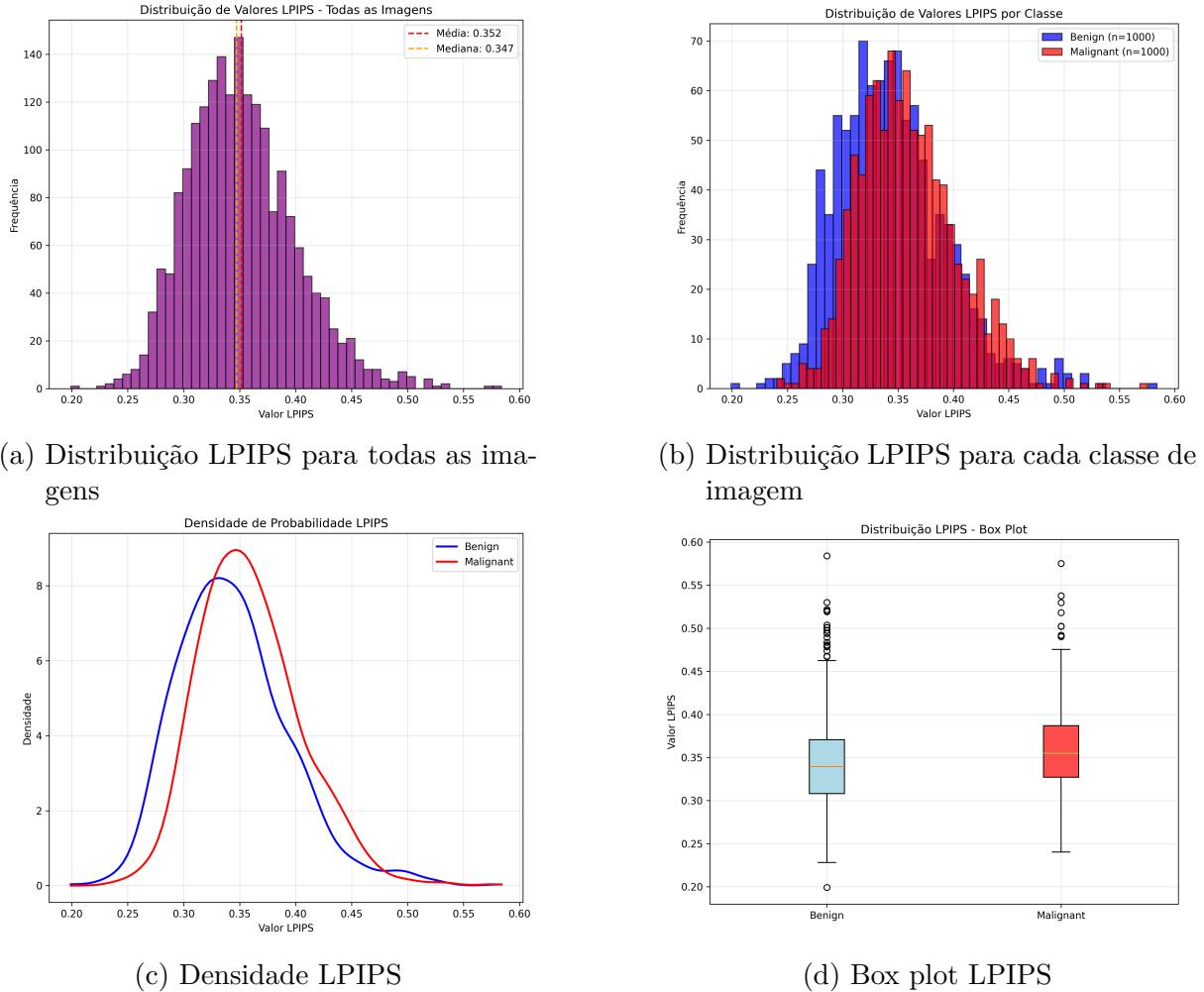


Figura 20 – Gráficos dos valores do LPIPS na geração de 1000 imagens para cada classe usando WGAN-GP. Em laranja estão representadas as classes malignas e em azul as benignas.

Considerando essas limitações e o objetivo de aprimorar a qualidade perceptual das imagens, optou-se pela adoção do *StyleGAN2-ADA* condicional. Essa escolha foi motivada pela capacidade do *StyleGAN2-ADA* de representar detalhes morfológicos complexos com maior realismo, aliando a abordagem *style-based* à técnica de *Adaptive Discriminator Augmentation* (ADA), o que possibilita um treinamento estável mesmo com conjuntos de dados limitados. Assim, a nova arquitetura busca superar as restrições observadas no cWGAN-GP, produzindo imagens de mamografia com maior fidelidade perceptual e melhor separabilidade entre as classes.

O gerador condicional do *StyleGAN2-ADA*, ilustrado na Figura 21, recebe um vetor latente combinado à condição da classe e o transforma progressivamente em uma imagem sintetizada por meio de blocos de convolução baseados em *style modulation*. Essa abordagem permite controlar características de alto e baixo nível durante o processo de geração, conferindo maior precisão na representação de estruturas morfológicas presentes em mamografias.

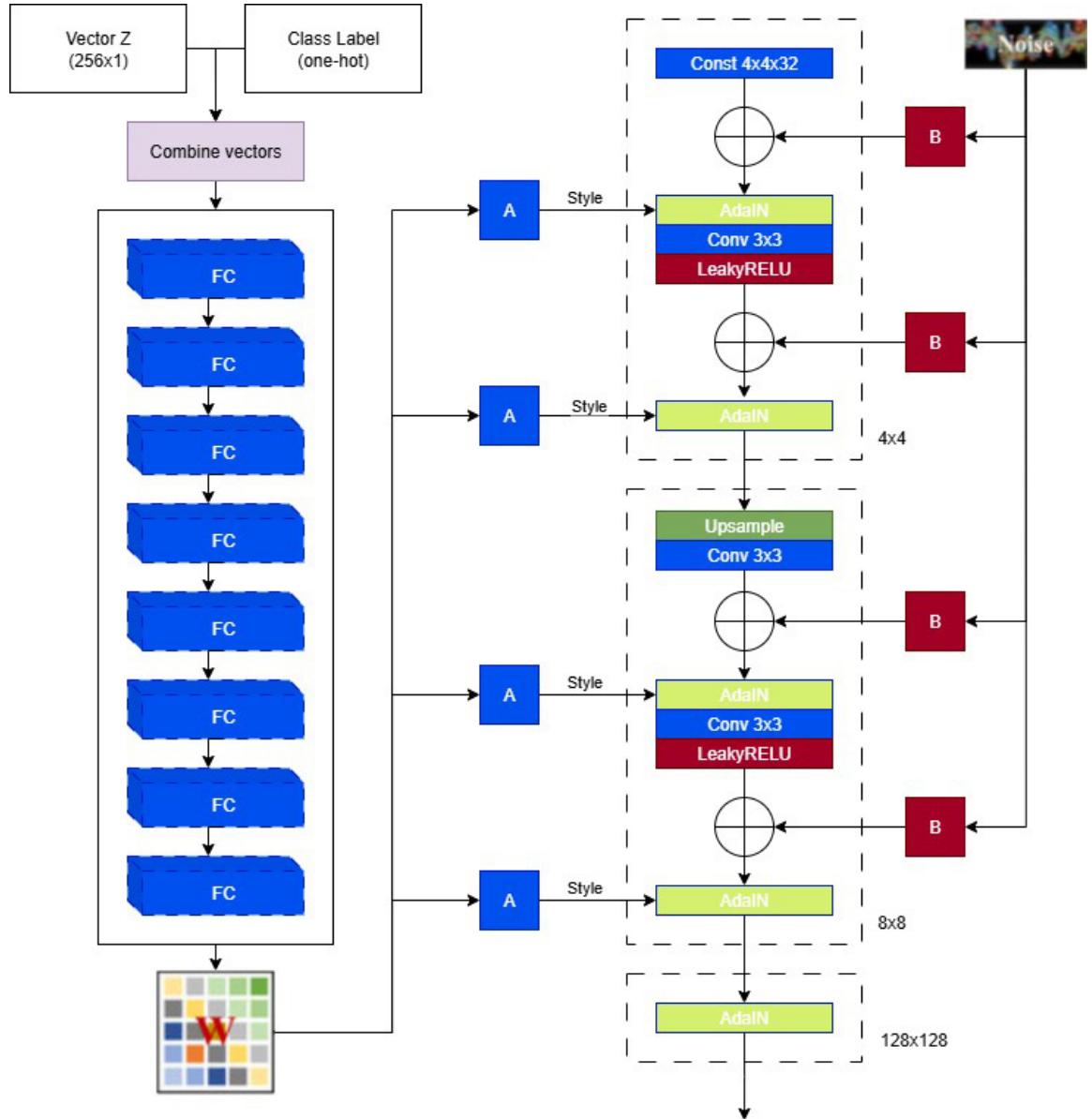


Figura 21 – Arquitetura do gerador

O discriminador condicional, por sua vez, recebe a imagem gerada concatenada à condição da classe e aplica uma sequência de blocos convolucionais com ativação Leaky ReLU, integrando os princípios de R1 regularization e ADA para melhorar a estabilidade e evitar sobreajuste (*overfitting*), Figura 22. Durante o treinamento, a arquitetura permite que o gerador aprenda características detalhadas das mamografias, enquanto o discrimi-

nador fornece *feedback* contínuo sobre realismo e consistência com a classe fornecida.

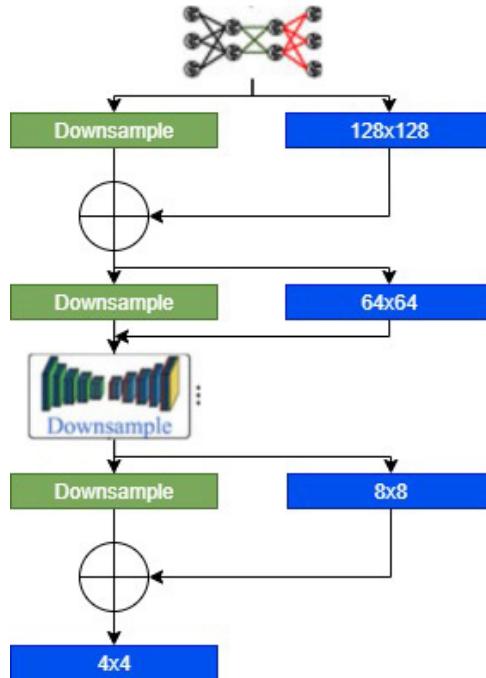


Figura 22 – Arquitetura do discriminador

O modelo StyleGAN2-ADA condicional demonstrou superioridade em comparação às arquiteturas previamente testadas, produzindo imagens sintéticas de alta qualidade, com melhor preservação de texturas e estruturas anatômicas críticas. A concatenação das informações de classe ao vetor latente possibilitou um controle mais preciso sobre o conteúdo gerado, garantindo que o modelo reproduzisse características específicas de mamografias benignas e malignas. Essa abordagem consolidou o StyleGAN2-ADA condicional como a arquitetura escolhida para a síntese de imagens neste estudo, estabelecendo uma base sólida para futuras etapas de aumento de dados e treinamento de modelos de classificação.

3.3.2 Pré-processamento de imagens

Para garantir que as mamografias estivessem em um formato adequado ao treinamento do modelo *StyleGAN2-ADA* condicional, todas as imagens foram primeiramente convertidas para escala de cinza, resultando em um único canal de entrada. Em seguida, as imagens foram redimensionadas para 128×128 pixels utilizando interpolação bilinear. Essa resolução inicial foi escolhida como um compromisso entre a qualidade visual das imagens e a limitação de memória da GPU, permitindo o treinamento com lotes (*batch size*) de 8 imagens.

Os valores dos pixels foram então convertidos em tensores do PyTorch e normalizados para o intervalo $[-1, 1]$, o que é compatível com as funções de ativação Leaky

ReLU empregadas nas redes do gerador e do discriminador. Cada imagem recebeu um rótulo binário indicando se a amostra era benigna (0) ou maligna (1), com base nas informações presentes no arquivo CSV de metadados. Esses rótulos foram posteriormente convertidos em vetores *one-hot* de dimensão 2, permitindo que o gerador criasse imagens condicionadas à classe correspondente.

Durante o treinamento, aplicou-se uma estratégia de aumento de dados adaptativa (*Adaptive Discriminator Augmentation, ADA*) para aumentar a diversidade das imagens e reduzir o sobreajuste. As transformações utilizadas incluíram *flip* horizontal com probabilidade de 50% e pequenas rotações de até 5°. A probabilidade de aplicação do ADA foi ajustada dinamicamente em função do desempenho do discriminador, visando uma meta de 60%.

O pré-processamento também considerou a preparação eficiente para a GPU, com carregamento de dados em lotes de 8 imagens, utilizando duas *threads* e memória pinada quando disponível. No total, o *dataset* consistiu em todas as imagens presentes no CSV fornecido, organizadas de maneira a garantir equilíbrio entre as classes e uniformidade na entrada da rede.

3.3.3 Treinamento do modelo de geração

O modelo do gerador foi inicializado com pesos distribuídos normalmente ($\mu = 0$, $\sigma = 1$), enquanto o discriminador recebeu configuração similar. O vetor latente $z \in \mathbb{R}^{256}$ foi combinado com os vetores *one-hot* das classes (benigno ou maligno) para formar a entrada condicionada do gerador. A rede mapeadora (*Mapping Network*) processou z para gerar o vetor $A \in \mathbb{R}^{256}$, que modulou as camadas do gerador através de *Adaptive Instance Normalization (AdaIN)*. Cada bloco do gerador aplicou convoluções, ruído estocástico e funções de ativação Leaky ReLU, permitindo a síntese progressiva de imagens de maior resolução.

O treinamento seguiu o esquema de otimização adversarial com *Adam* para o gerador e o discriminador, usando taxas de aprendizado de 0.0025 e parâmetros $\beta_1 = 0.0$ e $\beta_2 = 0.99$. As funções de perda empregadas foram a *logistic loss* para ambos os modelos e a penalidade R1 aplicada ao discriminador a cada 32 iterações. Além disso, o treinamento utilizou precisão mista (*automatic mixed precision*) para reduzir consumo de memória e acelerar o processamento.

O modelo também incorporou a estratégia de aumento de dados adaptativa (*ADA*), ajustando dinamicamente a probabilidade de aplicar transformações como *flip* horizontal e rotações pequenas, visando manter uma meta de 60% de eficácia no discriminador. Para monitoramento, *checkpoints* foram salvos, incluindo pesos do gerador e discriminador, otimizadores e histórico das perdas, enquanto amostras geradas fixas eram armazenadas para visualização da evolução do treinamento.

O treinamento foi conduzido por 400 épocas, com cada época percorrendo todo o *dataset*, garantindo que tanto o gerador quanto o discriminador fossem atualizados em todas as iterações. Essa configuração permitiu que o gerador aprendesse a produzir imagens médicas de alta fidelidade condicionadas à classe, minimizando sobreajuste e mantendo diversidade visual.

3.3.4 Uso do LPIPS para avaliação de qualidade das imagens geradas

Diversas métricas foram avaliadas na literatura para mensurar a qualidade de imagens geradas por redes neurais, dentre as quais destacam-se o PSNR (*Peak Signal-to-Noise Ratio*), o SSIM (*Structural Similarity Index*) e o LPIPS (*Learned Perceptual Image Patch Similarity*). Enquanto o PSNR e o SSIM baseiam-se em diferenças de intensidade e estrutura de pixels, o LPIPS utiliza representações extraídas de redes neurais profundas, permitindo uma avaliação perceptual mais próxima da visão humana.

Pawar et al. [78] mostraram que modelos de geração de imagens com aparência mais realista, como o SRGAN, obtiveram baixos valores de PSNR e SSIM, mas se destacaram significativamente em LPIPS. Isso indica que as métricas tradicionais nem sempre refletem a qualidade visual percebida, especialmente quando o objetivo é gerar dados sintéticos visualmente convincentes.

Dessa forma, neste trabalho optou-se pelo uso do LPIPS como métrica principal de avaliação, por apresentar maior correlação com a percepção humana e melhor sensibilidade a diferenças perceptuais sutis, mostrando-se, portanto, adequado para mensurar a fidelidade visual e o realismo das imagens sintéticas geradas. A aplicação dessa métrica permitiu avaliar a qualidade perceptual das mamografias produzidas em comparação com as imagens reais, conforme ilustrado na Figura 23, que apresenta amostras benignas, e na Figura 24, que exibe amostras malignas, em ambos os casos, a imagem (a) corresponde à amostra real e a imagem (b) à amostra sintética.

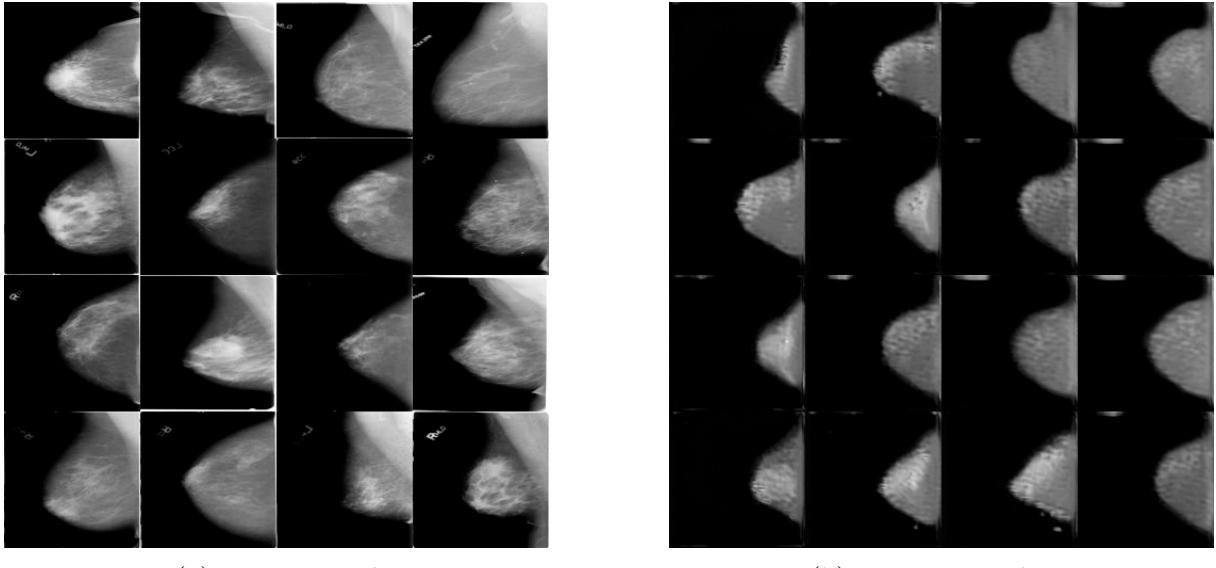


Figura 23 – Mamografias benignas: (a) real e (b) sintética.

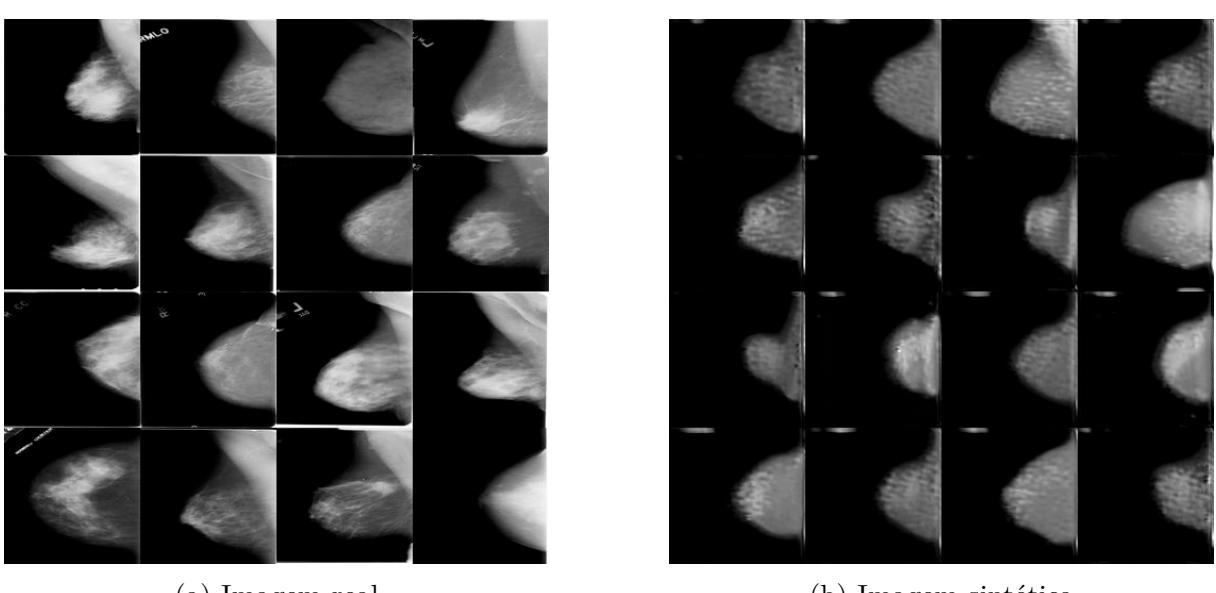


Figura 24 – Mamografias malignas: (a) real e (b) sintética.

Para compor o conjunto de dados sintéticos, foram selecionadas apenas as imagens cuja métrica LPIPS apresentou valor inferior a 0.2 quando comparadas ao conjunto dos dados reais, garantindo alta similaridade perceptiva entre as amostras reais e geradas. Esse limiar foi adotado como critério de qualidade, de modo a incluir apenas imagens sintéticas com fidelidade visual e coerência estrutural adequadas.

As imagens que atenderam a esse critério foram então utilizadas no treinamento da rede EfficientNet-B0, com o objetivo de aumentar a diversidade do conjunto de dados e melhorar a capacidade de generalização do modelo, sem comprometer a qualidade perceptual das amostras utilizadas.

3.4 Modelo de classificação

3.4.1 Arquitetura do modelo

Para a tarefa de classificação de mamografias, foi utilizada a *EfficientNet-B0*. Optou-se pela utilização desse modelo, devido ao seu melhor equilíbrio entre acurácia, eficiência computacional e capacidade de generalização.

O modelo foi inicializado com pesos pré-treinados no conjunto de dados *ImageNet*, permitindo o aproveitamento de representações visuais previamente aprendidas. A camada final da rede (*classifier*) foi substituída por uma camada totalmente conectada com um único neurônio de saída e função de ativação sigmoide implícita, adequada para a classificação binária entre amostras benignas e malignas. Para acelerar o treinamento e reduzir o risco de sobreajuste, os parâmetros da base convolucional (*features*) foram inicialmente congelados.

As imagens de entrada foram redimensionadas para 128×128 pixels e normalizadas de acordo com as médias e desvios padrão utilizados no treinamento do *ImageNet*. A rede foi treinada utilizando o otimizador *Adam*, com taxa de aprendizado inicial de 1×10^{-3} e função de perda binária *Binary Cross-Entropy with Logits (BCEWithLogitsLoss)*. O processo de validação cruzada foi realizado em 5 dobras (*5-fold Cross-Validation*), garantindo uma avaliação robusta do desempenho do modelo.

A escolha da *EfficientNet-B0* mostrou-se adequada por apresentar maior estabilidade durante o treinamento e resultados superiores nas métricas de acurácia, F1-Score e AUC. Além disso, a arquitetura é otimizada para eficiência em GPUs modernas, permitindo melhor aproveitamento dos recursos computacionais sem comprometer a qualidade da classificação das imagens mamográficas.

3.4.2 Seleção das imagens sintéticas para o modelo de classificação

Após o treinamento do modelo de geração, foi realizado um processo de seleção das imagens sintéticas a serem utilizadas no treinamento do modelo de classificação. O objetivo dessa etapa foi garantir que apenas amostras sintéticas com alta qualidade perceptual e diversidade fossem incluídas, reduzindo o risco de introduzir artefatos ou imagens redundantes no conjunto de dados.

Para isso, utilizou-se o LPIPS, o cálculo foi feito entre cada imagem sintética gerada e um subconjunto de imagens reais de mesma classe (benigna ou maligna).

Cada imagem gerada foi comparada com as amostras reais, e o valor mínimo de LPIPS obtido foi utilizado como medida de similaridade. Apenas imagens cuja menor distância perceptual fosse inferior a um limite pré-estabelecido ($\text{LPIPS} < 0.2$) foram mantidas. Esse limiar foi definido empiricamente após testes iniciais, equilibrando diversidade

e realismo visual das amostras selecionadas.

As imagens que atenderam ao critério foram salvas em diretórios separados por classe e posteriormente integradas ao conjunto real durante o treinamento supervisionado do classificador. Esse procedimento permitiu aumentar o volume de dados de forma controlada, mantendo a coerência estatística entre os domínios real e sintético.

3.4.3 Divisão entre dados reais e sintéticos

Após a geração das imagens sintéticas pelo modelo *StyleGAN2-ADA*, foi realizada a definição das proporções entre dados reais e sintéticos que seriam utilizadas no treinamento do classificador *EfficientNet-B0*. O objetivo dessa etapa foi investigar o impacto da incorporação de imagens artificiais no desempenho do modelo de classificação, buscando um equilíbrio entre a diversidade introduzida pelos dados gerados e a fidelidade das amostras reais.

Para isso, diferentes proporções de dados foram consideradas, variando desde o uso exclusivo de imagens reais até combinações em que o número de amostras sintéticas superava o de amostras reais. As razões testadas incluíram 1:1, 2:1, 3:1 e 4:1 entre imagens sintéticas e reais, mantendo o mesmo número total de amostras por classe em cada configuração. Todas as divisões foram realizadas de forma estratificada, assegurando que as classes benigna e maligna estivessem balanceadas em cada conjunto de treinamento e validação.

A etapa de divisão entre dados reais e sintéticos teve papel definindo a base comparativa para a análise do impacto do uso de dados gerados artificialmente no desempenho da rede *EfficientNet-B0*.

3.4.4 Pré-processamento de imagens reais e sintéticas

Antes do treinamento do modelo de classificação *EfficientNet-B0*, foi necessário realizar o pré-processamento das imagens reais e sintéticas de modo a garantir compatibilidade com a arquitetura e consistência entre as amostras. Todas as imagens foram redimensionadas para 128×128 pixels, mantendo a resolução utilizada na etapa de geração de imagens sintéticas, o que favoreceu a padronização do fluxo de dados entre os modelos. Diferentemente do modelo gerador, as imagens foram convertidas para o formato RGB, totalizando três canais de entrada, conforme exigido pelos pesos pré-treinados no conjunto *ImageNet*.

As transformações de pré-processamento incluíram a normalização dos valores de pixel utilizando a média [0.485, 0.456, 0.406] e o desvio padrão [0.229, 0.224, 0.225], parâmetros padrão do *ImageNet*. Essa normalização foi fundamental para estabilizar o treinamento e manter a coerência estatística entre o domínio original e o domínio médico das

mamografias. Durante o treinamento, foram aplicadas técnicas de aumento de dados (*data augmentation*) com o objetivo de reduzir o sobreajuste (*overfitting*) e aumentar a robustez do modelo frente à variação visual. Entre as transformações aplicadas, destacam-se o *flip* horizontal aleatório e rotações de até 5°, aplicadas de forma independente a cada amostra.

A etapa de carregamento das imagens foi implementada com o *DataLoader* do PyTorch, utilizando lotes de 32 amostras e duas threads para leitura paralela de dados. Sempre que disponível, foi empregada a memória pinada para otimizar a transferência entre CPU e GPU. Esse procedimento garantiu maior eficiência no treinamento, aproveitando os recursos da GPU.

Por fim, as imagens foram organizadas em estruturas balanceadas de classes, correspondentes aos rótulos binários 0 (benigna) e 1 (maligna), conforme definido nos arquivos CSV de metadados. O conjunto de dados foi dividido de forma estratificada entre treino e validação dentro de cada *fold* da validação cruzada, assegurando que a proporção entre as classes permanecesse constante e que o modelo recebesse uma representação equilibrada das categorias em todas as iterações de treinamento, Figura 25.

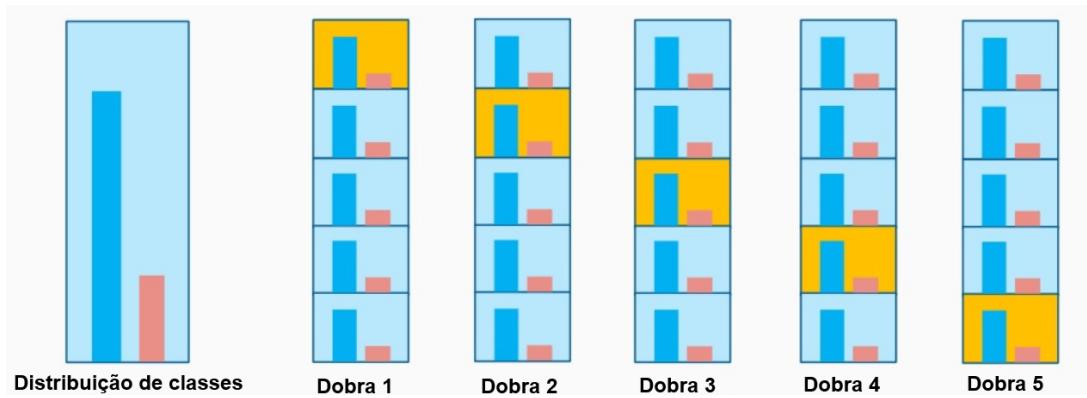


Figura 25 – Esquema de validação cruzada estratificada de 5 dobras.

3.4.5 Treinamento do modelo de classificação

O conjunto de dados real foi submetido a uma etapa de *downsampling* das amostras benignas, a fim de reduzir o desbalanceamento entre as classes. Especificamente, o número de imagens benignas foi limitado a 245 amostras, mantendo todas as instâncias malignas disponíveis. Essa redução foi realizada de forma aleatória, com *seed* fixa para garantir reproduzibilidade, ou seja, assegurar que os mesmos dados sejam selecionados em execuções futuras, mantendo a consistência experimental. Em seguida, as imagens sintéticas geradas pelo modelo *StyleGAN2-ADA* foram adicionadas apenas ao conjunto de treinamento, preservando o conjunto de validação composto exclusivamente por imagens reais.

Durante o treinamento, a rede *EfficientNet-B0* foi inicializada com pesos pré-treinados no *ImageNet*, e apenas a camada de classificação foi substituída por um neurônio

de saída com ativação sigmoide, apropriada para o problema binário (benigno/maligno). Inicialmente, as camadas convolucionais da base foram congeladas para preservar as representações gerais de baixo nível e acelerar a convergência. O treinamento foi conduzido com o otimizador *Adam*, taxa de aprendizado de 1×10^{-3} e função de perda *Binary Cross-Entropy with Logits (BCEWithLogitsLoss)*.

Cada época (*epoch*), que corresponde a uma passagem completa por todo o conjunto de treinamento, processou lotes de 32 imagens com aumento de dados aplicado em tempo real. As transformações incluíram rotações aleatórias de até 5° e inversão horizontal, contribuindo para maior robustez do modelo frente à variação posicional das estruturas mamárias. O carregamento dos dados foi realizado com duas *threads* paralelas e memória pinada (*pinned memory*), técnica que fixa os dados na RAM para otimizar a transferência entre CPU e GPU, reduzindo a latência no carregamento dos lotes durante o treinamento.

O treinamento foi executado por 30 épocas em cada *fold*, com monitoramento contínuo das métricas de acurácia, F1-score e AUC. Ao final de cada época, o modelo era avaliado no conjunto de validação, sendo armazenados os melhores pesos com base na acurácia obtida. Esse procedimento foi repetido para todos os *folds*, resultando em cinco modelos independentes e métricas médias representativas da performance geral. Toda a execução foi realizada em uma GPU, utilizando precisão mista (*mixed precision*) para reduzir o consumo de memória e acelerar o processamento.

3.4.6 Métricas de classificação

Para avaliar o desempenho do modelo *EfficientNet-B0* no problema de detecção de lesões mamográficas, foram empregadas métricas consolidadas na literatura de aprendizado de máquina para tarefas de classificação binária: acurácia, F1-score e área sob a curva ROC (AUC).

A acurácia mede a proporção total de previsões corretas em relação ao número total de amostras avaliadas, e é uma métrica global definida pela equação (3.1):

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3.1)$$

onde TP (verdadeiros positivos) representa o número de lesões malignas corretamente classificadas, TN (verdadeiros negativos) o número de amostras benignas corretamente identificadas, FP (falsos positivos) o número de amostras benignas classificadas incorretamente como malignas e FN (falsos negativos) o número de amostras malignas preditas como benignas.

O F1-score, por sua vez, é a média harmônica entre a precisão (*precision*) e a revocação (*recall*), sendo especialmente útil em cenários de classes desbalanceadas, como

é comum em bases médicas. Suas definições são dadas pela equação (3.2):

$$\text{Precisão} = \frac{TP}{TP + FP}, \quad \text{Revocação} = \frac{TP}{TP + FN}, \quad (3.2)$$

e o F1-score é calculado pela equação (3.3):

$$\text{F1-score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}. \quad (3.3)$$

Essa métrica, equação (3.3), penaliza fortemente classificações incorretas em qualquer uma das classes, equilibrando a capacidade do modelo de detectar casos positivos (revocação) e evitar alarmes falsos (precisão).

Já a área sob a curva ROC (AUC — *Area Under the Curve*) quantifica a capacidade discriminatória do modelo ao variar o limiar de decisão. A curva ROC (*Receiver Operating Characteristic*) é construída a partir da relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR), definidas pela equação (3.4):

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}. \quad (3.4)$$

A métrica AUC é então calculada como a integral da curva ROC, conforme mostrado na equação (3.5):

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}). \quad (3.5)$$

Essa integral representa a área sob a curva ROC, assumindo valores entre 0 e 1. Um modelo com $\text{AUC} = 0.5$ equivale a uma classificação aleatória, enquanto valores próximos de 1 indicam excelente separabilidade entre as classes.

As métricas foram calculadas individualmente para cada *fold* e posteriormente agregadas por média e desvio padrão, conforme a metodologia de validação cruzada implementada. Esses valores médios forneceram uma estimativa estável da performance geral do modelo.

4 RESULTADOS

Este capítulo apresenta e discute os resultados obtidos a partir dos experimentos realizados com os modelos de geração e classificação de imagens de mamografia. Inicialmente, são descritos os resultados referentes à geração de imagens sintéticas utilizando o modelo StyleGAN2-ADA, com análise baseada na métrica perceptual LPIPS para avaliar a qualidade e a similaridade das amostras geradas em relação às imagens reais.

Em seguida, são apresentados os resultados da classificação de mamografias empregando o modelo EfficientNet-B0, incluindo a avaliação do impacto da incorporação de imagens sintéticas no desempenho do classificador. As métricas de acurácia, F1-score e AUC são analisadas de forma comparativa, buscando identificar a proporção ideal entre dados reais e sintéticos que maximize a performance do modelo. Por fim, são discutidas as implicações dos resultados obtidos, bem como as limitações e observações relevantes quanto à viabilidade computacional dos experimentos realizados.

4.1 Geração de imagens de mamografia

O gerador StyleGAN2-ADA foi treinado com um conjunto relativamente reduzido de 553 imagens, sendo 245 classificadas como malignas e 308 como benignas. Essa limitação de amostras representa um desafio relevante, principalmente pela disparidade entre as classes, que pode afetar a capacidade do modelo de capturar adequadamente as variações presentes nas imagens malignas.

A análise da distribuição dos valores LPIPS, Figura 26, na geração de 1000 imagens para classe benigna e 1000 para maligna, mostra que a maioria das imagens geradas apresenta alta similaridade perceptual em relação às imagens reais. A distribuição geral apresenta média de 0.247 e mediana de 0.249, com concentração em torno de 0.25, indicando que o gerador manteve um padrão de qualidade estável ao longo das amostras. A proximidade entre média e mediana sugere uma distribuição simétrica, sem a presença de valores extremos significativos.

Ao observar a distribuição por classe, Figura 26b, nota-se que as imagens malignas tendem a apresentar valores de LPIPS ligeiramente mais altos em comparação às benignas. Esse comportamento pode estar diretamente relacionado à menor quantidade de exemplos malignos disponíveis no treinamento, o que limita a capacidade do modelo em generalizar os padrões dessa classe. Como consequência, o gerador produz imagens malignas com maior variabilidade estrutural e, portanto, diferenças perceptuais um pouco mais pronunciadas em relação às imagens reais. Essa tendência é corroborada pela curva de densidade de probabilidade, onde as imagens malignas exibem distribuição mais ampla,

indicando maior dispersão dos valores LPIPS.

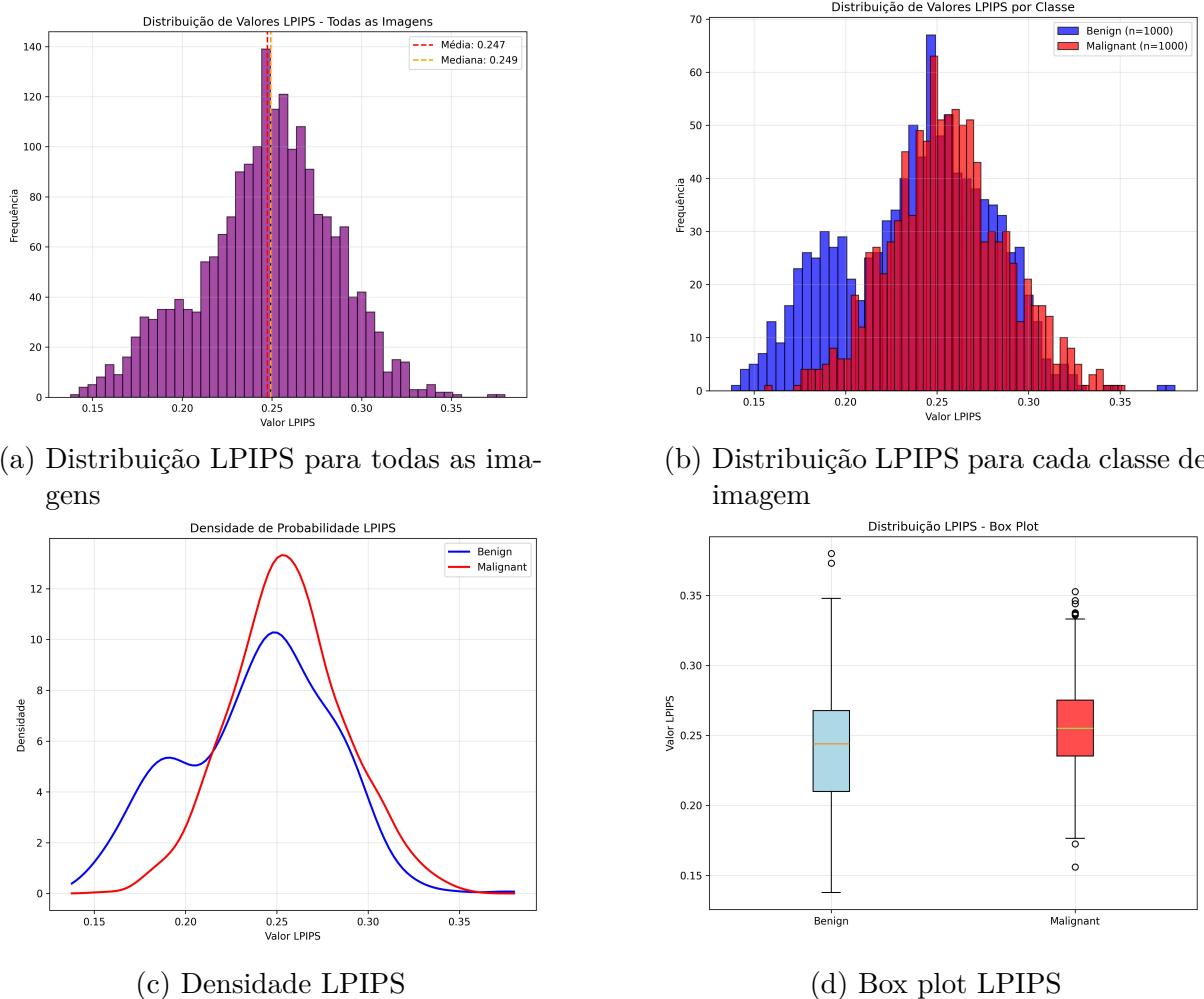


Figura 26 – Gráficos dos valores do LPIPS na geração de 1000 imagens para cada classe usando StyleGAN2-ADA. Em laranja estão representadas as classes malignas e em azul as benignas.

Mesmo com essa diferença, a sobreposição significativa entre as distribuições das duas classes e a proximidade das medianas no box plot, Figura 26d, demonstram que o modelo manteve uma qualidade perceptual consistente entre as categorias. A presença de poucos outliers e a amplitude interquartílica semelhante reforçam a estabilidade do processo de geração, apesar do número reduzido de imagens reais utilizadas para o treinamento.

4.2 Classificação de mamografias

Os resultados da avaliação perceptual indicam que o modelo gerativo foi capaz de reproduzir de forma consistente a distribuição visual das mamografias reais, gerando amostras sintéticas com elevada fidelidade e realismo. Essa coerência visual sugere

que as imagens produzidas preservam características estruturais relevantes das mamografias reais, tornando viável sua utilização como complemento no treinamento de modelos de classificação. Assim, procedeu-se à etapa de avaliação quantitativa, na qual se investigou o impacto da inclusão dessas imagens sintéticas no desempenho do classificador *EfficientNet-B0*.

Após a seleção das imagens sintéticas ($LPIPS < 0.2$), o impacto da sua utilização no treinamento do modelo *EfficientNet-B0* foi avaliado por meio de validação cruzada estratificada de cinco dobras. Esse procedimento garantiu que as proporções entre as classes benignas e malignas fossem preservadas em todas as partições, permitindo uma avaliação equilibrada do desempenho do classificador. Em cada iteração, o modelo foi treinado com a combinação de imagens reais e sintéticas, enquanto a avaliação foi realizada exclusivamente sobre imagens reais, assegurando que a mensuração de desempenho refletisse a capacidade de generalização do modelo.

Os resultados obtidos utilizando apenas imagens reais revelaram um desempenho inicial moderado, com **acurácia média de 0.606, F1-score médio de 0.573 e AUC de 0.665**. O desempenho é condizente com o tamanho reduzido e o desbalanceamento do conjunto original, composto por 553 imagens (308 benignas e 245 malignas). A partir da inclusão de imagens sintéticas geradas pelo *StyleGAN2-ADA*, observou-se uma tendência de melhoria nas métricas de desempenho, especialmente nas proporções mais equilibradas entre dados reais e sintéticos.

Tabela 2 – Resultados da validação cruzada (EfficientNet-B0).

Proporção	Acurácia Média ± DP	F1-Score Médio ± DP	AUC Média ± DP
Apenas reais	0.606 ± 0.035	0.573 ± 0.056	0.665 ± 0.034
1:1	0.643 ± 0.038	0.630 ± 0.046	0.703 ± 0.039
2:1	0.651 ± 0.030	0.658 ± 0.017	0.703 ± 0.029
3:1	0.618 ± 0.033	0.595 ± 0.065	0.667 ± 0.041
4:1	0.616 ± 0.037	0.583 ± 0.043	0.679 ± 0.061

A Tabela 2 mostra que o uso de imagens sintéticas foi benéfico até a proporção **2:1**, na qual se observaram os maiores valores de acurácia, F1-score e AUC. O aumento na proporção 1:1 já resultou em uma melhora perceptível, indicando que a inclusão de dados gerados pelo *StyleGAN2-ADA* contribuiu para ampliar a variabilidade visual do conjunto de treinamento, permitindo ao modelo aprender representações mais robustas das características mamográficas. O ganho mais expressivo foi observado na proporção 2:1, com um incremento de aproximadamente 7,4% em acurácia e 14,8% em F1-score em relação ao cenário com apenas dados reais.

A partir da proporção 3:1, contudo, nota-se um declínio no desempenho do classificador. Esse comportamento sugere que o aumento excessivo de amostras sintéticas pode gerar redundância ou introduzir ruído perceptual, desviando o modelo das distribuições estatísticas originais das imagens reais. Assim, embora as imagens geradas apresentem boa fidelidade visual conforme demonstrado pela análise de LPIPS, seu uso em excesso tende a reduzir a especificidade do aprendizado, impactando negativamente a acurácia final.

Além disso, observa-se que os desvios padrão permanecem baixos em todas as proporções, o que indica consistência e estabilidade nos resultados obtidos nas diferentes dobras de validação. Essa estabilidade demonstra que o comportamento do modelo é reproduzível, sem dependência de partições específicas do conjunto de dados.

Em síntese, os resultados do EfficientNet-B0 demonstram que a inclusão moderada de imagens sintéticas contribui de forma significativa para o aprimoramento do desempenho do classificador. O cenário de melhor desempenho foi alcançado na proporção **2:1 (sintético:real)**, na qual o modelo apresentou simultaneamente maior acurácia, equilíbrio entre precisão e sensibilidade (F1-score) e maior área sob a curva ROC (AUC).

Para garantir a viabilidade de treinamento e inferência, foram adotadas escolhas de arquitetura que equilibram complexidade e desempenho, como a utilização do **StyleGAN2-ADA condicional** com parâmetros reduzidos e do **EfficientNet-B0** como classificador, ambos reconhecidos por sua boa performance em sistemas com recursos limitados. A inclusão de imagens sintéticas foi planejada levando em consideração o tamanho reduzido do conjunto de dados real, com o objetivo de aumentar a diversidade amostral sem sobreregar a memória ou causar saturação do processo de treinamento.

5 CONCLUSÃO

A limitação no acesso a grandes bases de imagens médicas rotuladas representa um dos principais desafios para o treinamento eficaz de modelos de aprendizado profundo aplicados à detecção de lesões em mamografias. Diante desse problema, este trabalho teve como objetivo investigar o uso de imagens sintéticas geradas por modelos generativos para ampliar conjuntos de dados reais e aprimorar o desempenho de classificadores baseados em redes neurais convolucionais.

Para atingir esse objetivo, foi proposta uma estratégia que combinou o uso do modelo StyleGAN2-ADA na geração de imagens de mamografia com a aplicação da métrica LPIPS como filtro de similaridade perceptual, garantindo maior fidelidade entre as amostras sintéticas e as reais. As imagens geradas foram então integradas ao treinamento do classificador EfficientNet-B0, permitindo avaliar o impacto das proporções entre dados sintéticos e reais no desempenho do modelo.

Os resultados demonstraram que a incorporação controlada de imagens sintéticas contribuiu de forma significativa para a melhoria das métricas de acurácia, F1-score e AUC-ROC, especialmente nas proporções de 1:1 e 2:1 entre imagens reais e geradas. Esses achados indicam que a geração sintética de dados é uma alternativa para mitigar o problema da escassez de amostras, ampliando a capacidade do modelo de generalizar padrões relevantes das classes benignas e malignas. Contudo, verificou-se que o aumento excessivo de dados artificiais (proporção 3:1) resultou em leve queda de desempenho, possivelmente devido à introdução de redundâncias ou inconsistências perceptuais no conjunto de treinamento.

A execução dos experimentos também demonstrou a viabilidade computacional da proposta. Mesmo com uma infraestrutura moderada, GPU RTX 4060 de 8 GB, processador Intel Core i5-12400F e 16 GB de RAM, foi possível realizar o treinamento completo dos modelos dentro de prazos adequados, sem prejuízo ao desempenho. Isso reforça a aplicabilidade da abordagem em contextos acadêmicos e institucionais com recursos limitados.

Como dificuldades enfrentadas, destacam-se o tempo elevado de treinamento do modelo gerador, a necessidade de ajustes finos nos hiperparâmetros para evitar artefatos visuais, a limitação do conjunto original que restringiu a diversidade das amostras e a complexidade na escolha de arquiteturas adequadas para o trabalho, que exigiu um equilíbrio entre desempenho e viabilidade computacional. Para trabalhos futuros, sugere-se a ampliação da base de dados com imagens provenientes de diferentes fontes e equipamentos, além da exploração de arquiteturas mais recentes e eficientes.

REFERÊNCIAS

- [1] NASTESKI, V. An overview of the supervised machine learning methods. *Horizons*, b, v. 4, n. 51-62, p. 56, 2017.
- [2] WU, Y.-c.; FENG, J.-w. Development and application of artificial neural network. *Wireless Personal Communications*, Springer, v. 102, p. 1645–1656, 2018.
- [3] DING, B.; QIAN, H.; ZHOU, J. Activation functions and their characteristics in deep neural networks. In: IEEE. *2018 Chinese control and decision conference (CCDC)*. [S.l.], 2018. p. 1836–1841.
- [4] MASSAOUDI, A. et al. Machine learning models for modeling the biosorption of fe (iii) ions by activated carbon from olive stone. *Neural Computing and Applications*, Springer, v. 36, n. 21, p. 13357–13372, 2024.
- [5] RGUIBI, Z. et al. Cxai: Explaining convolutional neural networks for medical imaging diagnostic. *Electronics*, MDPI, v. 11, n. 11, p. 1775, 2022.
- [6] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.
- [7] RAKSHIT, R. D. et al. Cross-resolution face identification using deep-convolutional neural network. *Multimedia Tools and Applications*, Springer, v. 80, p. 20733–20758, 2021.
- [8] AHMED, T.; SABAB, N. H. N. Classification and understanding of cloud structures via satellite images with efficientunet. *SN Computer Science*, Springer, v. 3, n. 1, p. 99, 2022.
- [9] PATIL, M.; PATIL, M. M.; AGRAWAL, S. Wgan for data augmentation. In: *GANs for Data Augmentation in Healthcare*. [S.l.]: Springer, 2023. p. 223–241.
- [10] LI, Y.; DONG, B.; QIU, Y. Conditional generative adversarial network (cgan) for generating building load profiles with photovoltaics and electric vehicles. *Energy and Buildings*, Elsevier, v. 335, p. 115584, 2025.
- [11] FENG, Y. et al. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes & diseases*, Elsevier, v. 5, n. 2, p. 77–106, 2018.
- [12] PRECOCE, D. Controle do câncer de mama-documento de consenso. *Rev Bras Cancerol*, v. 50, n. 2, p. 77–90, 2004.
- [13] AZEVEDO, R. L.; GERÓTICA, R. M. G.; SANCHES, T. P. A importância da mamografia no diagnóstico precoce do câncer de mama. *UNILUS Ensino e Pesquisa*, v. 13, n. 30, p. 251, 2016.
- [14] OBUCHOWICZ, R.; OSZUST, M.; PIORKOWSKI, A. Interobserver variability in quality assessment of magnetic resonance images. *BMC Medical Imaging*, Springer, v. 20, n. 1, p. 109, 2020.

- [15] ERICKSON, B. J. Machine learning: discovering the future of medical imaging. *Journal of digital imaging*, Springer, v. 30, n. 4, p. 391–391, 2017.
- [16] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015.
- [17] MENDEL, K. et al. Transfer learning from convolutional neural networks for computer-aided diagnosis: a comparison of digital breast tomosynthesis and full-field digital mammography. *Academic radiology*, Elsevier, v. 26, n. 6, p. 735–743, 2019.
- [18] GONZALEZ, R. C. *Digital image processing*. [S.l.]: Pearson education india, 2009.
- [19] SZELISKI, R. *Computer vision: algorithms and applications*. [S.l.]: Springer Nature, 2022.
- [20] LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Springer, v. 60, p. 91–110, 2004.
- [21] BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. In: SPRINGER. *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006. Proceedings, Part I* 9. [S.l.], 2006. p. 404–417.
- [22] LI, S. Z. *Markov random field modeling in image analysis*. [S.l.]: Springer Science & Business Media, 2009.
- [23] KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, 2012.
- [24] JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, Ieee, v. 22, n. 1, p. 4–37, 2000.
- [25] JANAI, J. et al. Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and trends® in computer graphics and vision*, Now Publishers, Inc., v. 12, n. 1–3, p. 1–308, 2020.
- [26] KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. *Computers and electronics in agriculture*, Elsevier, v. 147, p. 70–90, 2018.
- [27] LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017.
- [28] GREENSPAN, H.; GINNEKEN, B. V.; SUMMERS, R. M. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1153–1159, 2016.
- [29] SHEN, D.; WU, G.; SUK, H.-I. Deep learning in medical image analysis. *Annual review of biomedical engineering*, Annual Reviews, v. 19, n. 1, p. 221–248, 2017.
- [30] HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

- [31] MITCHELL, T. M.; MITCHELL, T. M. *Machine learning*. [S.l.]: McGraw-hill New York, 1997. v. 1.
- [32] ALPAYDIN, E. *Introduction to machine learning*. [S.l.]: MIT press, 2020.
- [33] BISHOP, C. M.; NASRABADI, N. M. *Pattern recognition and machine learning*. [S.l.]: Springer, 2006. v. 4.
- [34] GÉRON, A. Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts. *Aurélien Géron-Google Kitaplar*, yy <https://books.google.com.tr/books>, 2019.
- [35] HASTIE, T. et al. *The elements of statistical learning*. [S.l.]: Citeseer, 2009.
- [36] REDDY, C. K. *Data clustering: algorithms and applications*. [S.l.]: Chapman and Hall/CRC, 2018.
- [37] TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to data mining*. [S.l.]: Pearson Education India, 2016.
- [38] HAYKIN, S. *Neural Networks and Learning Machines*. 3. ed. [S.l.]: Pearson Education, 2009.
- [39] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.
- [40] ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, v. 65, n. 6, p. 386–408, 1958.
- [41] AGGARWAL, C. C. *Neural Networks and Deep Learning: A Textbook*. [S.l.]: Springer, 2018.
- [42] GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*. [S.l.: s.n.], 2010. p. 249–256.
- [43] NWANKPA, C. et al. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018.
- [44] LECUN, Y. et al. Efficient backprop. In: *Neural Networks: Tricks of the Trade*. [S.l.]: Springer, 2012. p. 9–48.
- [45] RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *Nature*, v. 323, p. 533–536, 1986.
- [46] LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, 2015.
- [47] BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006.
- [48] LI, Y.; ZHANG, L.; HE, K. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, v. 33, n. 10, p. 4943–4964, 2021.

- [49] NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. p. 807–814, 2010.
- [50] SCHERER, D.; MÜLLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. In: SPRINGER. *International Conference on Artificial Neural Networks*. [S.l.], 2010. p. 92–101.
- [51] TAN, M.; LE, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: PMLR. *International conference on machine learning*. [S.l.], 2019. p. 6105–6114.
- [52] GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. v. 27, p. 2672–2680.
- [53] CRESWELL, A. et al. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, IEEE, v. 35, n. 1, p. 53–65, 2018.
- [54] GOODFELLOW, I. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.
- [55] RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.
- [56] KARRAS, T.; LAINE, S.; AILA, T. A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 4401–4410.
- [57] LEDIG, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4681–4690.
- [58] ZHU, J.-Y. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2017. p. 2223–2232.
- [59] ELGAMMAL, A. et al. Can: Creative adversarial networks, generating art by learning about styles and deviating from style norms. In: *ICCC*. [S.l.: s.n.], 2017. p. 96–103.
- [60] FRID-ADAR, M. et al. Synthetic data augmentation using gan for improved liver lesion classification. In: IEEE. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. [S.l.], 2018. p. 289–293.
- [61] ARJOVSKY, M.; BOTTOU, L. Towards principled methods for training generative adversarial networks. In: *International Conference on Learning Representations (ICLR)*. [S.l.: s.n.], 2017.
- [62] SALIMANS, T. et al. Improved techniques for training gans. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2016. p. 2234–2242.
- [63] GULRAJANI, I. et al. Improved training of wasserstein gans. *Advances in neural information processing systems*, v. 30, 2017.

- [64] ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein generative adversarial networks. In: PMLR. *International conference on machine learning*. [S.l.], 2017. p. 214–223.
- [65] MIRZA, M. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- [66] KARRAS, T. et al. Training generative adversarial networks with limited data. *Advances in neural information processing systems*, v. 33, p. 12104–12114, 2020.
- [67] HORE, A.; ZIOU, D. Image quality metrics: Psnr vs. ssim. *2010 20th International Conference on Pattern Recognition*, p. 2366–2369, 2010.
- [68] WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.
- [69] ZHANG, R. et al. The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 586–595.
- [70] DESAI, S. D. et al. Breast cancer detection using gan for limited labeled dataset. In: IEEE. *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*. [S.l.], 2020. p. 34–39.
- [71] JOSEPH, A. J. et al. Prior-guided generative adversarial network for mammogram synthesis. *Biomedical Signal Processing and Control*, Elsevier, v. 87, p. 105456, 2024.
- [72] GUAN, S.; LOEW, M. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *Journal of Medical Imaging*, Society of Photo-Optical Instrumentation Engineers, v. 6, n. 3, p. 031411–031411, 2019.
- [73] SHAH, D.; KHAN, M. A. U.; ABRAR, M. Reliable breast cancer diagnosis with deep learning: Dcgan-driven mammogram synthesis and validity assessment. *Applied Computational Intelligence and Soft Computing*, Wiley Online Library, v. 2024, n. 1, p. 1122109, 2024.
- [74] JIMÉNEZ-GAONA, Y. et al. Gan-based data augmentation to improve breast ultrasound and mammography mass classification. *Biomedical Signal Processing and Control*, Elsevier, v. 94, p. 106255, 2024.
- [75] 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.
- [76] SHEN, T. et al. Mass image synthesis in mammogram with contextual information based on gans. *Computer Methods and Programs in Biomedicine*, Elsevier, v. 202, p. 106019, 2021.

- [77] AWSAF49. *CBIS-DDSM - Breast Cancer Image Dataset*. 2021. <<https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset/data>>. Accessed: 2025-11-03.
- [78] PAWAR, N. M. et al. Esm data downscaling: a comparison of super-resolution deep learning models. *Earth Science Informatics*, Springer, v. 17, n. 4, p. 3511–3528, 2024.