



ARTHUR ALEXSANDER MARTINS TEODORO

**UMA ABORDAGEM DE CLASSIFICAÇÃO DE CÂNCER DE
PELE USANDO GAN E MECANISMO DE ATENÇÃO
BASEADO EM ROI**

LAVRAS – MG

2021

ARTHUR ALEXSANDER MARTINS TEODORO

**UMA ABORDAGEM DE CLASSIFICAÇÃO DE CÂNCER DE PELE USANDO GAN E
MECANISMO DE ATENÇÃO BASEADO EM ROI**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação para obtenção do título de Mestre.

Prof. DSc. Demóstenes Zegarra Rodríguez
Orientador

Prof. DSc. Bruno de Abreu Silva
Coorientador

**LAVRAS – MG
2021**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Teodoro, Arthur Alexsander Martins Teodoro

Uma Abordagem de Classificação de Câncer de Pele
usando GAN e Mecanismo de Atenção baseado em RoI /
Arthur Alexsander Martins Teodoro. – 2021.

82 p. : il.

Orientador(a): Prof. DSc. Demóstenes Zegarra Rodríguez.

Coorientador(a): Prof. DSc. Bruno de Abreu Silva.

Dissertação (mestrado acadêmico)– Universidade Federal
de Lavras, 2021.

Bibliografia.

1. Redes Neurais Convolucionais. 2. CNN. 3. Câncer de
Pele. I. Rodríguez, Demóstenes Zegarra. II. Silva, Bruno de
Abreu. III. Título.

ARTHUR ALEXSANDER MARTINS TEODORO

**UMA ABORDAGEM DE CLASSIFICAÇÃO DE CÂNCER DE PELE USANDO GAN E
MECANISMO DE ATENÇÃO BASEADO EM ROI**

**A SKIN CANCER CLASSIFICATION APPROACH USING GAN AND ROI-BASED
ATTENTION MECHANISM**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação para obtenção do título de Mestre.

APROVADA em 30 de Novembro de 2021.

Prof. DSc. Eric Fernandes de Mello Araujo DCC/UFLA
Prof. DSc. Miguel Arjona Ramirez PSI/USP

Prof. DSc. Demóstenes Zegarra Rodríguez
Orientador

Prof. DSc. Bruno de Abreu Silva
Coorientador

**LAVRAS – MG
2021**

AGRADECIMENTOS

Agradeço primeiramente à minha família, que sempre me apoiou e me incentivou em todos os momentos da minha vida, que, com toda humildade e simplicidade, me ensinaram a ser uma pessoa decente, a respeitar e buscar meus sonhos honestamente e dentro do meu tempo, mesmo que seja com muito trabalho árduo. Vocês são os responsáveis por quem sou hoje, e agradeço imensamente por isso.

Gostaria de agradecer minha noiva e futura esposa Victória, que esteve sempre do meu lado durante toda essa jornada, nos momentos de comemoração por um artigo aprovado, ou nos momentos de dificuldades, com noites em claro e nas dificuldades pessoais que o tempo de pandemia de COVID-19 trouxe. Sem você, não estaria onde estou.

Gostaria também de agradecer os amigos que me acompanharam nessa jornada, desde a faculdade até o mestrado, dividindo sonhos e anseios, ajudando a relaxar com um café no Saulo ou conversando no “buteco”.

Também gostaria de agradecer aos professores Demóstenes e Bruno, por aceitarem participar dessa jornada comigo, sempre guiando para traçar o melhor caminho, dando conselhos e ajudando nas dificuldades encontradas.

Por fim, agradeço à UFLA, a todos os seus colaboradores, pela excelente estrutura que tornou este trabalho possível. Por fim, agradeço à FAPEMIG pelo apoio financeiro para o desenvolvimento deste trabalho.

Apenas que, busquem conhecimento.

E.T. Bilu

RESUMO

O câncer de pele é um problema de saúde mundial, tratando-se do tipo de câncer mais comum, sendo dividido em duas categorias, o câncer não melanoma, mais comum e menos letal, e o câncer melanoma, menos comum, porém, com uma grande taxa de mortalidade. O diagnóstico precoce é a melhor forma de combater o câncer de pele, evitando métodos invasivos no câncer de pele não melanoma e aumentando a taxa de cura e a sobrevida no caso do câncer do tipo melanoma. Diversas técnicas de visão computacional estão sendo empregadas na área médica, a fim de auxiliar os profissionais com diagnóstico, recomendação de tratamento, entre outros. Um desses métodos é o uso de redes neurais convolucionais, também chamadas pela sigla CNN, para a classificação de imagens de lesões e tumores, com estudos demonstrando uma capacidade de acerto até maior que de médicos treinados. Entretanto, alguns problemas são encontrados quando se trata da classificação de imagens de lesão de pele, como o desbalanceamento do conjunto de imagens. Para isso, podem ser usadas diversas técnicas para a geração de imagens a fim de balancear o conjunto, porém, uma que tem ganhado destaque é o uso de redes generativas adversárias, também chamadas pela sigla GAN, que são capazes de gerar imagens sintéticas com alta qualidade baseado em um conjunto previamente usado em um processo de treinamento. Este trabalho tem como objetivo buscar métodos que levem a um aumento no desempenho de redes CNN para a classificação de lesão de pele. Para isso, este trabalho propõe uma arquitetura CNN baseada na rede EfficientNetB0, denominada EfficientAttentionNet, para classificação de lesões cutâneas, especificamente melanoma e não melanoma. Em primeiro lugar, o conjunto de dados da imagem original, da *International Society for Digital Skin Imaging* (ISDIS), é pré-processado para eliminar os pelos ao redor da lesão cutânea. Posteriormente, um modelo GAN gerou imagens sintéticas para equilibrar o número de amostras por classe no conjunto de treinamento. Um modelo U-net é usado para criar máscaras com a região de interesse da imagem. Finalmente, é apresentado o modelo EfficientAttentionNet proposto para classificar o câncer de pele usando mecanismos de atenção com a máscara. Os resultados mostraram que o modelo de classificação proposto atingiu resultados de alto desempenho, obtendo acurácia de 0,979, precisão de 0,945, *recall* de 0,995 e ROCAUC de 0,976, servindo como referência para pesquisas na área de classificação de lesões cutâneas.

Palavras-chave: Redes neurais convolucionais. CNN. Câncer de Pele. Redes generativas adversárias. GAN.

ABSTRACT

Skin cancer is a global health problem, being the most common type of cancer, being divided into two categories, non-melanoma cancer, more common and less lethal, and melanoma cancer, less common, but with a high mortality rate. Early diagnosis is the best way to fight skin cancer, avoiding invasive methods in non-melanoma skin cancer and increasing the cure rate and survival in the case of melanoma type cancer. Several computer vision techniques are being used in the medical field, in order to help professionals with diagnosis, treatment recommendation, among others. One of these methods is the use of convolutional neural networks, also called by the acronym CNN, for the classification of images of lesions and tumors, with studies showing an even greater accuracy capacity than trained physicians. However, some problems are encountered when it comes to skin lesion image classification, such as image set imbalance. For this, several techniques can be used for the generation of images in order to balance the set, however, one that has gained prominence is the use of generative adversarial networks, also called by the acronym GAN, which are capable of generating synthetic images with high quality based on a set previously used in a training process. This work aims to search for methods that lead to an increase in the performance of CNN networks for the classification of skin lesions. For that, this work proposes a CNN architecture based on the EfficientNetB0 network, called EfficientAttentionNet, for classification of skin lesions, specifically melanoma and non-melanoma. First, the original image dataset, from the *International Society for Digital Skin Imaging* (ISDIS), is pre-processed to eliminate hairs around the skin lesion. Subsequently, a GAN model generated synthetic images to balance the number of samples per class in the training set. A U-net template is used to create masks with the region of interest of the image. Finally, the proposed EfficientAttentionNet model to classify skin cancer using mask attention mechanisms is presented. The results showed that the proposed classification model achieved high performance results, obtaining accuracy of 0.979, precision of 0.945, *recall* of 0.995 and ROCAUC of 0.976, serving as a reference for research in the area of classification of skin lesions.

Keywords: Convolutional neural network. CNN. Skin Cancer. Generative adversarial networks. GAN

LISTA DE FIGURAS

Figura 2.1 – Regra do ABCDE para autoexame de câncer de pele.	19
Figura 2.2 – Equipamento para dermatoscopia.	20
Figura 2.3 – Representação de um neurônio biológico.	23
Figura 2.4 – Representação de um neurônio artificial.	25
Figura 2.5 – Rede neural perceptron multicamada totalmente conectada.	27
Figura 2.6 – Representação dos sinais funcional e de erro.	28
Figura 2.7 – Representação de saídas das camadas de uma CNN. Representações obtidas através de uma CNN VGG-16 usando pesos pré-treinados com o conjunto de dados ImageNet.	29
Figura 2.8 – Arquitetura de uma CNN.	30
Figura 2.9 – Camada convolucional com o volume de entrada e de saída.	31
Figura 2.10 – Funcionamento da operação de convolução.	32
Figura 2.11 – Exemplo de imagens geradas usando <i>data augmentation</i>	34
Figura 2.12 – Comparação do desempenho e quantidade de parâmetros das EfficientNets e outras CNNs.	35
Figura 2.13 – Arquitetura básica da EfficientNet-B0.	35
Figura 2.14 – Exemplo de uma matriz de confusão.	36
Figura 2.15 – Valores de VP, VN, FP e FN em uma matriz de confusão.	37
Figura 2.16 – Funcionamento de uma rede GAN.	38
Figura 2.17 – Arquitetura da rede StyleGAN.	40
Figura 2.18 – Arquitetura de uma rede U-Net.	42
Figura 2.19 – Arquitetura de uma rede TransUNet.	44
Figura 2.20 – Arquitetura de uma rede Attention U-Net.	45
Figura 2.21 – Diagrama representando as métricas de IoU e DC.	46
Figura 4.1 – Metodologia usada para o desenvolvimento deste trabalho.	53
Figura 4.2 – Exemplo de imagens do conjunto de dados construído.	54
Figura 4.3 – Processamento realizado sobre o conjunto de imagens criado.	55
Figura 4.4 – Exemplo do processo de redimensionamento da imagem.	56
Figura 4.5 – Esquemático do detector usada para detecção de pelos.	56
Figura 4.6 – Comparação entre uma imagem dermatoscópica após a remoção de pelos.	57
Figura 4.7 – Exemplo das operações realizadas sobre as imagens para a GAN.	58

Figura 4.8 – Arquitetura da GAN usada para a geração das imagens sintetizadas.	58
Figura 4.9 – Exemplo das imagens já processadas do <i>dataset</i> ISIC 2017.	60
Figura 4.10 – Diagrama da rede U-Net implementada.	61
Figura 4.11 – Bloco de atenção proposto.	62
Figura 4.12 – Arquitetura da CNN proposta.	63
Figura 5.1 – Comparação entre as matrizes de confusão dos treinamentos realizados com e sem pelo no <i>dataset</i>	69
Figura 5.2 – Exemplos de resultados da GAN com a métrica FID.	70
Figura 5.3 – Valores da métrica FID durante o processo de treinamento da GAN.	70
Figura 5.4 – Exemplos de máscaras preditas pela CNN U-Net.	72

LISTA DE TABELAS

Tabela 3.1 – Comparação entre os trabalhos do estado da arte e a proposta do presente trabalho.	52
Tabela 5.1 – Comparação do desempenho da CNN com imagens com e sem pelo.	68
Tabela 5.2 – Resultados dos treinamentos das redes U-Net.	71
Tabela 5.3 – Resultados da avaliação de desempenho obtidos pelos modelos em diferentes estágios.	73

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Objetivos	14
1.1.1	Objetivos específicos	14
1.2	Organização do trabalho	14
2	REFERENCIAL TEÓRICO	16
2.1	Câncer de Pele	16
2.1.1	Tipos de Câncer de Pele	16
2.1.1.1	Câncer de pele não melanoma	16
2.1.1.2	Câncer de pele melanoma	17
2.1.2	Sintomas	18
2.1.3	Diagnóstico	20
2.1.4	Tratamento	20
2.2	Inteligência Artificial e Redes Neurais Artificiais	22
2.2.1	Redes Neurais Artificiais	22
2.2.1.1	Neurônio Biológico	23
2.2.1.2	Neurônio Artificial	24
2.3	CNN (<i>Convolutional Neural Network</i>)	28
2.3.1	Inspiração Biológica	29
2.3.2	Arquitetura da CNN	29
2.3.2.1	Camada de Entrada	31
2.3.2.2	Camada Convolutacional	31
2.3.2.3	Camada de <i>Pooling</i>	32
2.3.2.4	Camada Totalmente Conectada	33
2.3.3	<i>Data Augmentation</i>	33
2.3.4	EfficientNet	34
2.3.5	Métricas de desempenho para CNNs	36
2.4	GAN (<i>Generative Adversarial Networks</i>)	38
2.4.1	<i>StyleGAN</i>	39
2.4.2	Métricas de desempenho para GANs	40
2.5	U-Net	41
2.5.1	TransUNet	43

2.5.2	Attention U-Net	44
2.5.3	Métricas de desempenho para U-Nets	46
2.6	Considerações Finais	46
3	TRABALHOS RELACIONADOS	48
4	MÉTODO E SOFTWARES USADOS	53
4.1	Conjunto de Dados	53
4.1.1	Pré-processamento	54
4.2	<i>Generative Adversarial Network</i>	57
4.3	Segmentação de lesões de pele	59
4.4	Mecanismo de Atenção e EfficientAttentionNet	61
4.5	Software utilizados	64
4.5.1	Python	64
4.5.2	TensorFlow	65
4.5.3	ktrain	66
4.5.4	Google Colab	66
4.6	Considerações Finais	67
5	RESULTADOS	68
5.1	Resultados da remoção de pelo	68
5.2	Resultados da GAN	69
5.3	Resultados do treinamento da U-Net	71
5.4	Resultados do treinamento da CNN EfficientAttentionNet	72
5.5	Considerações Finais	74
6	CONCLUSÕES	75
	REFERÊNCIAS	77

1 INTRODUÇÃO

Atualmente, o câncer de pele é um problema de saúde pública em todo o mundo, sendo o quarto câncer mais comum no ano de 2020, segundo dados da Organização Mundial de Saúde (OMS) (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020), se somado o tipo melanoma e não melanoma. Ainda segundo a OMS, o número de novos casos de câncer de pele no ano de 2020 em todo o mundo foi de 1.522.708 casos, além de 120.774 mortes decorrentes de complicações da doença (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020). Nos Estados Unidos, mais de 9.500 pessoas são diagnosticadas com câncer de pele a cada dia, além de que, mais de duas pessoas morrem a cada hora de câncer de pele (SKIN CANCER FOUNDATION, 2019).

Segundo o levantamento da OMS sobre a incidência e mortalidade do câncer no ano de 2018, os países que lideram a classificação mundial de casos por 100.000 habitantes em idade normalizada (taxa ASR) no caso do câncer de pele, é a Austrália e a Nova Zelândia, com incidência de 36,6 e 31,6 respectivamente para o câncer do tipo melanoma, além de 140,0 e 127,5 respectivamente para o não melanoma. No caso da taxa de mortalidade, os países que lideram a classificação são Papua Nova Guiné e Namíbia, com mortalidade de 5,1 e 3,8 respectivamente para câncer do tipo não melanoma. Já no caso do tipo melanoma, os países com maiores taxas de mortalidade são a Nova Zelândia e a Noruega, com 4,7 e 3,2 respectivamente (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020).

No Brasil, o câncer de pele corresponde a 33% dos casos diagnosticados, com o Instituto Nacional do Câncer (INCA) registrando a cada ano 180 mil novos casos. O câncer de pele não melanoma, o tipo mais comum da doença, apesar de possuir baixa letalidade, apresenta um elevado número de casos. O câncer do tipo melanoma, apesar de apresentar uma taxa menor de casos, se trata de uma doença de maior gravidade, com alto índice de mortalidade (SBD, 2018).

Além do problema de saúde pública, o câncer de pele é também um problema econômico. Nos Estados Unidos, o valor gasto anualmente com o tratamento do câncer de pele é estimado em \$8,1 bilhões de dólares, sendo \$4,8 bilhões com câncer do tipo não melanoma e \$3,3 bilhões com melanoma (SKIN CANCER FOUNDATION, 2019). No Brasil, somente no estado de São Paulo, no ano de 2007, foram gastos cerca de R\$ 172 milhões de reais do setor público e privado. Com o câncer de pele não melanoma foram gastos cerca de R\$ 63 milhões de reais, sendo R\$ 37 milhões do setor público. Já no melanoma, o valor gasto pelo setor público foi de R\$ 33 milhões (SOUZA et al., 2011).

O câncer de pele do tipo não melanoma possui impacto no custo do sistema público de saúde do Brasil equivalente ao câncer do tipo melanoma. Noventa e oito por cento dos pacientes diagnosticados com câncer de pele não melanoma são identificados nos estágios precoces, contabilizando 97% dos custos de tratamento deste tipo de câncer. Enquanto isso, no câncer melanoma, somente 63% dos casos são diagnosticados precocemente, representando somente 3% do custo de tratamento dos casos de câncer de pele melanoma(SOUZA et al., 2011).

O diagnóstico precoce das duas categorias de câncer de pele é de extrema importância. No caso do câncer de pele do tipo não melanoma, o diagnóstico precoce significa tratamentos menos agressivos, trazendo uma qualidade de vida melhor para o paciente (SOUZA et al., 2011). Já no caso do câncer de pele do tipo melanoma, estudos comprovam que a detecção nos primeiros estágios desta categoria de câncer reduz significativamente a taxa de mortalidade do câncer do tipo melanoma (SBD, 2018).

Atualmente, o diagnóstico do câncer de pele é realizado utilizando dermatoscopia, usado por 81% dos dermatologistas nos Estados Unidos. Para o diagnóstico do tipo melanoma, o uso do dermatoscópio incrementa a acurácia do diagnóstico de 4,7 a 5,6 vezes em relação a análise a olho nu. Já no caso do tipo não melanoma, a sensibilidade é aumentada de 67% para 87% e a especificidade de 97% para 98%. Porém, a dermatoscopia é um exame dependente do operador, necessitando bastante treinamento e experiência para alcançar bons resultados (REITER et al., 2019).

Recentemente, técnicas de visão computacional e inteligência artificial têm sido empregadas na área médica. O uso destas técnicas tem aumentado significativamente a velocidade do diagnóstico, além de diminuir os riscos de erros humanos (ZHANG et al., 2020). Um grande exemplo é a plataforma da IBM, IBM Watson Health, que, com suas soluções de inteligência artificial, impactaram mais de 15.000 clientes e parceiros, 80.000 profissionais e 185.000 pacientes (IBM, 2020).

Um método de inteligência artificial que vem sendo empregado em várias áreas, inclusive na criação de sistemas de auxílio médico, são as chamadas Redes Neurais Convolucionais, ou CNNs (do inglês, *Convolutional Neural Networks*). Esta rede neural artificial, recebe como entrada um arquivo de imagem, realiza um processamento e entrega como resultado uma tabela de probabilidades associadas aos rótulos predefinidos no treinamento.

Diversos trabalhos mostram que, usando CNNs, é possível realizar uma classificação igual, ou melhor, que dermatologistas (HAENSSLE et al., 2020), recebendo como entrada ima-

gens de lesões pigmentadas, “manchas” ou “pintas” anormais, provindas do exame de dermatoscopia. Além disso, esse resultado é obtido com a CNN recebendo como entrada somente uma única imagem, enquanto o dermatologista, possui informações de várias fontes, como risco do paciente, evolução da lesão, etc. (HAENSSLE et al., 2020). O trabalho de Aldwgeri e Abu-backer (2019) utiliza um modelo conjunto de CNNs para obter resultados superiores ao usando somente um único modelo.

Além disso, outro método que tem ganhado destaque é o uso de Redes Generativas Adversárias, ou GANs (do inglês, *Generative Adversarial Network*), para o aumento do conjunto de imagens disponível. Essa categoria de rede neural é capaz de gerar imagens sintéticas. Para isso, é realizado previamente um treinamento, e então, baseado nesse treinamento, a rede consegue entregar uma imagem fiel ao conjunto de imagens treinado.

Alguns trabalhos mostram a aplicação de redes generativas adversárias para a realização da tarefa de *data augmentation*, ou seja, aumentar a quantidade de imagens que constitui um *dataset*. Qin et al. (2020b) propõe a realização de imagens sintéticas usando uma GAN do tipo *style-based GAN*. Já Abdelhalim, Mohamed e Mahdy (2021) apresentam uma GAN utilizando método de atenção e treinamento progressivo para a construção de um conjunto de imagens maior para um posterior treinamento. Em ambos os trabalhos, o resultado alcançado foi superior ao alcançado pela CNN sem a inclusão das imagens sintéticas.

Outra abordagem que tem sido bastante integrado às CNNs são os chamados métodos de atenção. Neles, alguma forma de processamento é realizado no interior da CNN, visando definir qual área da imagem que está sendo analisada é de maior valia para a classificação, dando mais “atenção” a essa parte, sendo essa a origem do nome. Esse método é bastante aplicado principalmente nas CNNs usadas para segmentação de imagens, porém, tem ganhado destaque nas suas aplicações para classificação. Os trabalhos de Xue, Kamata e Luo (2021) e Aggarwal, Das e Sreedevi (2019) mostram cenários onde a aplicação de métodos de atenção para classificação trouxe uma grande melhora no desempenho das CNNs propostas.

Como citado anteriormente, o câncer de pele é um dos cânceres com maior incidência no mundo, atingindo mais de um milhão de pessoas. O diagnóstico precoce é a chave para a cura da doença, dado que, quando diagnóstica nos estágios iniciais, a chance de vida do paciente é consideravelmente aumentada. A utilização de métodos de inteligência artificial pode servir como suporte para o dermatologista na realização do diagnóstico. Sendo assim, este trabalho se

justifica visto que busca encontrar métodos que colaborem para uma classificação mais assertiva de redes neurais convolucionais para a classificação de lesões de pele.

1.1 Objetivos

O objetivo principal deste projeto é a melhora na taxa de acerto no diagnóstico de lesões de pele usando métodos avançados de inteligência artificial, como redes neurais convolucionais, redes generativas adversárias e método de atenção. De modo a alcançar esse objetivo, foi proposto uma abordagem faseada, cujo objetivo era a melhora das métricas de avaliação de desempenho após o uso de cada técnica. A partir dos resultados alcançados no treinamento, foram realizados testes comparativos de cada uma dessas etapas, de modo a confirmar a melhora esperada. Com isso, esperamos que o processo de classificação de lesão de pele possa ser mais confiável, trazendo assim uma melhora no diagnóstico de lesões pigmentas, o que pode implicar em um aumento na chance de cura desta lesão, que, no caso do câncer de pele do tipo melanoma, pode significar na redução da taxa de mortalidade da doença.

1.1.1 Objetivos específicos

São os objetivos específicos deste trabalho de mestrado:

- a) implementar uma rede GAN para geração de imagens sintéticas de lesão de pele;
- b) implementar uma rede CNN para segmentação de lesão de pele;
- c) implementar um método de atenção proposto baseado em Região de Interesse;
- d) implementar uma CNN que utilize o método de atenção proposto, bem como as imagens sintéticas geradas pela GAN;
- e) avaliar os resultados obtidos.

1.2 Organização do trabalho

Este documento é dividido da seguinte maneira: o referencial teórico, apresentado no Capítulo 2, aborda os principais conceitos necessários ao entendimento do projeto, como inteligência artificial e redes neurais artificiais, com principal foco em redes neurais convolucionais, redes adversárias generativas, redes para segmentação e métodos de atenção. As conclusões da revisão da literatura para classificar imagens de câncer de pele e seus métodos para melho-

rar as métricas de avaliação são apresentadas no Capítulo 3 enquanto o Capítulo 4 apresenta a metodologia de pesquisa usada neste trabalho de mestrado. No Capítulo 5, são apresentados os resultados deste trabalho. Por fim, no Capítulo 6, são apresentadas as considerações finais acerca do trabalho apresentado nesta Dissertação de Mestrado.

2 REFERENCIAL TEÓRICO

Neste Capítulo, serão apresentados conceitos fundamentais para o desenvolvimento do projeto. O Capítulo apresenta os conceitos sobre câncer de pele na Seção 2.1, conceitos sobre inteligência artificial e redes neurais artificiais, enquanto na Seção 2.2, é dado foco principal nas redes neurais convolucionais na Seção 2.3. As redes generativas adversárias são apresentadas na Seção 2.4 e por fim, na Seção 2.5 são discutidas as redes de segmentação U-Net.

2.1 Câncer de Pele

O câncer de pele é o tumor que atinge a pele, o mais comum no Brasil e no mundo, segundo o Ministério da Saúde (MINISTÉRIO DA SAÚDE, 2018). Atinge principalmente a população com mais de 40 anos sendo considerado raro em crianças e pessoas negras. O caso mais comum de câncer de pele, o câncer de pele não melanoma, possui uma baixa letalidade, porém, o número de casos é alto. O câncer de pele é provocado pelo crescimento descontrolado e anormal das células que compõem a pele, que se dispõem formando camadas e, de acordo com qual camada é afetada, são definidos tipos diferentes de câncer (SBD, 2018).

2.1.1 Tipos de Câncer de Pele

Segundo o Ministério da Saúde (MINISTÉRIO DA SAÚDE, 2018), o câncer de pele acontece quando as células que fazem parte da pele se multiplicam sem controle, podendo ser dividido em duas classes:

- a) **câncer de pele melanoma**: ocorre nas células produtoras de melanina, substância responsável por determinar a cor da pele, sendo mais frequente em adultos brancos;
- b) **câncer de pele não melanoma**: tipo mais frequente de câncer de pele no Brasil, responsável por 30% dos tumores malignos registrados.

2.1.1.1 Câncer de pele não melanoma

O câncer de pele do tipo não melanoma é o tipo mais comum no Brasil e no mundo, possuindo alta chance de cura quando detectado e tratado precocemente, porém pode deixar mutilações bastante expressivas se não tratado corretamente (MINISTÉRIO DA SAÚDE, 2018).

Segundo estimativa do Instituto Nacional de Câncer (INCA), é esperado que o número de casos de câncer de pele não melanoma em cada ano do triênio 2020 – 2022 seja de 83.770 em homens e 93.160 em mulheres, correspondendo a um risco estimado de 82,12 casos novos a cada 100 mil homens e 86,65 casos novos a cada 100 mil mulheres (INCA, 2019).

O câncer de pele não melanoma apresenta diversos tipos de tumores, sendo os dois mais comuns:

- a) **Carcinoma basocelular (CBC):** é o tipo prevalente entre todos, surgindo nas células basais, que se encontram na camada mais profunda da epiderme (camada superior da pele). Possui baixa letalidade e alta chance de cura no caso de um diagnóstico precoce. Ocorrem com maior frequência em regiões expostas ao sol, como face, orelhas, pescoço, couro cabeludo, ombros e costas. Em alguns casos, podem existir outros fatores que desencadeiam seu surgimento. O tipo mais comum é o CBC nódulo-ulcerativo, sendo uma pápula vermelha, brilhosa, com uma crosta central e pode sangrar com facilidade (SBD, 2018);
- b) **Carcinoma espinocelular (CEC):** segundo tipo mais comum de câncer de pele, manifestando-se nas células escamosas, que constituem a maior parte das camadas superiores da pele. Pode ocorrer em qualquer parte da célula, porém, sua prevalência ocorre em áreas expostas ao sol. Nessas regiões, a pele apresenta sinais de dano solar, como enrugamento, mudança de pigmentação e perda de elasticidade. A ocorrência de CEC é maior em homens do que em mulheres. A exposição ao sol é a principal causa do surgimento do CEC, porém não é a única. Alguns casos são associados a feridas crônicas e cicatrizes na pele, uso de drogas antirrejeição de órgãos transplantados e a certos agentes químicos ou a radiação. Normalmente, possuem coloração avermelhada, apresentando-se como machucados e feridas espessos e descamativos, que não se cicatrizam e sangram ocasionalmente, podendo se assemelhar a verrugas (SBD, 2018).

2.1.1.2 Câncer de pele melanoma

O câncer de pele do tipo melanoma pode ocorrer em qualquer parte do corpo, na pele ou mucosas, expressando-se na forma de pintas, manchas ou sinais. Em pessoas de pele negra, é mais comum nas partes claras, como palma das mãos e planta dos pés. Embora o câncer de pele seja o mais comum no Brasil, o tipo melanoma representa somente 3% das neoplasias malignas

do órgão (MINISTÉRIO DA SAÚDE, 2018). Segundo relatório do INCA, os anos do triênio 2020 – 2022, o número de casos de câncer de pele do tipo melanoma em cada ano é de 4.200 em homens e 4.250 em mulheres, representando um risco estimado de 4,03 e 3,94 em 100 mil, respectivamente (INCA, 2019).

Embora o diagnóstico de câncer de pele traga medo e apreensão, as chances de cura quando descoberto nos primeiros estágios são de 90%. O melanoma em geral tem aparência de uma pinta ou sinal na pele, em tons acastanhados ou enegrecidos. Porém, a “pinta” ou “sinal” geralmente muda de cor, de formato e/ou tamanho, podendo sangrar. As lesões costumam surgir nas pernas em mulheres, tronco nos homens, e pescoço, e rosto em ambos os sexos (SBD, 2018).

Pessoas com pele clara e que se queimam facilmente quando se expõem ao sol, com fotótipo I e II, possui um maior risco de desenvolver a doença, que também pode ocorrer em indivíduos negros ou de fotótipo mais elevado, ainda que raramente. O melanoma ocorre nos melanócitos, células que produzem a melanina, pigmento que dá cor à pele. Nos estágios iniciais, o melanoma se desenvolve somente na camada superior da pele, facilitando a remoção cirúrgica e a cura do tumor. Nos estágios mais avançados da doença, a lesão é mais profunda e espessa, o que aumenta a chance de ocorrer um espalhamento para outros órgãos, a chamada metástase, diminuindo assim a chance de cura. Dado isso, é de extrema importância o rápido diagnóstico do melanoma. Embora o diagnóstico seja pior, avanços na medicina possibilitam que pessoas com melanoma avançado hoje tenham um aumento de sobrevida e na qualidade de vida (SBD, 2018).

Um papel central no desenvolvimento do melanoma é a hereditariedade. Assim, pessoas com familiares diagnosticados com a doença devem se submeter a exames preventivos regularmente. O risco de desenvolvimento do melanoma aumenta quando existem casos registrados em familiares de primeiro grau (SBD, 2018). Hoje em dia, testes genéticos são capazes de determinar quais mutações levam ao desenvolvimento do melanoma avançado (como BRAF, cKIT, NRAS, CDKN2A, CDK4), possibilitando a escolha do melhor tratamento para cada paciente (SBD, 2018).

2.1.2 Sintomas

O câncer de pele se assemelha com pintas, eczemas ou lesões benignas. Visto isso, é importante, na hora de detectar qualquer irregularidade, conhecer bem a própria pele e saber

em quais regiões existem pintas. Entretanto, somente um exame clínico feito por um médico especializado ou uma biópsia podem identificar o câncer de pele, porém, deve-se ficar atento diante da ocorrência de alguns sintomas (SBD, 2018):

- uma lesão na pele com aparência elevada e brilhante, translúcida, avermelhada, castanha, rósea ou multicolorida, com crosta central e sangra com facilidade;
- uma pinta preta ou castanha que muda de cor, textura, se torna irregular nas bordas e cresce de tamanho;
- uma mancha ou ferida que não se cicatriza, que continua a crescer, coça, apresenta crosta, erosões ou sangramento.

Figura 2.1 – Regra do ABCDE para autoexame de câncer de pele.

	benigno	maligno
A ssimetria	 simétrico	 assimétrico
B ordas	 bordas regulares	 bordas irregulares
C or	 uma só cor	 mais de uma cor
D iâmetro	 menor que 6 mm	 maior que 6 mm
E volução	 mudança rápida na aparência (tamanho, forma, cor, etc.)	

Fonte: Adaptado de Ministério da Saúde (2018).

Além destes sintomas, melanomas metastáticos podem apresentar outras, variando conforme a área para onde o câncer avançou. Isso inclui nódulos na pele, inchaço nos gânglios linfáticos, falta de ar ou tosse, dores abdominais e de cabeça, por exemplo (SBD, 2018).

Várias sociedades de dermatologistas, como a Sociedade Brasileira de Dermatologia e a *American Cancer Society* defendem o uso da Regra do ABCDE, vista na Figura 2.1, guia para a identificação de sinais sugestivos do melanoma. Entretanto, a Agência Internacional de Pesquisa do Câncer chamou atenção para as limitações desta regra quando aplicada em lesões

pequenas, visto que pequenos melanomas podem não apresentar tais divisões tão claramente (MINISTÉRIO DA SAÚDE, 2018).

2.1.3 Diagnóstico

Segundo o Ministério da Saúde (2018), o diagnóstico do câncer de pele é feito por um dermatologista através de um exame clínico. Em algumas situações, é possível que o dermatologista utilize um exame chamado dermatoscopia, que consiste no uso de um equipamento especializado e pode ser visto na Figura 2.2, que permite a visualização de camadas da pele não visíveis a olho nu, capaz de permitir uma melhoria de 49% na acurácia do diagnóstico se comparado com um diagnóstico realizado a olho nu (KITTLER et al., 2002), e, em certas situações, é necessário realizar uma biópsia.

Figura 2.2 – Equipamento para dermatoscopia.



Fonte: (COMMONS, 2009).

A biópsia é o exame indicado para a confirmação do câncer de pele. O material deve ser encaminhado para o laboratório de anatomia patológica que emitirá o laudo. Outros exames podem ser necessários para realizar a determinação do estadiamento da doença e decidir qual o melhor tratamento. Por meio desses exames, é possível identificar se o câncer de pele é melanoma ou não melanoma e seus tipos (MINISTÉRIO DA SAÚDE, 2018).

2.1.4 Tratamento

Segundo a Sociedade Brasileira de Dermatologia (SBD, 2018), todos os casos de câncer de pele devem ser diagnosticados e tratados precocemente, incluindo os casos com baixa letalidade, que podem provocar lesões mutilantes ou desfigurações em áreas expostas do corpo.

Existem diversas formas de tratamento para o câncer de pele não melanoma, com a modalidade variando conforme o tipo e a extensão da doença, porém, a maior parte dos carcinomas basocelulares e espinocelulares podem ser tratados com procedimentos simples, sendo eles:

- a) **cirurgia excisional:** consiste na remoção do tumor com um bisturi, e de uma borda adicional de pele sadia, como margem de segurança. O tecido removido é analisado para aferir que todas as células cancerosas foram removidas. Tal técnica possui altos índices de cura e pode ser empregada em casos de tumores recorrentes;
- b) **curetagem e eletrodissecção:** empregada em tumores menores, é realizada a raspagem da lesão com cureta, enquanto um bisturi elétrico é usado para destruir as células cancerígenas, repetindo o procedimento várias vezes de modo a não deixar vestígios de células tumorais. Entretanto, tal procedimento não é recomendado para tumores mais invasivos;
- c) **criocirurgia:** destrói o tumor por meio do congelamento com nitrogênio líquido. A técnica possui uma taxa de cura menor que a cirurgia excisional, porém pode ser uma boa alternativa em casos de tumores pequenos ou recorrentes. Trata-se de um tratamento que não envolve cortes ou sangramentos, porém também não é recomendada para tumores mais invasivos;
- d) **cirurgia micrográfica de Mohs:** nessa cirurgia, é retirado o tumor e um fragmento de pele ao redor com uma cureta, com o material sendo analisado. Esse procedimento é repetido sucessivamente, até que não existam vestígios de células tumorais. Essa técnica preserva boa parte dos tecidos saudáveis, indicada para tumores mal-delimitados ou em áreas críticas, principalmente no rosto, onde cirurgias mais amplas levam a cicatrizes extensas e desfigurações;
- e) **terapia fotodinâmica (PDT):** é aplicado um agente fotossensibilizante, como o ácido 5-aminolevulínico (5-ALA) na pele lesada. Após algumas horas, a área é exposta a uma luz intensa que ativa o 5-ALA, destruindo as células tumorais, com danos mínimos aos tecidos saudáveis.

Além das cirurgias, a radioterapia, quimioterapia, imunoterapia e as medicações orais e tópicas são outras opções de tratamento.

Já no caso do câncer de pele do tipo melanoma, conforme a extensão, agressividade e localização do tumor, bem como a idade e estado de saúde do paciente, o tratamento é escolhido. As intervenções mais comuns são a cirurgia excisional e a Cirurgia Micrográfica de Mohs. Já no

caso do melanoma metastático, não existe cura, por isso, é de extrema importância a detecção e tratamento precoce da doença. Embora sem cura, hoje é possível viver por mais tempo e com mais qualidade, realizando o controle da doença a longo prazo. Por isso, é importante que os pacientes realizem testes genéticos para determinar quais mutações apresentam (como BRAF, cKIT, NRAS, CDKN2A, CDK4), possibilitando assim a escolha de tratamento que pode trazer melhores resultados para cada caso. Por exemplo, mais de 90% dos pacientes que apresentam a mutação BRAF podem se beneficiar do tratamento com terapia-alvo oral, que retarda a progressão do melanoma e melhora a qualidade de vida do paciente. Além disso, outros tratamentos podem ser recomendados, isoladamente ou em combinação para o tratamento de melanomas avançados, incluindo quimioterapia, radioterapia e imunoterapia (SBD, 2018).

2.2 Inteligência Artificial e Redes Neurais Artificiais

Ao longo da história da humanidade, o desenvolvimento da Inteligência Artificial é uma das conquistas mais emblemáticas já alcançadas. O desenvolvimento de aplicações capazes de dirigir, conversar através de um celular, acionar eletrodomésticos e, em certo nível, simular o pensamento ou ação humana envolvem uma diversidade de algoritmos e dispositivos criados desde o início da década de 1950, quando os primeiros estudos sobre Inteligência Artificial foram iniciados (MEDEIROS, 2018).

As pesquisas na área de Inteligência Artificial são bastante diversificadas entre si, podendo ser baseadas no funcionamento do cérebro humano; no processo de seleção natural; baseada na coletividade de animais, além de serem aplicadas nas mais diversas áreas, como medicina, classificação de imagens e mercado financeiro (LUGER, 2013).

2.2.1 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs), são consideradas um dos maiores adventos da área da inteligência artificial. O funcionamento das RNAs possibilita a aprendizagem de padrões que tem como base a interligação de pequenos elementos simples baseados nos neurônios biológicos (MEDEIROS, 2018).

De acordo com Haykin (2007), os trabalhos na área de redes neurais artificiais podem ter sido motivados desde o começo graças ao conhecimento da forma completamente diferente que o cérebro humano processa informação em comparação com um computador. Ainda segundo

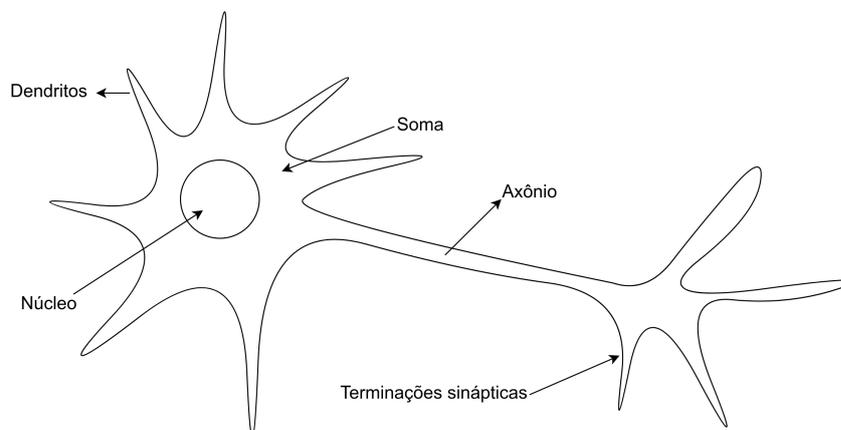
ele, o cérebro humano pode organizar os neurônios de tal forma a realizar tarefas como reconhecimento de padrões, percepção e controle motor de forma mais rápida que o computador mais potente existente.

Os trabalhos na área de RNA buscam realizar a emulação da representação do cérebro vivo, com base nas configurações descobertas da neurociência. Entretanto, mesmo com avanços por parte dos neurocientistas, no que tange ao funcionamento fisiológico de um conjunto de neurônios, isto ainda é uma tarefa desafiadora (MEDEIROS, 2018).

2.2.1.1 Neurônio Biológico

O cérebro humano possui pequenas unidades responsáveis pelo processamento das informações, sendo essas unidades, chamadas neurônios, as unidades fundamentais dos tecidos do sistema nervoso. O funcionamento do neurônio se resume em conduzir impulsos elétricos sob certas condições. O neurônio biológico, Figura 2.3, pode ser dividido em três partes básicas: os dendritos; o corpo celular (também chamado de soma); e o axônio.

Figura 2.3 – Representação de um neurônio biológico.



Fonte: Adaptado de Silva, Spatti e Flauzino (2010).

No corpo celular se encontram as organelas responsáveis pelo metabolismo do neurônio, como o núcleo, mitocôndrias e outras organelas comuns no restante das células. Os dois componentes predominantemente únicos dos neurônios são os dendritos e o axônio. Os dendritos são extensões ramificadas do neurônio, responsáveis por receber sinais de outros neurônios. Eles podem possuir vários e complexos formatos, dependendo da sua localização e tipo. Dendritos podem se parecer com árvores, como os encontrados nas células de Purkinje do cerebelo; ou ser estruturas simplistas, como os dendritos dos neurônios motores da coluna vertebral (GAZZANIGA; IVRY; MANGUN, 2019).

Já o axônio é uma única extensão do corpo celular, sendo a estrutura de saída do neurônio. Sinais elétricos viajam pelo axônio até as terminações sinápticas, onde o neurônio transmite este sinal para outros neurônios ou outros alvos. Esta transmissão ocorre na sinapse, uma estrutura especializada onde dois neurônios estão próximos um do outro, por onde os impulsos elétricos são enviados de um para o outro (GAZZANIGA; IVRY; MANGUN, 2019).

Segundo Medeiros (2018), o impulso nervoso é transmitido a partir de uma complexa reação eletroquímica. São lançadas das sinapses substâncias químicas transmissoras, enviadas aos dendritos, abaixando ou subindo o potencial elétrico do neurônio. Quando um limite é atingido, um pulso elétrico ou potencial de ação é enviado ao axônio, onde tal pulso se espalha e chega eventualmente em outras sinapses, liberando neurotransmissores no corpo destas células.

As sinapses podem ser divididas em duas categorias: as excitatórias, que incrementam o potencial de outros neurônios; e as inibitórias, que decrementam este potencial. Os neurônios podem realizar novas conexões com outros neurônios, com tais mecanismos formando a base do aprendizado do cérebro (MEDEIROS, 2018).

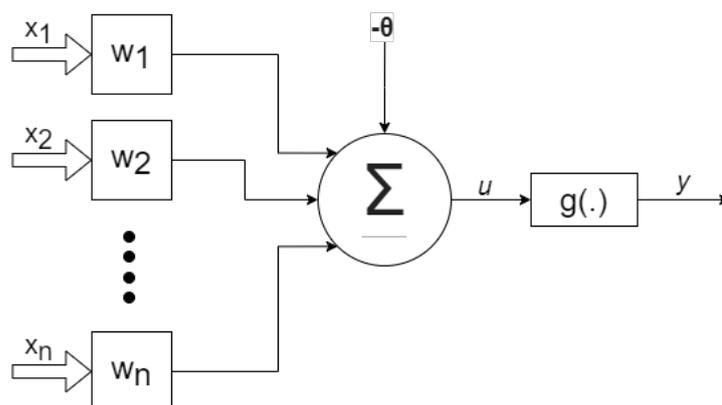
Assim, uma rede biológica pode ser abstraída e ter seu comportamento simulado, dando origem ao neurônio artificial. Assim, uma rede neural artificial é um sistema massivamente paralelo e distribuído, constituído de unidades simples, com a propensão de armazenar conhecimento e torná-lo disponível para uso (HAYKIN, 2007).

2.2.1.2 Neurônio Artificial

Estimulados pelas descobertas sobre os neurônios biológicos e os avanços tecnológicos criados por Alan Turing e John Von Neumann, Warren McCulloch e Walter Pitts publicaram o primeiro modelo de um neurônio biológico (MCCULLOCH; PITTS, 1943). Segundo Silva, Spatti e Flauzino (2010), mesmo se tratando de um modelo simplista, este possui todas as características de um neurônio biológico, ou seja, alta conectividade e paralelismo, ainda sendo um modelo usado em diversas arquiteturas de RNAs.

No modelo de McCulloch e Pitts, representado na Figura 2.4, podem ser vistas partes que realizam tarefas semelhantes ao neurônio biológico. Um exemplo disso são os valores de entrada espelhados no conjunto $\{x_1, x_2, \dots, x_n\}$, análogos aos impulsos coletados pelos dendritos. Outro exemplo disso são as ponderações realizadas pelas sinapses, que no modelo artificial,

Figura 2.4 – Representação de um neurônio artificial.



Fonte: Adaptado de Silva, Spatti e Flauzino (2010).

correspondem ao conjunto de pesos sinápticos $\{w_1, w_2, \dots, w_n\}$. A relevância de cada entrada x_i é dada pela multiplicação do seu valor com o peso sináptico w_i .

No modelo do neurônio artificial, sete partes principais são encontradas, sendo elas: (i) sinais de entrada do ambiente (x_i), cada uma correspondendo a um parâmetro distinto da aplicação; (ii) pesos sinápticos (w_i), sendo os responsáveis por ponderar cada entrada da rede neural, classificando parâmetros de maior relevância; (iii) combinador linear (Σ) cuja função é agregar os sinais de entradas já ponderados de modo a produzir um valor de potencial de ativação; (iv) limiar de ativação (θ) que especifica qual o valor de potencial de ativação deve ser alcançado para enviar um valor à saída do neurônio; (v) o potencial de ativação (u) correspondendo ao valor da diferença entre o combinador linear e o potencial de ativação, que, caso positivo, o neurônio produzirá um valor excitatório, caso contrário, inibitório; (vi) função de ativação (g), responsável por limitar a saída do neurônio a um intervalo que corresponda a imagem funcional dele e (vii), saída do neurônio (y), sendo o valor final produzido pelo neurônio em relação aos valores de entrada, podendo ser o resultado da aplicação ou servir de entrada a outros neurônio (SILVA; SPATTI; FLAUZINO, 2010).

Tal modelo de neurônio artificial pode ser sumarizado em duas expressões simples:

$$u = \sum_{i=1}^n w_i \cdot x_i - \theta \quad (2.1)$$

$$y = g(u) \quad (2.2)$$

A estrutura de um cérebro artificial, em tese, deveria ser mais rápida que um biológico, porém, esta não é a realidade. Tipicamente, o tempo de processamento de um neurônio bi-

ológico acontece na ordem de milissegundos (10^{-3} s), enquanto circuitos digitais operam na ordem de nanossegundos (10^{-9} s). O cérebro biológico compensa este tempo com uma estrutura super maciça de conexões. Estima-se que o cérebro humano possua cerca de 10 bilhões de neurônios e 60 trilhões de sinapses e conexões, fazendo com que seja extremamente eficiente. Além disso, a eficiência energética do cérebro humano é estimada em $10^{-16} J$ (*joules*) por operação por segundo, enquanto computadores possuíam em 2007, eficiência energética média de $10^{-6} J$, segundo Haykin (2007). Outra razão que explica o cérebro biológico ser mais rápido do que o artificial é que as redes neurais artificiais são baseadas em modelos computacionais que envolvem a propagação de uma variável a partir de uma unidade de processamento a outra, ou seja, um sistema serial e centralizado, mesmo usando paralelismo. Enquanto isso, redes neurais biológicas se comunicam através de pulsos, com o tempo entre cada pulso sendo usado para transmitir as informações e realizar cálculos, ou seja, um sistema paralelo e distribuído (ELUYODE; AKOMOLAFE, 2013).

O modelo de McCulloch e Pitts foi posto a prova por Rosenblatt (1958), quando este desenvolveu o modelo de rede neural artificial Perceptron, usando como base o modelo do neurônio artificial de McCulloch e Pitts.

Haykin (2007) cita seis propriedades que caracterizam uma rede neural artificial, sendo elas:

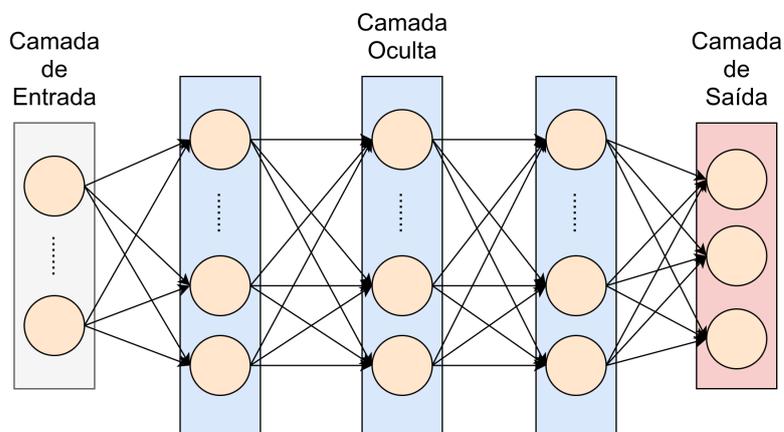
- a) **não linearidade:** os neurônios podem ser não lineares, permitindo aproximações de funções de mapeamento não lineares;
- b) **mapeamento de entrada-saída:** a rede aprende usando exemplos, realizando um mapeamento de padrões apresentados na entrada com as saídas dadas pelos exemplos na fase de aprendizagem;
- c) **adaptabilidade:** as redes neurais artificiais são treinadas e armazenam o conhecimento aprendido nos pesos sinápticos, que podem ser alterados caso o conjunto de entradas se altere com o passar do tempo;
- d) **resposta a evidências:** a rede neural, além de realizar uma tarefa de classificação, também pode informar o grau de confiança no padrão escolhido;
- e) **informação contextual:** a própria estrutura e ativação da rede é responsável pelo armazenamento do conhecimento;
- f) **tolerância a falhas:** redes neurais artificiais implementadas em hardware podem ser tolerantes a falhas em casos em que neurônios ou conexões forem danificadas; ou mesmo

redes implementadas em software que utilizam técnicas de podas, reduzindo a quantidade de neurônios e/ou sinapses, mantendo a condição de desempenho.

RNAs constituídas de um único neurônio possuem a capacidade de classificar dois grupos que sejam linearmente separáveis, ou seja, traça uma reta entre às duas classes, diferenciando-as por qual lado da reta um item se encontra. Entretanto, nem todos os problemas são linearmente separáveis, como a operação lógica XOR, problema este apresentado por Minsky e Papert (1969), mostrando a limitação do perceptron simples para tratar problemas complexos.

As redes perceptron multicamadas se diferenciam das redes perceptron simples por possuir camadas ocultas na rede neural. Com o uso de camadas ocultas, o número de pesos e sinapses cresce consideravelmente, permitindo assim um melhor mapeamento das entradas e saídas. Como dito anteriormente, um perceptron simples é capaz de tratar somente problemas linearmente separáveis, já a rede perceptron multicamadas, tem condições de lidar com conjunto de dados de forma com que, ao final do treinamento, consiga separá-los de forma não linear (MEDEIROS, 2018).

Figura 2.5 – Rede neural perceptron multicamada totalmente conectada.



Fonte: Adaptado de Khan et al. (2018).

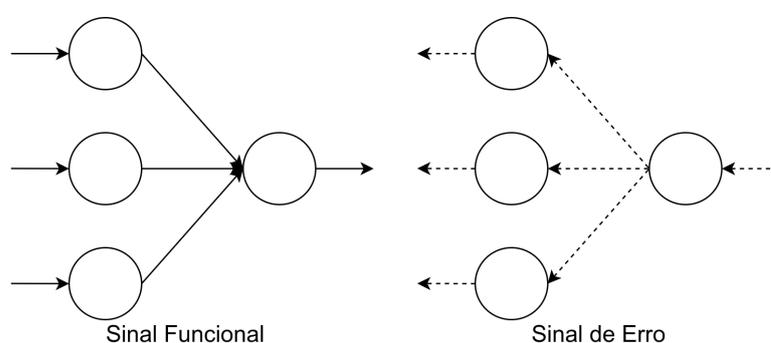
Uma rede neural com múltiplas camadas possui alguns aspectos destacados. Inicialmente, as redes possuem uma divisão em níveis ou camadas, que consistem em um número de neurônios. Tipicamente, o valor de entrada flui da camada de entrada para a camada de saída, responsável por realizar a predição (KHAN et al., 2018). Na Figura 2.5, é demonstrado um exemplo de uma rede neural com uma camada de entrada, três camadas ocultas e uma camada de saída, esta com três neurônios, ou seja, esta rede consegue realizar a separação de até 8 classes.

O uso de camadas ocultas implica no uso de algoritmos de aprendizagem que realizem a atualização dos pesos das camadas ocultas. O processo de ativação se dá primeiro nas camadas

internas, para depois chegar à camada de saída. A retroalimentação do erro também acontece nos pesos que conectam as camadas ocultas (MEDEIROS, 2018).

Segundo Haykin (2007), as redes neurais de múltiplas camadas envolvem o processamento de dois sinais, mostrados na Figura 2.6, que se propagam na rede, sendo eles o sinal funcional, que corresponde ao sinal enviado à camada de entrada, sendo o vetor de entrada da rede propagado com destino à saída da rede, passando por todas as camadas, ativando os neurônios até a saída da rede; e o sinal de erro, com origem na saída da rede, propagando-se até a entrada, responsável por ajustar os pesos dos neurônios ou sinapses.

Figura 2.6 – Representação dos sinais funcional e de erro.



Fonte: Adaptado de Medeiros (2018).

O algoritmo de retropropagação (*backpropagation*), popularizado por Rumelhart, Hinton e Williams (1986), envolve o processo chamado descida de gradiente, visando calcular o gradiente local de erro, ou seja, a direção para onde tende a crescer o valor do erro médio calculado, utilizando-o para realizar a correção dos pesos sinápticos na direção contrária do gradiente, assim, buscando o erro mínimo local.

2.3 CNN (*Convolutional Neural Network*)

Redes Neurais Convolucionais, também chamadas de CNN (do inglês *Convolutional Neural Network*), é uma das categorias mais conhecidas de redes neurais artificiais, especialmente na área de dados com múltiplas dimensões, como imagens e vídeos. As CNNs funcionam similarmente às redes neurais convencionais, tendo como principal diferença que cada camada de uma CNN possui duas ou mais dimensões de um filtro de convolução aplicado à entrada desta camada. Isso é de grande importância quando é desejado aprender padrões em dados com múltiplas dimensões, como imagens e vídeo (KHAN et al., 2018).

As redes neurais convolucionais são um modelo útil para a classe de paradigmas supervisionados e não supervisionados. O paradigma supervisionado é aquele onde são conhecidos a

entrada e o seu rótulo esperado, e o modelo aprende um mapeamento entre os dois. Já no paradigma não supervisionado, os rótulos verdadeiros não são conhecidos, e o modelo visa estimar a distribuição das amostras de dados das entradas. Um exemplo de aprendizado supervisionado é mostrado na Figura 2.7, no caso, classificação de imagem. Neste caso, a CNN aprende a mapear uma imagem para uma determinada categoria, detectando características abstratas, variando das mais simples às mais complexas. Estes recursos são usados na rede para prever a categoria correta da imagem de entrada (KHAN et al., 2018).

Figura 2.7 – Representação de saídas das camadas de uma CNN. Representações obtidas através de uma CNN VGG-16 usando pesos pré-treinados com o conjunto de dados ImageNet.



Fonte: Do Autor (2021).

2.3.1 Inspiração Biológica

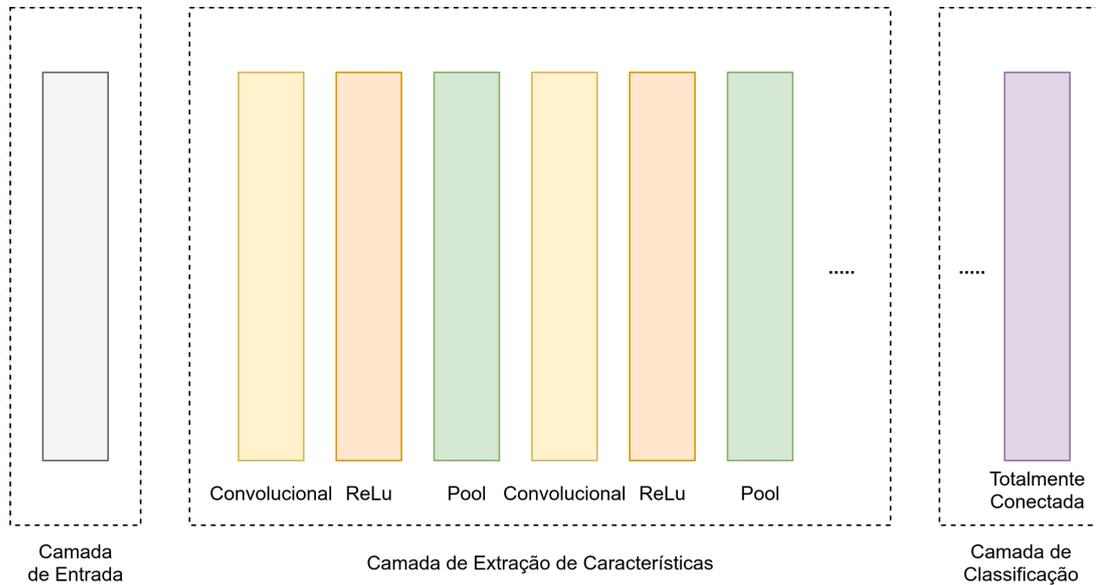
As redes neurais convolucionais são inspiradas no córtex visual dos animais. As células do córtex visual são sensíveis a pequenas sub-regiões de entrada, chamadas campo visual. Estas pequenas sub-regiões são agrupadas, cobrindo todo o campo visual. Estas células são adequadas para explorar a correlação espacial encontrada entre as imagens processadas no cérebro e funcionam como filtros locais na entrada. Existem duas classes nesta região do cérebro, uma com células mais simples, ativadas quando detectam padrões como bordas, e outra, mais complexa, ativadas quando possuem um campo receptivo maior e são invariantes à posição do padrão (EICKENBERG et al., 2017).

2.3.2 Arquitetura da CNN

As redes neurais convolucionais transformam os dados de entrada em um conjunto de pontuações, representando a chance de aquela imagem pertencer a uma determinada classe, através das camadas conectadas. Atualmente, existem diversas arquiteturas de CNN, porém

todas elas são baseadas no padrão de camadas, que pode ser observado na Figura 2.8 (PATTERSON; GIBSON, 2017).

Figura 2.8 – Arquitetura de uma CNN.



Fonte: Adaptado de Patterson e Gibson (2017).

A partir da Figura 2.8, é possível verificar três principais grupos: camada de entrada, camadas de extração de características e, por fim, camadas de classificação. A camada de entrada geralmente recebe um dado espacial, possuindo largura, altura e profundidade, representando os canais de cores, sendo normalmente três para imagens RGB. As camadas de extração de características têm um padrão de repetição de uma camada de convolução, uma ativação e uma camada de *pooling*. Estas camadas são responsáveis por encontrar um número de características na imagem e progressivamente constroem características de ordem maior. Isto corresponde ao tema de aprendizado profundo, visto que essas características são aprendidas automaticamente. Por fim, são encontradas as camadas de classificação em que existem uma ou mais camadas totalmente conectadas para obter os recursos de ordem maior e produz probabilidades ou pontuações de classe. Como os próprios nomes das camadas indicam, estas camadas são totalmente conectadas a todos os neurônios da camada anterior. Estas camadas produzem uma saída bidimensional $[b \times N]$, onde b é o número de imagens no lote em que foi processado, e N , correspondendo ao número de classes desejadas em pontuar (PATTERSON; GIBSON, 2017).

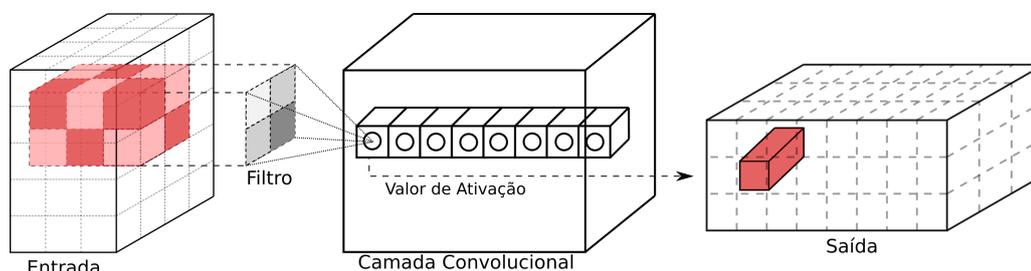
2.3.2.1 Camada de Entrada

A camada de entrada é a responsável por carregar e armazenar o dado bruto da imagem para ser processada pela rede neural artificial. Este dado é especificado por largura, altura e o número de canais, tipicamente três para uma imagem RGB (PATTERSON; GIBSON, 2017).

2.3.2.2 Camada Convolutiva

As camadas convolucionais são o núcleo de funcionamento de uma rede neural convolutiva. A Figura 2.9 ilustra a transformação dos dados de entrada usando um conjunto de neurônios conectados localmente à camada anterior. A camada calculará o produto escalar entre as regiões dos neurônios da camada de entrada e os pesos aos quais estão conectados localmente na saída (PATTERSON; GIBSON, 2017). Na Figura 2.9, é possível ver um fragmento de informação obtido na entrada processado pelo filtro, que possui pesos adaptáveis, sendo enviado a um conjunto de valores de ativação, que, por fim, são enviados à saída.

Figura 2.9 – Camada convolutiva com o volume de entrada e de saída.



Fonte: Adaptado de Patterson e Gibson (2017).

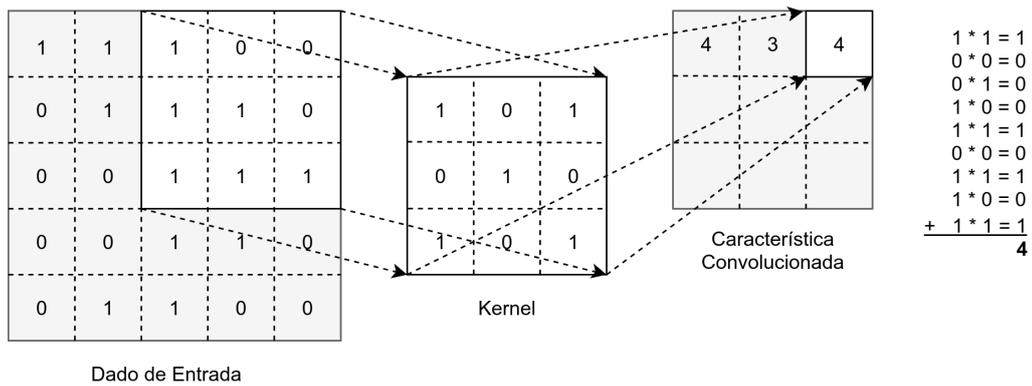
Normalmente, a saída resultante possui as mesmas dimensões espaciais, ou dimensões menores, porém, existem casos onde o número de elementos da terceira dimensão (profundidade) pode aumentar (PATTERSON; GIBSON, 2017).

A operação chave da camada convolutiva, chamada convolução, é definida como uma operação matemática que descreve a forma de como mesclar dois conjuntos de informação. É muito importante nas áreas da matemática e física e define uma ligação entre os domínios do espaço/tempo e frequência, através do uso de transformadas de Fourier. Nela, uma entrada é aplicada a um núcleo (comumente chamado *kernel*) de convolução e fornece um mapa de características como saída (PATTERSON; GIBSON, 2017).

A operação de convolução demonstrada na Figura 2.10 é conhecida como detecção de características de uma CNN. Como entrada para a convolução podem ser dados brutos, ou seja,

a imagem RGB, ou ainda o mapa de características de outra camada convolucional. Normalmente, é interpretado como um filtro em que o *kernel* seleciona dados para uma certa informação, como exemplo, um *kernel* de borda permite passar apenas informações da borda de uma imagem (PATTERSON; GIBSON, 2017).

Figura 2.10 – Funcionamento da operação de convolução.



Fonte: Adaptado de Patterson e Gibson (2017).

A Figura 2.10 demonstra como o *kernel* caminha através da entrada para produzir a saída. Em cada passo, o *kernel* multiplica sua entrada com seus pesos, criando uma única entrada no mapa de características de saída. Ou seja, na prática, a saída é um valor grande se a característica que esteja sendo procurada for detectada na entrada (PATTERSON; GIBSON, 2017).

2.3.2.3 Camada de Pooling

As camadas de *pooling* são normalmente inseridas entre sucessivas camadas de convolução sucessivas. É desejada a inserção de camadas de *pooling* entre camadas de convolução para gradativamente reduzir o tamanho espacial (largura e altura) dos dados. As camadas de *pooling* diminuem a representação de dados através da rede neural, ajudando no controle do sobreajuste, ou seja, prevenindo que a rede neural se ajuste muito bem ao conjunto de dados usado no aprendizado, mas com deficiência em prever novos resultados (PATTERSON; GIBSON, 2017).

Ainda segundo Patterson e Gibson (2017), as camadas de *pooling* utilizam uma função $\max()$, mostrada na Equação 2.3, para reduzir os dados de entrada espacialmente, operação chamada *max pooling*. Com um *kernel* de 2×2 , a função de $\max()$ retorna o maior valor entre os 4 valores analisados pelo filtro. A operação de *max pooling* não afeta a dimensão de profundidade da entrada.

$$\max(a, b) = \begin{cases} a, & \text{se } a \geq b \\ b, & \text{se } b > a \end{cases} \quad (2.3)$$

As camadas de *pooling* utilizam filtros para executar a subamostragem do volume de entrada. A configuração mais comum para a camada de *pooling* é de um filtro de dimensão 2×2 . Este filtro reduzirá cada canal da entrada em um fator de dois os seus valores de largura e altura, resultando no descarte de 75% das ativações (PATTERSON; GIBSON, 2017).

2.3.2.4 Camada Totalmente Conectada

Segundo Patterson e Gibson (2017), esta camada é usada para a classificação, sendo a saída da rede neural, ou seja, é a última camada da rede neural convolucional. Normalmente, possui dimensão de $[1 \times 1 \times N]$, com o valor de N correspondendo a quantidade de classes que serão avaliadas. Esta camada possui conexão entre todos os seus neurônios e todos os neurônios da camada anterior.

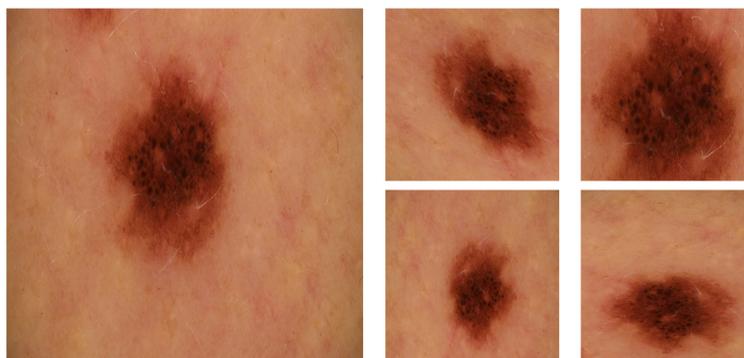
2.3.3 Data Augmentation

Segundo Aldwgeri e Abubacker (2019), um dos principais problemas enfrentados no aprendizado de máquina é o desbalanceamento de dados. O desbalanceamento de dados ocorre quando a quantidade de imagens por classe de um *dataset* é diferente. O desbalanceamento de dados afeta diretamente o desempenho do método de aprendizagem de máquina, tendendo a focar somente na classe com mais exemplos, ignorando as outras classes (MA; XU; ZHANG, 2018). É importante salientar que no ramo médico, existem desafios e limitações associadas à coleta de imagens, o que acaba gerando um desbalanceamento no *dataset*. Em primeiro lugar, a coleta de dados leva muito tempo. Em segundo lugar, a coleta de grandes volumes de dados sobre pacientes com certas doenças costuma ser um desafio devido à baixa prevalência da doença (MAZUROWSKI et al., 2008).

Para contornar esse problema, pode ser usada a técnica de *data augmentation*. Essa técnica consiste na criação de exemplos artificialmente, adicionando essas imagens à classe da imagem original (VASILEV, 2019). As principais operações para a criação de dados artificiais em imagens, também utilizada neste trabalho, é a rotação da imagem, inverter a imagem hori-

zontal ou verticalmente, adicionar *zoom* à imagem, deslocamento, entre outros. A Figura 2.11 apresenta um exemplo de imagem após passar pelo processo de *data augmentation*.

Figura 2.11 – Exemplo de imagens geradas usando *data augmentation*.



Fonte: Do Autor (2021).

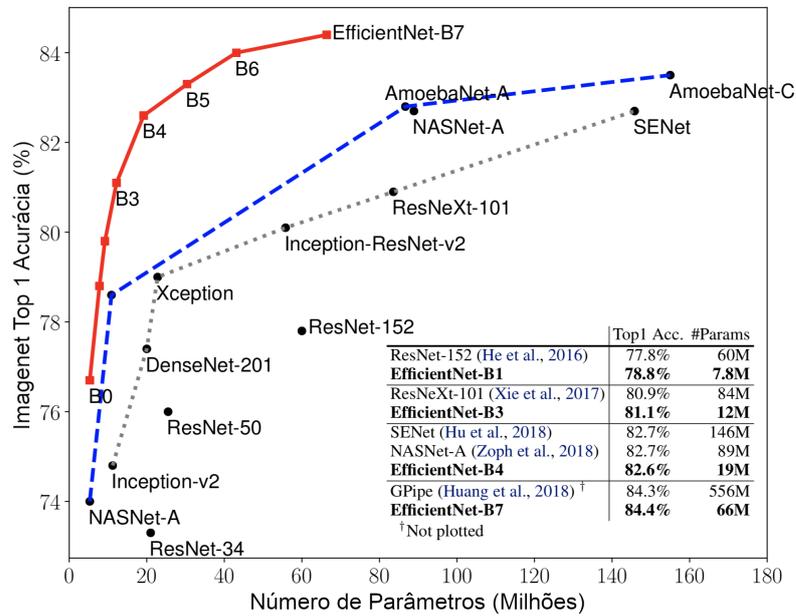
É importante frisar que algumas dessas transformações não devem ser aplicadas dependendo do seu *dataset*. Como citado por Goodfellow (2016), rotações de 180° não devem ser aplicadas em um conjunto de caracteres, uma vez que assim, existiriam exemplos gerados que pertenceriam a outra classe, como “b” se transformaria em “d”. Sendo assim, as transformações que impactariam na imagem da lesão, como formato, cor e textura não devem ser aplicadas neste contexto.

2.3.4 EfficientNet

Uma das arquiteturas de CNN que tem ganhado destaque são as chamadas EfficientNet. Essa arquitetura, proposta por Tan e Le (2020), usa um método de escala que aumenta uniformemente todas as dimensões de profundidade, largura e resolução eficientemente, usando um coeficiente composto simples. Com base nisso, foi criada uma família de 8 CNNs, chamadas EfficientNet-B0 até EfficientNet-B7. A EfficientNet-B7 em específico obteve 84,3% de acurácia top-1 no *dataset* ImageNet, enquanto é 8,4 vezes menor e 6,1 vezes mais rápida que a melhor rede convolucional previamente existente.

Como é possível observar na Figura 2.12, as redes EfficientNets foram as que obtiveram melhor resultado na classificação de imagens do *dataset* ImageNet, além de ser as que contêm menos parâmetros. Além disso, é possível ver que mesmo a menor configuração existente para as EfficientNet, a chamada EfficientNet-B0, alcançou resultados superiores aos de redes já consagradas, como a ResNet-50, consumindo uma quantidade consideravelmente menor de parâmetros.

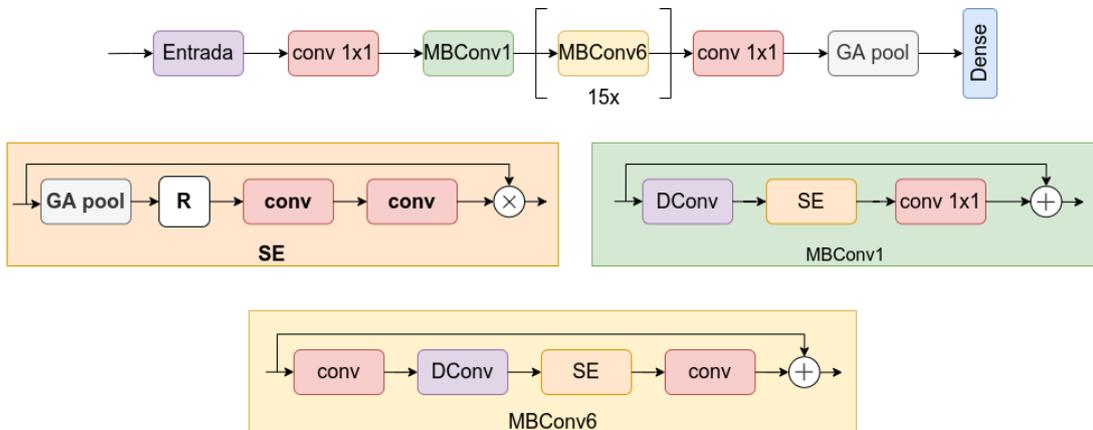
Figura 2.12 – Comparação do desempenho e quantidade de parâmetros das EfficientNets e outras CNNs.



Fonte: Adaptado de Tan e Le (2020).

Tal resultado é alcançado através do uso de três blocos básicos para construção da CNN. O primeiro é o bloco SE, baseado no bloco *Squeeze-and-Excitation*, proposto por Hu, Shen e Sun (2017), para recalibrar as características de cada canal. Por fim, temos os dois blocos de construção básicos responsáveis pela convolução, chamados MBConv6 e MBConv1. Estes blocos são em sua essência iguais, entretanto, o que diferem entre si, é a quantidade de operações de convolução que existem em cada bloco. A Figura 2.13 mostra tais blocos, bem como a arquitetura da rede EfficientNet-B0 para classificação de imagens do *dataset* ImageNet.

Figura 2.13 – Arquitetura básica da EfficientNet-B0.

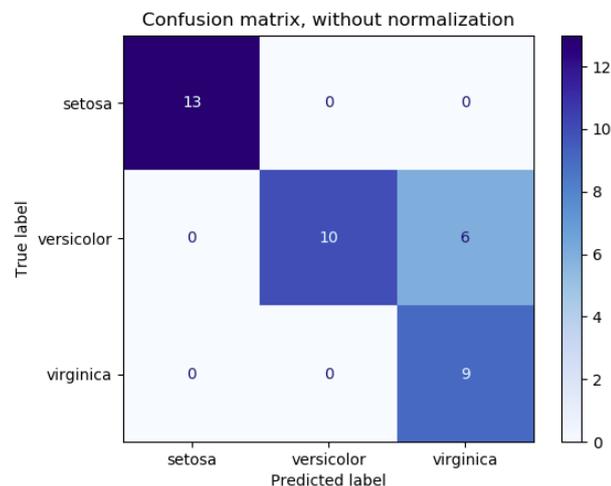


Fonte: Do Autor (2021).

2.3.5 Métricas de desempenho para CNNs

No que tange métricas de desempenho para redes neurais convolucionais, diversos estudos encontrados na literatura descrita no Capítulo 2 usam as métricas precisão, *recall* e pontuação F1 (*F1 score*), dentre outras, descritas por Buckland e Gey (1994) para a avaliação de algoritmos de aprendizagem de máquina. Uma ferramenta importante para o cálculo de métricas de cada classe é a matriz de confusão (do inglês *confusion matrix*), demonstrada na Figura 2.14. Nela, os valores na diagonal principal representam a quantidade de exemplos corretamente classificados, e os outros valores, classificados incorretamente. As linhas da matriz representam as classes e as colunas, as previsões (ZHONG et al., 2020).

Figura 2.14 – Exemplo de uma matriz de confusão.



Fonte: (SCIKIT-LEARN, 2018).

A partir da matriz de confusão é possível obter os valores de verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN), usados para o cálculo de outras métricas. A forma que tais valores podem ser obtidos é demonstrada na Figura 2.15.

Neste trabalho, as principais métricas empregadas para a avaliação das redes neurais convolucionais foram a acurácia e a área sob a curva ROC, do inglês, *Receiver Operating Characteristic*, (ROCAUC). Tais métricas foram escolhidas por serem utilizadas em vários artigos que tratam sobre esse assunto (ALDWGERI; ABUBACKER, 2019; MAHBOD et al., 2019; MAHBOD et al., 2020; GESSERT et al., 2018; CODELLA et al., 2019).

A métrica de acurácia é descrita pela Equação 2.4.

$$acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.4)$$

Figura 2.15 – Valores de VP, VN, FP e FN em uma matriz de confusão.

0.37	0.27	0.23	0.0	0.03	0.1	0.0
0.0	0.49	0.11	0.0	0.4	0.0	
0.01	0.03	0.45	0.0	0.1	0.49	0.0
0.0	0.12	0.0	0.0	0.0	0.88	0.0
0.09	0.04	0.04	0.02	0.2	0.0	
0.0	0.01	0.01	0.0	0.01	0.97	0.0
0.0	0.0	0.0	0.0	0.0	0.23	0.77

Verdadeiro Negativo

Falso Positivo

Falso Negativo

VP

FN

VN

FP

VN

Fonte: Adaptado de (LOUKAS, 2020).

Além disso, outras métricas foram utilizadas, de modo a gerar uma melhor comparação entre as redes neurais convolucionais implementadas, sendo elas:

- a) **recall ou sensibilidade**: medida de quantos dos verdadeiros positivos o modelo pode identificar, descrito pela Equação 2.5;

$$recall = \frac{VP}{(VP + FN)} \quad (2.5)$$

- b) **especificidade** : medida de quantos dos falso positivos o modelo pode identificar, descrito pela Equação 2.6;

$$especificidade = \frac{VN}{(VN + FP)} \quad (2.6)$$

- c) **precisão**: medida de quão precisas são as previsões positivas, descrita pela Equação 2.7;

$$precisão = \frac{VP}{(VP + FP)} \quad (2.7)$$

- d) **F1 score**: combina precisão e *recall* para fornecer uma avaliação geral do desempenho, mostrado na Equação 2.8.

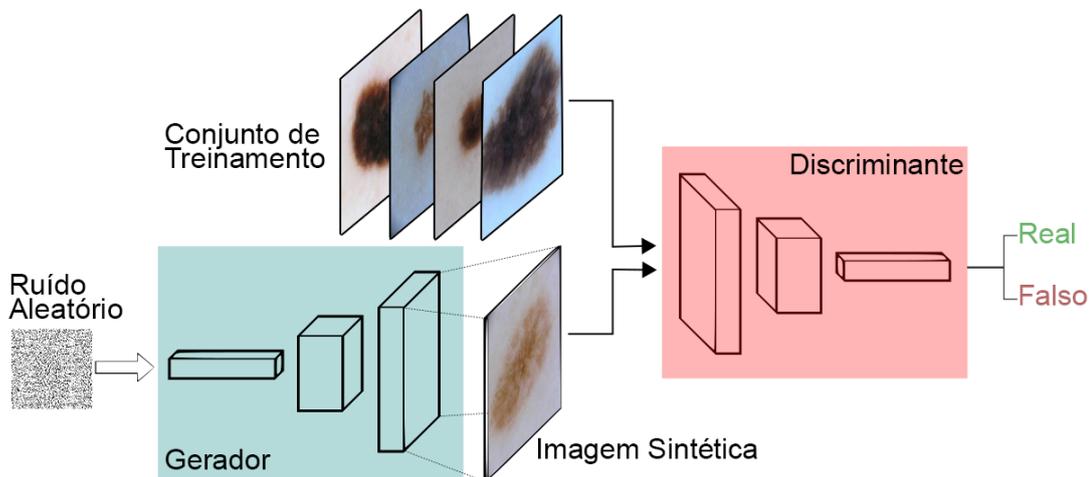
$$F1 = \frac{2 \times precisão \times recall}{precisão + recall} \quad (2.8)$$

2.4 GAN (*Generative Adversarial Networks*)

Além das redes classificatórias, outra categoria de redes neurais existentes são as redes generativas. Elas, ao contrário das redes classificadoras, não rotulam uma imagem, e sim, geram imagens. Em trabalho proposto por Goodfellow et al. (2014), as redes adversárias generativas, também chamadas GANs, do inglês, *Generative Adversarial Networks*, têm ganhado grande popularidade, com diversos trabalhos propondo melhorias e otimizações. Apesar de que as GANs possam gerar qualquer tipo de dados, seu principal uso tem sido para a geração de imagens.

As redes adversárias generativas são compostas por dois componentes, sendo eles, mostrados na Figura 2.16:

Figura 2.16 – Funcionamento de uma rede GAN.



Fonte: Adaptado de (DATA SCIENCE ACADEMY, 2021).

- a) **gerador:** É a parte generativa da GAN. É o responsável por gerar uma imagem realista com base em uma entrada aleatória;
- b) **discriminador:** Esta rede recebe duas entradas alternativamente: uma imagem real do *dataset* de treinamento ou uma imagem gerada pelo gerador. Essa rede visa determinar se a entrada se trata de uma imagem original ou falsa.

Segundo Vasilev (2019), as duas redes neurais são treinadas em conjunto. A rede discriminador distingue entre uma imagem real e uma imagem gerada. Por outro lado, a rede generativa cria imagens cada vez mais reais de modo a “enganar” a rede discriminador.

Em uma analogia, criada em Goodfellow et al. (2014), a rede generativa é como um falsificador, que produz dinheiro falso. Já a rede discriminador é como oficiais de polícia,

tentando encontrar o dinheiro falso, enquanto os dois constantemente tentam ganhar um do outro, daí o nome adversárias.

O objetivo final do sistema consiste em fazer com que a rede generativa alcance um resultado tão bom a ponto de a rede discriminador não conseguir distinguir entre as imagens reais e as geradas. Além disso, como o discriminador não realiza nenhuma classificação, o processo de treinamento de uma GAN é não supervisionado, não sendo necessário assim a rotulação prévia das imagens a serem geradas.

2.4.1 *StyleGAN*

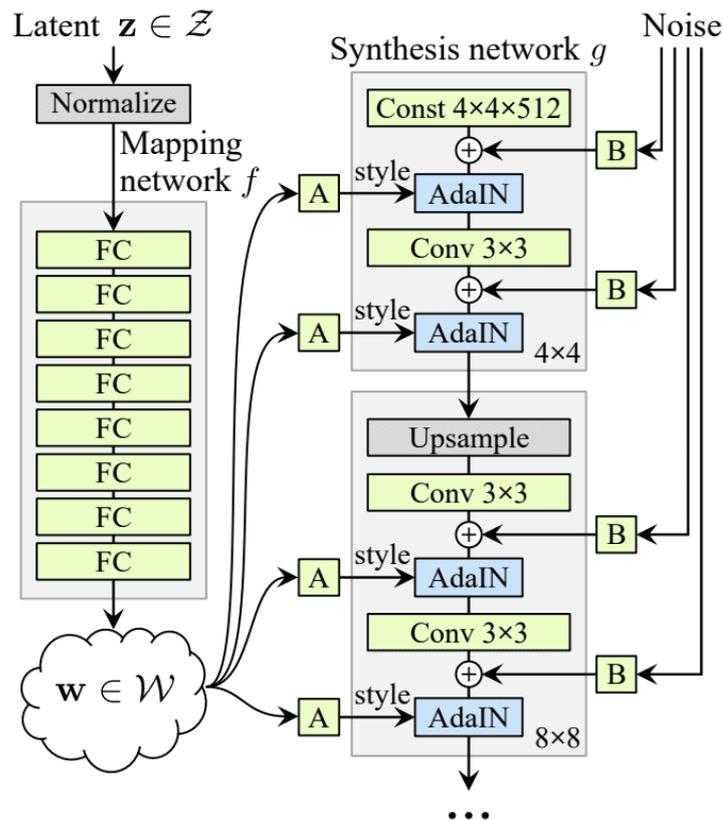
Atualmente, existem diversas arquiteturas de redes adversárias generativas, possuindo cada uma pequenas diferenças com as outras, melhorando o processo de geração de imagens, trabalhando na parte generativa ou na parte discriminador. Uma dessas arquiteturas que tem ganhado destaque é a *StyleGAN* (KARRAS; LAINE; AILA, 2019). Essa arquitetura utiliza um arcabouço de técnicas para alcançar resultados superiores em uma menor quantidade de tempo. Uma dessas técnicas é uma arquitetura de rede generativa para controlar melhor os níveis de detalhes de áreas grosseiras para áreas mais detalhadas. Além disso, é usada a técnica de crescimento da GAN, proposto em Karras et al. (2017), onde inicialmente a rede é treinada em pequenas resoluções, como 4×4 , e então vão-se adicionando camadas com maiores resoluções quando as inferiores estiverem estabilizadas. Isso faz com que o treinamento ocorra de forma mais rápida além de ser mais estável.

Além disso, a *StyleGAN* traz grandes mudanças na parte geradora da GAN. É usada uma rede de mapeamento, visando pegar um ponto amostrado no espaço latente como entrada e gerar um vetor de estilo. A rede de mapeamento é composta por oito camadas totalmente conectadas, sendo assim uma rede perceptron multicamadas. Então, o vetor de estilo é transformado e incorporado após as camadas convolucionais por meio da operação chamada cálculo da normalização de instância adaptativa, ou AdaIN (do inglês, *Adaptive Instance Normalization*). Essa operação constitui na normalização do vetor de características para uma distribuição normal gaussiana, seguida da adição do vetor de estilo como um *bias*.

Além dessa alteração no gerador, foi feita a alteração para fazer com que o gerador não use um ponto no espaço latente como entrada. No lugar, é usado um vetor constante de dimensão $4 \times 4 \times 512$ para a síntese das imagens.

Outra mudança realizada é a adição de ruído à GAN. O resultado obtido após a execução da camada de convolução da GAN é um bloco de mapas de ativação. Esse ruído, do tipo gaussiano, é adicionado em cada mapa de ativação, antes de ser processado pela operação AdaIN. O objetivo da inserção desse ruído é a introdução da variação de um nível de estilo em determinado nível de detalhe.

Figura 2.17 – Arquitetura da rede StyleGAN.



Fonte: (KARRAS; LAINE; AILA, 2019).

A arquitetura final da rede StyleGAN é apresentada na Figura 2.17. A inserção dessas modificações no gerador da GAN trouxe como resultado uma grande capacidade na geração de imagens realistas, principalmente de rostos, sendo este o conjunto de imagens usado para treinamento no trabalho de Karras, Laine e Aila (2019), além do controle de estilo da imagem gerada em diferentes níveis de detalhe.

2.4.2 Métricas de desempenho para GANs

Em redes adversárias generativas, a principal medida entre o gerador e o discriminador é o quão bem estão um em relação ao outro, por exemplo, o quão bem o gerador está enganando

o discriminador. Entretanto, isso não é uma boa métrica para medir a qualidade da imagem e diversidade, sendo assim, necessário usar outras métricas de desempenho.

Em se tratando de GANs, é desejável que a probabilidade condicional $P(y | x)$ seja altamente previsível, ou seja, dada uma imagem de entrada, deve ser fácil saber o tipo de objeto.

As principais métricas usadas para medida da qualidade e diversidade são: *Inception Score* (IS) e *Fréchet Inception Distance* (FID). Na métrica IS, é usada uma rede Inception para classificar as imagens geradas e $P(y | x)$, onde y são os rótulos e x são as imagens geradas. Uma deficiência da métrica IS é que ela pode representar mal o desempenho se gerar apenas uma imagem por classe.

Já no FID, também é usada uma rede Inception, porém, ela é usada para extrair recursos de uma camada intermediária. Após isso, é feita uma modelagem de distribuição de dados para esses recursos extraídos, utilizando uma distribuição gaussiana multivariada com média μ e covariância Σ . Sendo assim, o FID entre as imagens reais x e as imagens geradas g é calculado como:

$$FID(x, g) = \|\mu_x - \mu_g\|_2^2 + Tr(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}) \quad (2.9)$$

onde Tr é a soma de todos os elementos da diagonal principal. Sendo assim, valores de FID mais baixos significam uma melhor qualidade e diversidade de imagem. Além disso, a FID é mais robusta que o IS em relação a ruído, sendo assim ela é a melhor medida para diversidade de imagens. A FID tem viés alto, porém, uma baixa variância. Se for realizado o cálculo com um *dataset* de treinamento e um de teste, em teoria é esperado uma FID igual a zero, dado que ambas as imagens são reais, entretanto não é isso que acontece, com a FID ficando em 2 na média.

2.5 U-Net

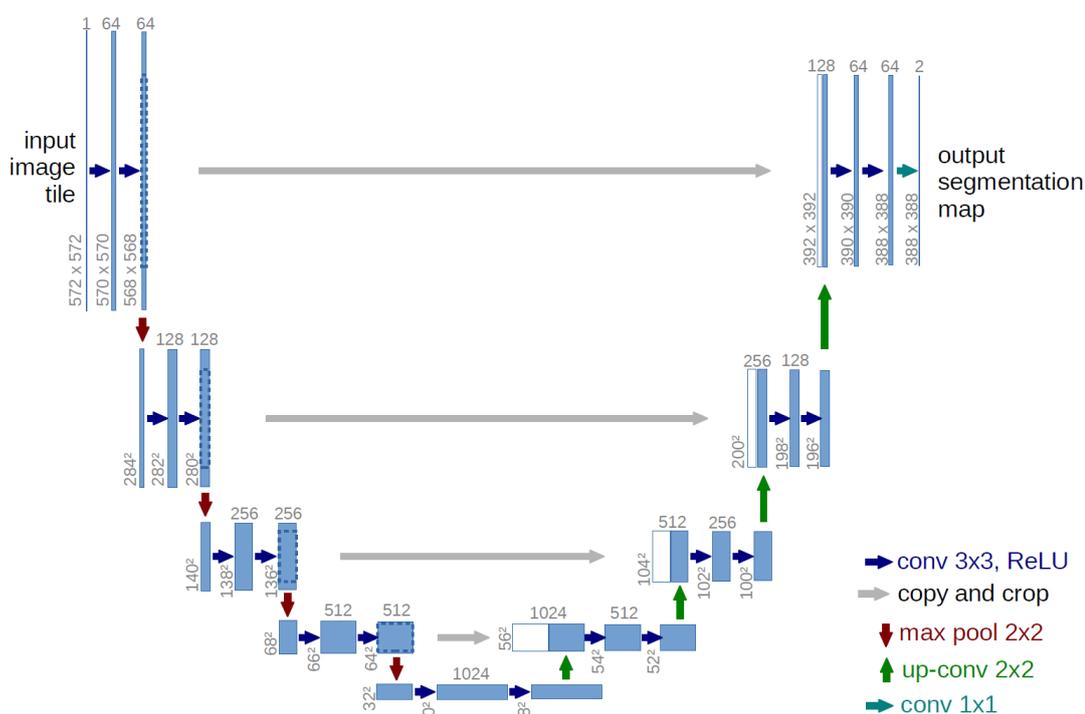
Além das redes para classificação de imagens e geração de imagens sintéticas, outra classe de redes neurais que existe são as redes usadas para segmentação de imagens. A segmentação de uma imagem nada mais é que, a partir de uma imagem, definir uma máscara que recorte um determinado item em específico na imagem.

Diversas técnicas têm sido propostas recentemente para a tarefa de segmentação de imagens. Porém, o método que mais se destaca são as redes U-Net. Proposta por Ronneberger,

Fischer e Brox (2015), esta rede tem sido amplamente usada para segmentação, porém, sua aplicação original é a segmentação de imagens biomédicas.

A rede U-Net é uma rede baseada nas redes para segmentação totalmente convolucionais, FCN, sigla para *Fully Convolutional Network* (LONG; SHELHAMER; DARRELL, 2015). A ideia é fazer com que os recursos de mais alto nível na camada do decodificador sejam mantidos. Uma rede U-Net é dividida em duas seções, a parte codificadora e a decodificadora, como pode ser observado na Figura 2.18.

Figura 2.18 – Arquitetura de uma rede U-Net.



Fonte: (RONNEBERGER; FISCHER; BROX, 2015).

A arquitetura da parte codificadora tem o funcionamento igual a uma rede neural convolucional. Na rede apresentada por Ronneberger, Fischer e Brox (2015), a seção codificadora consiste na aplicação repetida de duas convoluções com filtros de tamanho 3×3 . Cada convolução é seguida de uma função de ativação ReLU e uma normalização em lote. Em seguida, uma operação de *max pooling* 2×2 é aplicada de modo a realizar a redução das dimensões espaciais.

Já na parte decodificadora, cada etapa no caminho de expansão consiste na ampliação dos mapas de características, normalmente através do método de interpolação bi-cúbica em uma convolução de tamanho 2×2 . Essa operação é o inverso da realizada pelas camadas de *max pooling* da seção codificadora, fazendo com que as dimensões espaciais sejam aumentadas. Então, é realizada a concatenação com o mapa de características correspondente àquele nível

da U-Net. Tal concatenação é normalmente seguida por uma convolução de tamanho 3×3 , seguida por uma ativação do tipo ReLU. Na camada final da U-Net, é usada uma convolução 1×1 para mapear os canais para o número desejado de classes (RONNEBERGER; FISCHER; BROX, 2015).

Além disso, a rede U-Net pode ter a sua parte codificadora alterada para funcionar com outros modelos de CNN. Isso faz com que o desempenho da U-Net seja elevado, gerando um melhor mapa de segmentação (GHOSH et al., 2019). A rede U-Net ganhou popularidade graças ao seu grande desempenho, ganhando o desafio de rastreamento celular ISBI (MAŠKA et al., 2014). No *dataset* PhC-U373, a rede U-Net obteve 0,9203 em IoU, enquanto o segundo lugar 0,83. Já no conjunto DIC-HeLa, obteve Interseção sobre União, do inglês *Intersection over Union* (IoU) de 0,7756, significativamente melhor que o segundo lugar, que obteve 0,46 (GHOSH et al., 2019).

Além disso, a rede U-Net possui uma grande quantidade de redes que a utiliza como base para seu funcionamento, como as redes W-Net (XIA; KULIS, 2017), V-Net (MILLETARI; NAVAB; AHMADI, 2016), UNet++ (ZHOU et al., 2018), R2U-Net (ALOM et al., 2018), Attention U-Net (OKTAY et al., 2018), ResUNet-a (DIAKOGIANNIS et al., 2020), U²-Net (QIN et al., 2020a), UNet 3+ (HUANG et al., 2020), TransUNet (CHEN et al., 2021) e Swin-Unet (CAO et al., 2021).

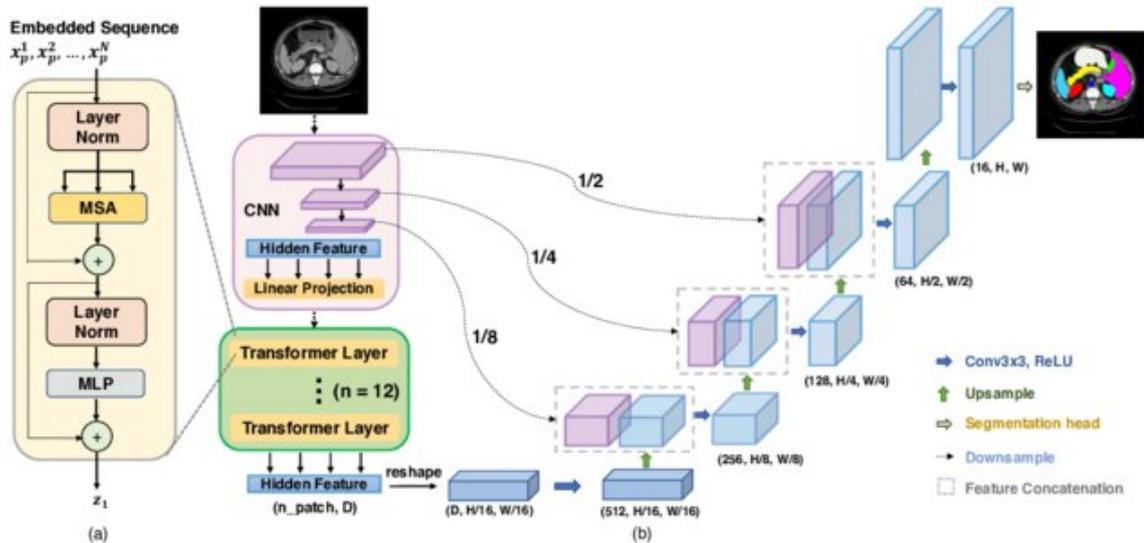
2.5.1 TransUNet

Como citado, a partir da rede U-Net, foram propostas várias arquiteturas que utilizam como base seu funcionamento. Uma dessas redes é a TransUNet, proposta por Chen et al. (2021).

Tem sido proposto como alternativa para as redes U-Net o uso de transformadores. Transformadores tem se mostrado como ótimas estruturas para várias tarefas da área de aprendizagem de máquina, como processamento de linguagem natural e classificação de imagens. Porém, transformadores têm como ponto fraco sua característica de se concentrar somente no contexto global, não conseguindo capturar informações detalhadas de localização.

Com base nisso, Chen et al. (2021) propuseram uma rede híbrida, usando tanto uma rede U-Net quanto transformadores. Essa rede híbrida foi batizada de TransUNet. A arquitetura da rede TransUNet pode ser observada na Figura 2.19.

Figura 2.19 – Arquitetura de uma rede TransUNet.



Fonte: (CHEN et al., 2021).

Como é possível observar na Figura 2.19, a rede TransUNet é composta por uma rede U-Net padrão. Entretanto, no último nível do codificador na U-Net, foi inserido o transformador Vision Transformer (ViT), sendo este transformador implementado como 12 camadas de transformação. Cada uma dessas camadas consiste em uma camada de normalização, uma camada de *Multiple Sequence Alignments* (MSA), outra camada de normalização e então uma rede *perceptron* multicamadas.

As imagens a serem processadas são enviadas à TransUNet, onde os recursos são extraídos a cada nível pela CNN, com a CNN funcionando como codificadora. Então, a projeção linear dessas características é enviada para o transformador, para obter o contexto global desses recursos. Então, a imagem de saída é gerada a partir das características de saída do transformador e as conexões com a seção codificadora da U-Net.

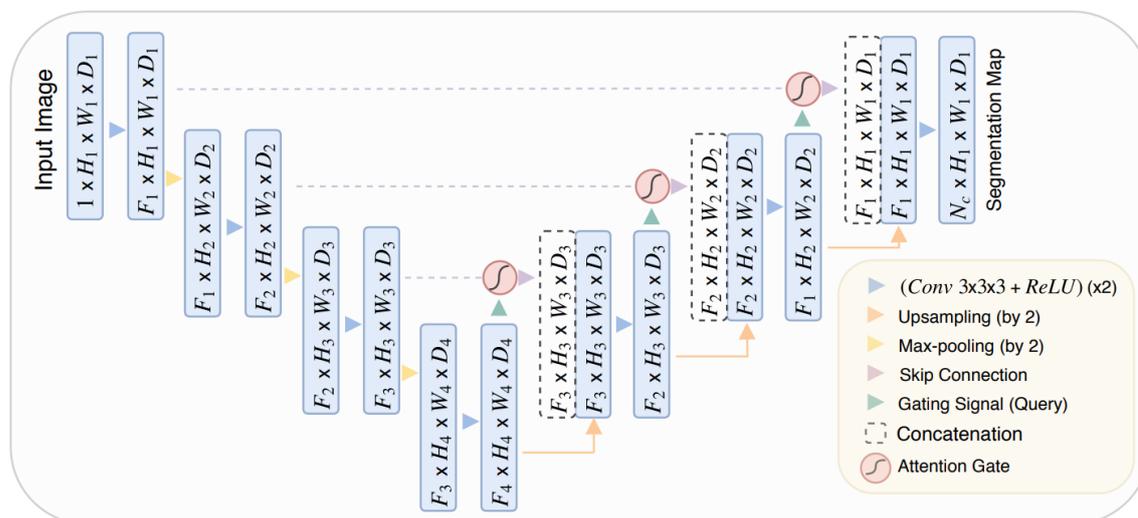
2.5.2 Attention U-Net

Outra melhoria que vem sendo proposta para as redes U-Net é a aplicação de métodos de atenção. No contexto de segmentação de imagens, os métodos de atenção visam destacar somente as regiões de maior interesse durante o processo de treinamento.

Durante o processo de recriação da informação em uma rede U-Net, a informação espacial é imprecisa. Para resolver este problema, as redes U-Net usam ligações diretas entre a rede codificadora e decodificadora. Entretanto, isso acaba fazendo com que características sejam redundantes. Logo, métodos de atenção podem ser usados para diminuir o número de caracte-

terísticas redundantes nas redes U-Net, visto que as características redundantes não relevantes para o treinamento são filtradas pelo método de atenção.

Figura 2.20 – Arquitetura de uma rede Attention U-Net.



Fonte: (OKTAY et al., 2018).

Oktay et al. (2018) propõem a rede Attention U-Net baseada neste princípio. Como pode ser observado na Figura 2.20. Nela, é possível observar que os blocos de atenção propostos são inseridos nas conexões da U-Net, com esse objetivo de filtrar as características redundantes.

Para a realização dessa filtragem é usado um bloco de atenção que funciona da seguinte maneira: o bloco de atenção recebe duas entradas, a entrada x , que consiste nas características vindas da seção codificadora da U-Net, e a entrada g , sendo estas as características do bloco inferior da seção decodificadora da U-Net. Ambas as entradas passam por uma convolução, para que possuam as mesmas dimensões, e são então multiplicadas. Após essa multiplicação, os elementos são novamente enviados a uma camada de convolução, reduzindo mais uma vez suas dimensões, e então é realizada a aplicação de uma função sigmoide às características resultantes dessa convolução, onde valores próximos a um representam características mais importantes. Então, é realizada a ampliação do vetor para as dimensões originais da entrada x , onde esses valores ampliados são multiplicados pela entrada x . Por fim, o resultado dessa multiplicação é enviada para a seção decodificadora da U-Net.

Com isso, Oktay et al. (2018) conseguiram diminuir a quantidade de características redundantes da rede U-Net, melhorando seu resultado.

2.5.3 Métricas de desempenho para U-Nets

Para a medida de desempenho de uma rede convolucional do tipo U-Net, métricas são usadas para classificar a qualidade das máscaras criadas. Essas métricas são compostas das métricas de acurácia, *recall* e precisão, empregadas normalmente nas redes convolucionais usadas para classificação.

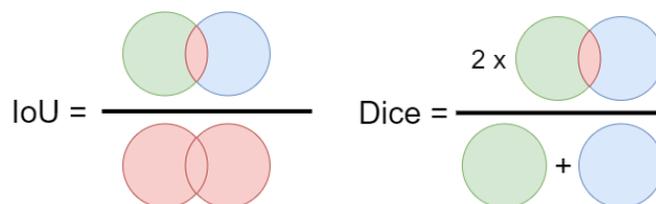
Entretanto, as principais métricas usadas para a avaliação de redes neurais para a segmentação de imagens são as métricas de Interseção sobre União, do inglês *Intersection over Union* (IoU) (2.10), e o Coeficiente de Dice (DC) (2.11). Ambas as métricas utilizam as operações sobre conjuntos para identificar o quão próximo uma máscara predita chegou da máscara verdadeira do *dataset*.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2.10)$$

$$DC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (2.11)$$

Por se tratar de métricas que utilizam operações sobre conjuntos, é possível realizar a representação de tais métricas usando Diagramas de Venn, permitindo assim uma melhor visualização de como tais métricas são calculadas, e quais seus principais pontos de análise. Por isso, na Figura 2.21 são mostrados visualmente os cálculos das métricas de avaliação IoU e DC.

Figura 2.21 – Diagrama representando as métricas de IoU e DC.



Fonte: Do Autor (2021).

2.6 Considerações Finais

Neste Capítulo foram apresentados os conhecimentos necessários para o desenvolvimento deste trabalho, como o que é o câncer de pele, as diferenças entre o câncer de pele do tipo melanoma e o não melanoma, bem como os sintomas, o diagnóstico e tratamento. Também

é citado sobre inteligência artificial, com foco em redes neurais artificiais, principalmente as do tipo CNN, usadas para a classificação de imagens, as redes GAN, usadas para a geração de imagens sintéticas e por fim, a rede U-Net, usada para a segmentação de imagens. No Capítulo seguinte são apresentados os trabalhos relacionados de maior relevância para este projeto.

3 TRABALHOS RELACIONADOS

Diversos trabalhos na literatura mostram o uso de CNNs para classificação de pele, bem como a aplicação de técnicas para alcançar um melhor desempenho nessas CNNs. Além disso, são encontrados trabalhos apontando o uso de GANs para a geração de imagens sintéticas para a criação de um novo conjunto de dados. Neste capítulo são apresentados trabalhos que, utilizando diferentes ferramentas e técnicas, obtiveram resultados significativos na implementação de uma CNN ou GAN para classificação de lesões pigmentadas.

Villa-Pulgarin et al. (2022) realizam o treinamento dos modelos DenseNet-201, Inception-ResNet-V2 e Inception-V3 para a classificação de lesão de pele usando os *datasets* HAM10000 (TSCHANDL; ROSENDAHL; KITTLER, 2018) e ISIC 2019 (TSCHANDL; ROSENDAHL; KITTLER, 2018; CODELLA et al., 2018; COMBALIA et al., 2019). Além disso, os autores propõem a utilização de métodos de *data augmentation* e ajuste dos modelos de CNN usados, de modo a encontrar um melhor resultado para os modelos propostos. Nesse trabalho, as principais regras usadas são: acurácia, precisão, *recall* e *F1-Score*. Inicialmente, Villa-Pulgarin et al. (2022) realizam um comparativo entre os três modelos, usando o *dataset* HAM10000 para avaliar a melhora do desempenho quando usado a técnica de *data augmentation*. Por meio desse experimento, os autores mostraram que a utilização de um conjunto de dados balanceado traz um melhor resultado no treinamento da CNN, obtida acurácia de 95% para a CNN DenseNet-201, 94% para Inception-ResNet-V2 e 95% para Inception-V3. Então, foi realizado um treinamento com os modelos ajustados, com os modelos DenseNet-201, Inception-ResNet-V2 e Inception-V3 obtendo acurácia de 98%, 97% e 96%, respectivamente. Por fim, Villa-Pulgarin et al. (2022) realizaram o treinamento do modelo vencedor, neste caso, a CNN DenseNet-201 ajustada para a realização de classificação de lesão de pele usando o conjunto de dados ISIC 2019. Neste treinamento, onde também foi empregada a técnica de *data augmentation*, foi obtido resultado de 93% em acurácia, e 94% no restante das métricas usadas.

Já o trabalho de Nersisson et al. (2021) apresenta uma configuração de CNN que, além de utilizar as características processadas pela CNN, também utiliza de características tradicionais, como textura e cores da área onde está contida a lesão de pele na imagem de entrada. Além das características obtidas através da CNN, é empregado um extrator de textura, usando os métodos GLCM e Gabor, enquanto para a extração de características da coloração da imagem, foi empregado o método CLCM. O modelo proposto funciona da seguinte maneira: inicialmente,

a imagem de entrada é enviada para a rede YOLO, onde o mapa de características é gerado, bem como pelos métodos GLCM, Gabor e CLCM. Os resultados dos últimos 3 métodos são então concatenados e tem sua dimensão reduzida através de um método de *Principal Component Analysis*. Então, mais uma vez é feita a concatenação dos mapas de características, porém, dessa vez, é realizada a concatenação entre o mapa proveniente da CNN bem como a saída do método PCA. Essa fusão de características é então enviada para uma rede totalmente conectada, que fica a cargo de classificar a entrada como maligna ou benigna. Após a realização do treinamento utilizando o conjunto de imagens ISIC 2016, usando a técnica de *data augmentation*, os autores obtiveram como resultado uma acurácia de 94%, precisão de 0,85, *recall* de 0,88 e ROCAUC de 0,95.

O trabalho de Qin et al. (2020b) apresenta a utilização de uma rede GAN para a geração de imagens sintéticas de modo a aumentar a quantidade de imagens existentes. Em um primeiro momento, foi proposto o desenvolvimento de uma rede GAN do tipo *style-based*. Os autores propõem uma modificação no método de controle do estilo e geração de ruído, tanto na parte geradora quanto discriminante da GAN, visando sintetizar com maior eficiência imagens de lesão de pele de alta qualidade. Os autores realizaram avaliações da rede GAN proposta, comparando-a com outros modelos de GAN existentes na literatura, com o modelo proposto superando-os nas métricas de *Inception Score* (IS), *Fréchet Inception Distance* (FID), precisão e *recall*, obtendo IS de 3,037, enquanto as redes GAN, DCGAN e StyleGAN obtiveram 1,727, 2,118 e 3,125 respectivamente. Já em relação a FID, foi obtido valor de 1,059, enquanto GAN, DCGAN e StyleGAN alcançaram valores de 3,275, 1,367 e 2,796. Então, com a rede geradora já treinada, Qin et al. (2020b) recorre a uma CNN do tipo ResNet50 pré-treinada usando o método de aprendizado por transferência para a verificação do impacto na adição de imagens de lesão de pele sintéticas ao conjunto de treinamento. Para isso, foi realizado o treinamento da rede usando um novo conjunto de imagens balanceado, conjunto este baseado no HAM10000. Ou seja, foram adicionadas imagens sintéticas ao *dataset* HAM10000 até que a quantidade de imagens em cada classe fosse igual. Nas métricas de acurácia, precisão e *recall*, os autores obtiveram resultados de 95,2%, 83,2% e 74,3% respectivamente utilizando o modelo proposto, uma melhora de 1,6%, 24,4% e 3,6%, em relação ao modelo de CNN base.

Abdelhalim, Mohamed e Mahdy (2021) também apresenta outro projeto onde é usado GAN para a realização da operação de *data augmentation*. Os autores propõem as redes batizadas de *Self-attention Progressive Growing of GANs* (SPGGANs) para a geração de imagens de

256 × 256 pixels de tamanho. Dois principais métodos são aplicados à GAN proposta de modo a melhorar o resultado encontrado, sendo eles, o método de atenção e o crescimento progressivo. No método de atenção, são usados blocos na rede de modo a reforçar as características de maior interesse para a geração da imagem específica, fazendo com que a qualidade final da imagem seja superior. Já o crescimento progressivo é uma técnica usada em GANs em que, a imagem de saída tem suas dimensões progressivamente aumentadas. Isso faz com que o esforço necessário para gerar uma imagem de alta qualidade seja diluído no tempo do treinamento, resultado em imagens mais estáveis. Além disso, os autores usaram a técnica de *Two-timescale update rule* (TTUR) de modo a compensar a aprendizagem lenta do discriminante, possibilitando um menor uso de etapas do gerador por etapa do discriminador, produzindo melhores resultados, tornando possível usar menos etapas do gerador por etapa do discriminador, produzindo melhores resultados. Após o treinamento da GAN proposto, Abdelhalim, Mohamed e Mahdy (2021) realizaram o treinamento de uma CNN do tipo ResNet-18 usando a técnica de aprendizagem por transferência para funcionar como um experimento dos resultados alcançados usando um conjunto de imagens composto pelo *dataset* HAM10000 e imagens sintetizadas. Os resultados utilizando o conjunto de imagens gerados pela GAN e HAM10000 encontrados pelos autores foi de 66,1% em acurácia, 79,3% em ROCAUC, 50,1% em precisão e 64,7% em *recall*, em contrapartida, dos valores de 60,6%, 76,2% 40,2% e 59,1% quando realizado o treinamento sem a adição das imagens sintetizadas.

Aggarwal, Das e Sreedevi (2019) usam a técnica de atenção para a melhora do resultado em classificação de imagens de pele utilizando CNNs. Nele, é proposto um método de atenção visando dar maior foco para as seções onde contêm a lesão de pele. Nesse trabalho, os autores utilizam a técnica para dar maior destaque aos *pixels* onde a lesão de pele está presente, bem como a redução do destaque das áreas onde não existe lesão de pele. Para isso, Aggarwal, Das e Sreedevi (2019) propõem um bloco de atenção que usa três convoluções para a redução de características, com tais características passando por uma ativação usando o método sigmoide, e então, passando por outra convolução para possuir as mesmas dimensões do mapa de características de entrada, e finalmente, passando por uma multiplicação, gerando assim o mapa de características com atenção. Na arquitetura de CNN desenvolvida pelos autores, esse bloco de atenção é inserido após a extração das características pela CNN, com seu mapa de características resultante sendo enviado para a parte totalmente conectada da CNN. Os autores aplicaram tal método de atenção em três diferentes tipos de CNN, sendo elas VGG16, InceptionV3 e

Xception. Para a verificação dos impactos obtidos com a aplicação do método de atenção, foi realizado o treinamento das CNNs usando múltiplos *datasets*, sendo eles HAM10000, DermNet, BioGPS e ISIC Images. A CNN que não usava a técnica de atenção que obteve melhor resultado foi a VGG16, com 74,7% de acurácia, 0,79 na métrica de precisão, 0,67 em *recall* e 0,73 na métrica *F1-Score*. Já recorrendo ao método de atenção, a rede InceptionV3 foi quem obteve melhores resultados, com acurácia de 85,8%, 0,87 em precisão, 0,86 em *recall* e 0,86 em *F1-Score*, mostrando uma melhora no desempenho da CNN quando utilizado o método de atenção.

Xue, Kamata e Luo (2021) propõem uma nova arquitetura de CNN, utilizando método de atenção para obter um melhor resultado na classificação de lesões de pele. A arquitetura proposta é composta por dois ramos. No primeiro ramo, é realizada a extração das características utilizando uma CNN padrão. Já no segundo ramo, os autores propõem um módulo batizado de DRPM, sigla para *Distinct Region Proposal Module*. Este módulo é composto por 4 blocos distintos, sendo eles: O bloco de atenção proposto, que utiliza os métodos de *channel-wise attention* e *spatial attention*. Então é usado o bloco de proposta da região de interesse, usado para criar as caixas delimitadoras. Tais caixas delimitadoras são enviados para o bloco de filtragem e pontuação, responsável por filtrar as caixas delimitadoras redundantes e/ou pouco descritivas. Por fim, existe um último bloco no módulo proposto, responsável por realizar a regulação do bloco de atenção usando como base as informações finais do bloco de filtragem. Então, neste segundo ramo, após definida as caixas delimitadoras da lesão de pele, é realizada a extração das características, essas multiplicadas com as obtidas no primeiro ramo, sendo então classificadas por uma rede totalmente conectada. De modo a melhorar o resultado obtido, os autores incluem no modelo proposto mais duas CNNs, sendo elas as redes ResNet-50 e VGG-16, para criar um modelo *ensemble*. Com este modelo final, treinado com o *dataset* ISIC 2017, Xue, Kamata e Luo (2021) obtiveram como resultados 89,5 de AUC para melanoma e 96,9 de AUC para não melanoma, resultados significativamente melhores que os obtidos por outros autores.

Além dos trabalhos citados acima, diversos outros estudos realizam uma comparação entre redes neurais convolucionais e dermatologistas para a classificação de lesões pigmentadas, como Esteva et al. (2017), Haenssle et al. (2018), Marchetti et al. (2018), Haenssle et al. (2020). Tais trabalhos mostram a viabilidade do uso de Redes Neurais Convolucionais para a classificação automática de lesões de pele pigmentadas a partir de imagens dermatoscópicas.

Tabela 3.1 – Comparação entre os trabalhos do estado da arte e a proposta do presente trabalho.

Autores	Uso de GAN	Uso de Método de Atenção	GAN Usada	CNN Usada
(VILLA-PULGARIN et al., 2022)	X	X	—	DenseNet-201
(NERSISSON et al., 2021)	X	X	—	YOLO
(QIN et al., 2020a)	✓	X	Style-based GAN	ResNet50
(ABDELHALIM; MOHAMED; MAHDY, 2021)	✓	X	SPGGANs	ResNet18
(AGGARWAL; DAS; SREEDEVI, 2019) (2019)	X	✓	—	InceptionV3
(XUE; KAMATA; LUO, 2021)	X	✓	—	WSFG-Net
Este Trabalho	✓	✓	StyleGAN2	EfficientNetB0

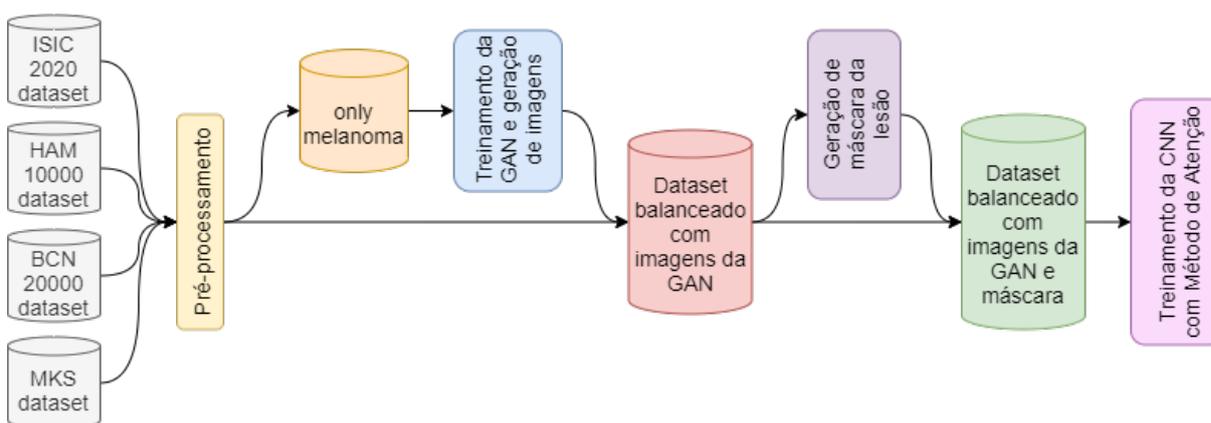
Fonte: Do Autor (2021).

Neste Capítulo foram apresentados os trabalhos relacionados a este trabalho de mestrado. A Tabela 3.1 mostra uma comparação entre os trabalhos levantados durante a fase de revisão de literatura. São apresentados trabalhos onde os autores usam GAN para a inclusão de novas imagens no conjunto de imagens, ou trabalhos que aplicam métodos de atenção para ressaltar detalhes da imagem recebida para classificação. Em ambos os casos, o uso de tais técnicas resultaram em uma melhora nas métricas de desempenho usadas. Entretanto, nenhum dos trabalhos relacionados anteriormente citados mostram o uso do processo de treinamento usando imagens sintéticas geradas por GAN e o emprego de um método de atenção em conjunto, mostrando o potencial do uso de ambas as técnicas combinadas. Sendo assim, mais uma vez se justifica o desenvolvimento deste projeto, onde trabalhos da área não abordam a classificação de imagens usando GAN e métodos de atenção. No próximo Capítulo, é apresentada a metodologia de desenvolvimento realizada neste trabalho.

4 MÉTODO E SOFTWARES USADOS

O trabalho proposto foi dividido em algumas etapas pré-definidas, servindo como marco para o desenvolvimento do trabalho. Inicialmente, foi realizada a etapa de aquisição e tratamento de imagens para composição do *dataset* usado para o desenvolvimento do trabalho. Após tal etapa ter sido finalizada com sucesso, e seu impacto validado empiricamente, foi realizado o treinamento de uma GAN, com o intuito de gerar um balanceamento no conjunto de imagens da etapa anterior. Em paralelo com esse treinamento, foi realizado um segundo treinamento, desta vez com uma rede do tipo U-Net, de modo a gerar as zonas de interesse da imagem, no caso deste trabalho, definir uma máscara que recorte a lesão de pele. Após tais passos terem sido completados com sucesso, a rede U-Net anteriormente treinada foi usada para a criação da máscara de recorte do *dataset* balanceado gerado a partir das imagens originais do conjunto de treinamento e das imagens sintéticas geradas pela GAN. Por fim, foi realizado o treinamento de uma rede baseada na CNN EfficientNet-B0, onde um método de atenção baseado em RoI foi empregado, de modo a melhorar o desempenho da rede proposta em relação à classificação de lesões de pele. A Figura 4.1 apresenta o esquema da metodologia usada neste trabalho, com cada etapa sendo explicada nas próximas seções.

Figura 4.1 – Metodologia usada para o desenvolvimento deste trabalho.



Fonte: Do Autor (2021).

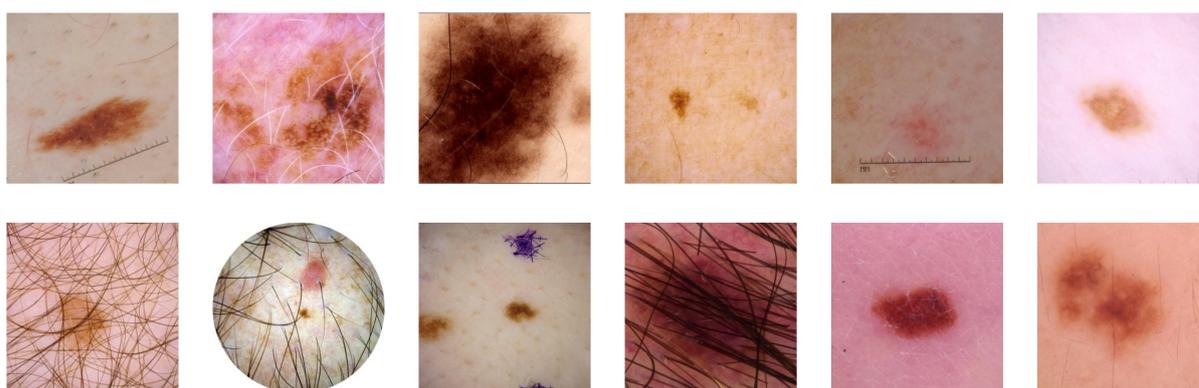
4.1 Conjunto de Dados

Para a realização do treinamento das redes neurais para classificação e/ou geração de imagens sintetizadas, é necessário o uso de uma base de dados padrão de modo a tornar possível a reprodutividade dos experimentos, normalmente chamados *dataset*. Para a realização

deste trabalho, foi usado não somente um *dataset*, porém, uma junção de alguns *datasets* populares na área de lesão de pele. Esses conjuntos de dados são disponibilizados pela *International Skin Imaging Collaboration* (ISIC), que organiza desafios de modo a fomentar trabalhos de inteligência artificial para lesão de pele. Para a construção do *dataset* utilizado, foram usadas imagens provenientes dos conjuntos *ISIC 2020* (ROTEMBERG et al., 2021), *HAM10000* (TSCHANDL; ROSENDAHL; KITTLER, 2018), *BCN20000* (COMBALIA et al., 2019) e *MKS Dataset* (CODELLA et al., 2019).

A junção destes conjuntos de dados totaliza um total de 59.039 imagens, gerados pela ISIC, com imagens provenientes do Hospital das Clínicas de Barcelona, Universidade Médica de Viena, Centro de Câncer Memorial Sloan Kettering, Instituto de Melanoma da Austrália, Universidade de Queensland, e Escola Universitária de Medicina de Atenas. Exemplos das imagens contidas no *dataset* podem ser observadas na Figura 4.2.

Figura 4.2 – Exemplo de imagens do conjunto de dados construído.



Fonte: (ROTEMBERG et al., 2021; TSCHANDL; ROSENDAHL; KITTLER, 2018; COMBALIA et al., 2019; CODELLA et al., 2019).

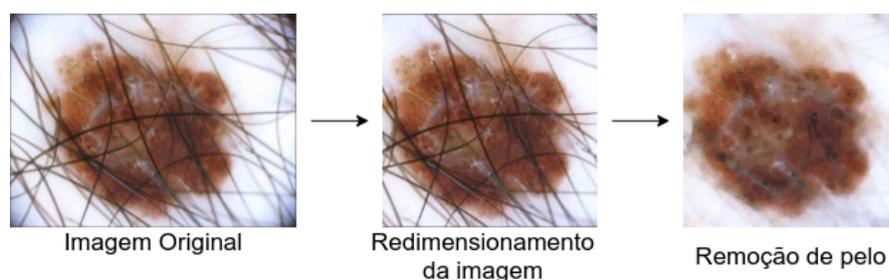
O conjunto de dados resultante possui um total de duas classes de imagens, sendo elas não melanoma, com um total de 53.352 imagens, e melanoma, com 5.637 imagens compondo o *dataset*. Através desses números, é possível ver claramente que o conjunto de dados criado é desbalanceado, ou seja, apresenta uma quantidade diferente de imagens para cada uma das classes, com a classe não melanoma contendo 9,46 vezes mais imagens que a classe melanoma.

4.1.1 Pré-processamento

Para o treinamento de uma rede neural, o *dataset* usado deve ser dividido em subconjuntos, sendo eles, treinamento e teste. O primeiro subconjunto, treinamento, é usado para efetivamente treinar a CNN, ou seja, é a partir dele que o modelo aprende. Já o subconjunto de

teste é usado para calcular as métricas de desempenho da rede neural. Isso é necessário devido ao fato de que, como a rede neural aprende com base em exemplos, é necessário que existam imagens desconhecidas pela rede, permitindo assim o cálculo de suas métricas reais. Neste trabalho, o conjunto de dados foi dividido em 80% para treinamento e 20% para teste.

Figura 4.3 – Processamento realizado sobre o conjunto de imagens criado.



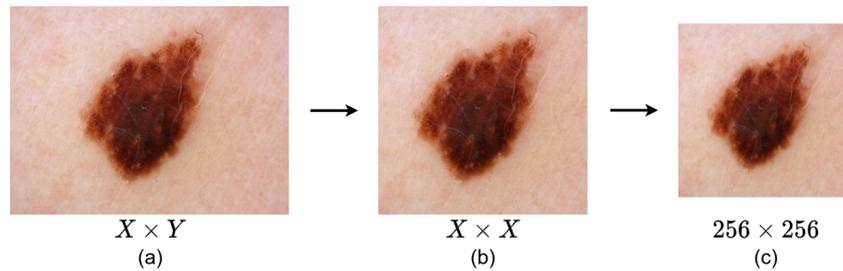
Fonte: Do Autor (2021).

Além disso, outras formas de processamento foram aplicadas para o tratamento das imagens de entrada para a GAN e para a CNN. A Figura 4.3 mostra esses passos, sendo eles, o redimensionamento das imagens para um tamanho determinado e remoção de pelo, resultando assim na imagem final, usada para o treinamento das redes neurais.

Inicialmente, como as imagens que compõe esse *dataset* são provenientes de diversas fontes, o tamanho das imagens era diferente. Logo, algum processamento era necessário ser realizado de modo a deixar o conjunto com todas as imagens com as mesmas dimensões. Para isso, foi criado um *script* em linguagem Python usando a biblioteca Pillow¹. Esse algoritmo lê a imagem original, bem como suas dimensões. De posse das dimensões da imagem, a imagem passa por um corte centralizado de tamanho $W \times W$, onde W é a menor dimensão da imagem, gerando assim uma imagem quadrada. Após esse corte, o tamanho da imagem era redimensionado para as dimensões 256×256 , usando interpolação bicúbica. A realização do corte centralizado se faz necessária dado que, como a imagem original poderia não ser quadrada, somente a aplicação do redimensionamento causaria uma imagem com proporções distorcidas. Sendo assim, após a execução desse *script* sobre o *dataset* por completo, todas as imagens que o compõe tinham dimensões 256×256 . Na Figura 4.4 é possível constatar as imagens resultantes de cada um dos passos até chegar à imagem final. A Figura 4.4 (a) mostra a imagem original, enquanto a imagem (b) apresenta a imagem após passar pelo corte centralizado de tamanho $W \times W$, e, por fim, a imagem (c) mostra a imagem após passar pelo processo de redimensionamento, para o tamanho final.

¹ Disponível em: <<https://github.com/python-pillow/Pillow>>

Figura 4.4 – Exemplo do processo de redimensionamento da imagem.

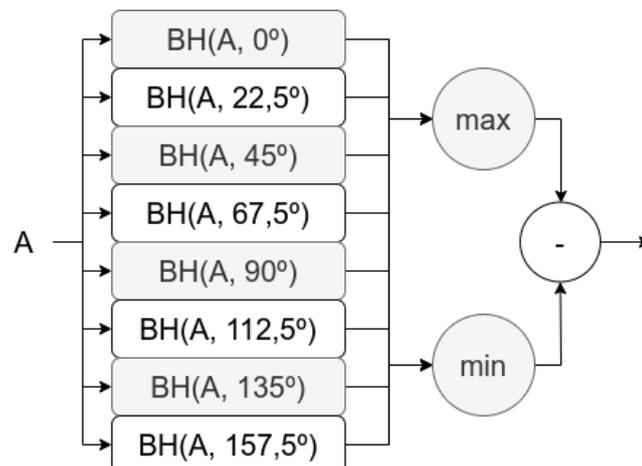


Fonte: Do Autor (2021).

Após a imagem já se encontrar com o tamanho correto, foi realizado um processamento de modo a remover os pelos da imagem. Como as lesões podem ocorrer em toda a pele, em alguns casos pode ocorrer de existir pelo na área da lesão, causando assim uma obstrução na imagem da lesão pigmentada. Sendo assim, de modo a resolver este problema, foi usado um algoritmo para remoção de pelos, baseado no trabalho de Bibiloni, González-Hidalgo e Massanet (2017). Este algoritmo é composto por dois principais passos: um detector de objetos para imagens coloridas, e por fim, um algoritmo de *inpainting* também para imagens coloridas. A fase de detecção de objetos é realizada através da combinação de transformações do tipo *soft color top-hat*. São usadas 8 transformações, cada uma usando uma orientação diferente dos filtros aplicados, sendo eles $0^\circ, 22,5^\circ, 45^\circ, \dots, 157,5^\circ$. De posse dessas transformações realizadas, a transformação com mais e a com menos elementos são escolhidas, realizada a subtração entre as duas. A Figura 4.5 mostra um esquemático do detector usado.

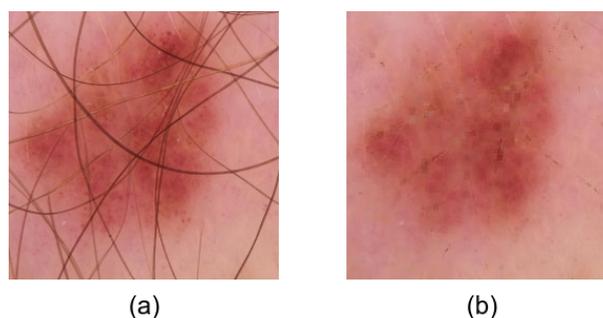
Com os pelos detectados, é realizada a remoção de tais pelos da imagem através da aplicação de um algoritmo de *inpainting* baseado em *soft color morphology*. A Figura 4.6 mostra uma imagem antes e após passar pelo processo de remoção de pelos que é aplicado em todo o *dataset*.

Figura 4.5 – Esquemático do detector usada para detecção de pelos.



Fonte: Do Autor (2021).

Figura 4.6 – Comparação entre uma imagem dermatoscópica após a remoção de pelos.



Fonte: Do Autor (2021).

4.2 Generative Adversarial Network

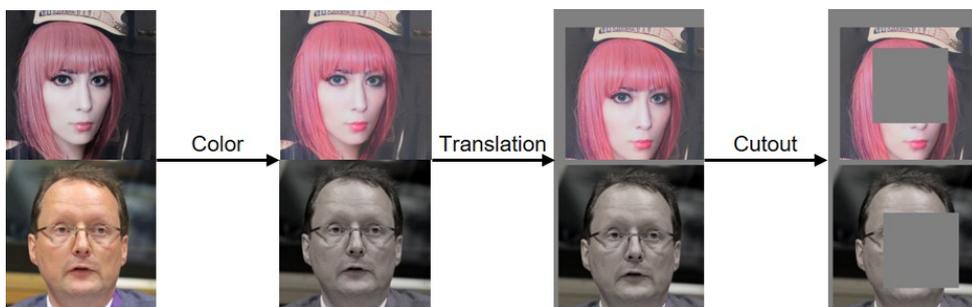
Como mostrado anteriormente, o conjunto de dados usados para classificação de lesões de pele é desbalanceado. Segundo Aldwgeri e Abubacker (2019), um dos principais problemas enfrentados no aprendizado de máquina é o desbalanceamento de dados. O desbalanceamento de dados ocorre quando a quantidade de imagens por classe de um *dataset* é diferente. O desbalanceamento de dados afeta diretamente o desempenho do método de aprendizagem de máquina, tendendo a focar somente na classe com mais exemplos, ignorando as outras classes (MA; XU; ZHANG, 2018). É importante salientar que no ramo médico, existem desafios e limitações associadas à coleta de imagens, o que acaba gerando um desbalanceamento no *dataset*. Em primeiro lugar, a coleta de dados leva muito tempo. Em segundo lugar, a coleta de grandes volumes de dados sobre pacientes com certas doenças costuma ser um desafio devido à baixa prevalência da doença (MAZUROWSKI et al., 2008).

De modo a contornar esse problema, uma abordagem que tem ganhado força é o uso de redes adversárias generativas, as chamadas GANs, do inglês *Generative Adversarial Network* para sintetização de imagens de modo a balancear o *dataset*. Tal técnica tem sido usada em diversos ramos, incluindo o uso em *datasets* médicos sobre lesão de pele, visto que, como o câncer do tipo melanoma é menos comum que o não melanoma, a quantidade de exemplares de imagens dermatoscópicas é menor, entretanto, a gravidade do câncer melanoma é maior, logo, a acurácia de um modelo de inteligência artificial para essa categoria de lesão deve ser alta.

Para a realização do processo de treinamento da GAN, foi criado um *dataset*, chamado *only_melanoma*. Este conjunto contém somente as imagens de câncer de pele do tipo melanoma do *dataset* mencionado na Seção 4.1. A escolha no uso de todas as imagens disponíveis de modo a treinar a GAN se deve ao fato de que, com um número maior de imagens, a qualidade da imagem sintetizada pela GAN é maior.

Entretanto, a quantidade de imagens disponíveis no *dataset only_melanoma* ainda é considerada pequena, sendo um conjunto de dados com menos de 10 mil imagens. Para amenizar esse problema, para este trabalho foi utilizada a rede StyleGAN2 em conjunto com *Differentiable Augmentation* (DiffAugment), proposto por Zhao et al. (2020). Neste trabalho, foi proposto o uso de técnicas de *data augmentation* para gerar um número maior de imagens para as redes geradoras e discriminatórias que compõe a GAN. A Figura 4.7 mostra um exemplo de quais operações são realizadas sobre as imagens do *dataset*.

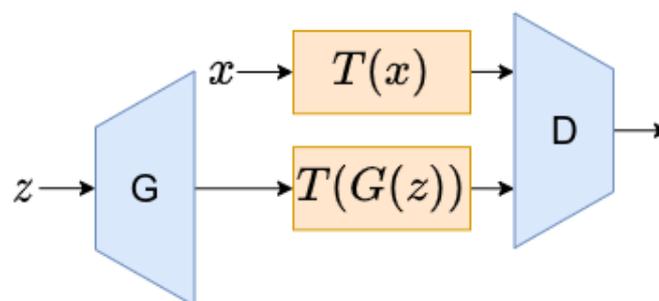
Figura 4.7 – Exemplo das operações realizadas sobre as imagens para a GAN.



Fonte: Zhao et al. (2020).

A Figura 4.8 mostra a arquitetura da GAN usada para a síntese das imagens de lesão de pele. Nela, o gerador G é responsável por gerar as imagens de melanoma a partir dos valores aleatórios z . Então, a partir dessas imagens, é aplicada a técnica de *data augmentation* T , demonstrada na Figura 4.7, resultando assim no conjunto de imagens $T(G(z))$. Essa mesma função de aumento de imagens é aplicada sobre as imagens reais de treinamento x , gerando assim o conjunto $T(x)$. Então, estes são os conjuntos de imagens enviados para o discriminante D , responsável por classificar as imagens como reais ou falsas. A aplicação da função de transformação T evita o *overfitting* do discriminante D , evitando assim ele tender a uma classe (reais/falsas), consequentemente levando a um melhor resultado da seção generativa da GAN.

Figura 4.8 – Arquitetura da GAN usada para a geração das imagens sintetizadas.



Fonte: Do Autor (2021).

De modo a acelerar o processo de treinamento da GAN, foi usada como ponto de partida uma GAN já treinada para a geração de imagens de rostos. Isso se deve ao fato de que, com

a rede iniciando o treinamento para lesão de pele a partir de sintetização de rostos, pequenas *nuances* como cor e textura de pele são “reaproveitadas”, diminuindo assim o tempo necessário para atingir uma determinada métrica se comparado com uma GAN iniciando a partir de pesos aleatórios.

Para a realização do treinamento da GAN, foi utilizada uma máquina equipada com 40 núcleos de processamento Intel Xeon E5-2650 v3, executando a 2,30 GHz, com um total 62 GB de memória RAM, além de ser equipado com uma placa gráfica Nvidia Tesla K20c, com 5 GB de memória, chegando a 3.524 TFLOPS em precisão simples. A resolução, tanto das imagens de entrada quanto de saída foi configurada para 256×256 pixels, com um *batch size* de 8, com o treinamento sendo finalizado quando a GAN chegasse à 3 milhões de imagens processadas. Como otimizador foi usado o algoritmo Adam (KINGMA; BA, 2015), *Leaky ReLU* como função de ativação, um método de regulação baseado em *style-mix* (KARRAS; LAINE; AILA, 2019). Tais parâmetros são padrões para o treinamento das redes StyleGAN2.

Após o treinamento da GAN, foi executado o algoritmo responsável pela geração das imagens sintéticas de lesões de pele. Tal algoritmo consiste em utilizar somente a parte geradora da GAN, obtendo assim imagens sintéticas de lesões de pele. De modo a igualar a quantidade de imagens de melanoma com imagens de não melanoma, foi gerado um total de 48.247 imagens. Com essa quantidade de imagens extra, o número de imagens presentes na classe melanoma e não melanoma se torna igual, obtendo assim um balanceamento no *dataset*.

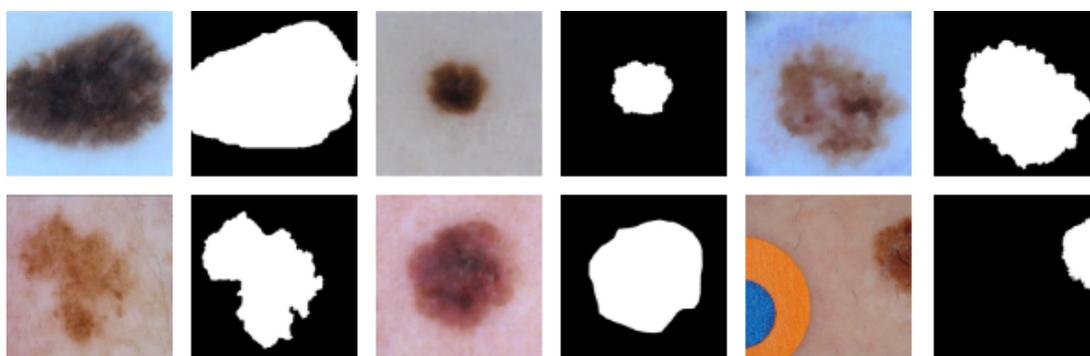
Neste trabalho foi utilizada como principal métrica para a GAN a métrica FID, devida sua maior robustez em relação a ruído, bem como seu melhor desempenho na medida de diversidade de imagens.

4.3 Segmentação de lesões de pele

Para a realização do processo de treinamento da rede de segmentação, foi necessária a obtenção de um novo conjunto de dados. Isso se deve ao fato de que, como o processo de treinamento de uma CNN do tipo U-Net é do tipo supervisionado, deve existir para cada entrada do *dataset* x , um valor de y correspondente, onde y é a máscara da lesão, no caso deste trabalho. Diversos *datasets* podem ser usados com essa finalidade, como ISIC 2017 (CODELLA et al., 2018), ISIC 2018 (CODELLA et al., 2019) e PH² (MENDONCA et al., 2013).

Para este trabalho, foi usado o conjunto de dados ISIC 2017. Isso se deve ao fato de que este *dataset* é usado em diversos trabalhos da área de segmentação de lesões de pele, permitindo assim uma melhor comparação dos resultados obtidos no processo de treinamento dessa etapa. O ISIC 2017 é composto por 2.000 imagens, cada uma contendo a sua respectiva máscara de recorte. Além disso, o processo de pré-processamento das imagens, descrito na Subseção 4.1.1 foi aplicado, ou seja, as imagens tiveram seus tamanhos reajustados para 256×256 *pixels*, além da realização do procedimento para remoção de pelo. A Figura 4.9 mostra um exemplo do conjunto de dados, bem como suas respectivas máscaras.

Figura 4.9 – Exemplo das imagens já processadas do *dataset* ISIC 2017.

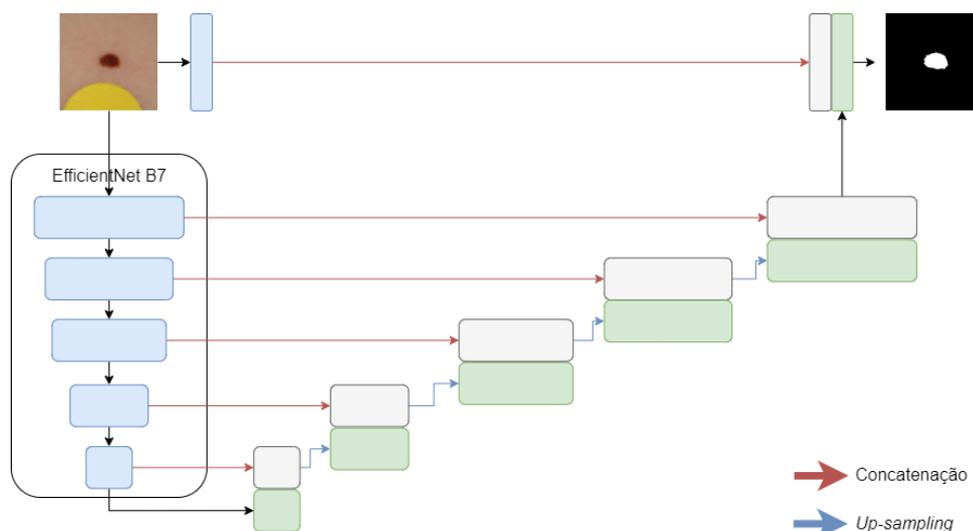


Fonte: Do Autor (2021).

Para o desenvolvimento deste trabalho, foi implementada uma CNN baseada na rede U-Net, usando como parte decodificadora a rede neural convolucional EfficientNet-B7 como *backbone*. Isso porque, dada as características dessa CNN, como velocidade devido ao reduzido número de parâmetros e robustez na classificação, sua utilização como *backbone* traria uma melhora no desempenho final da CNN U-Net. Para o desenvolvimento do trabalho, foram escolhidos os filtros de tamanho 64, 128, 256, 512 e 1024. Além disso, cada bloco de *up-sampling* é composto por duas camadas convolucionais, além de um processo de interpolação bilinear. A função de ativação ReLU é usada nas camadas internas da CNN, enquanto é usada uma função sigmoide como função de ativação de saída da U-Net. A Figura 4.10 mostra um diagrama simplificado da rede implementada.

Para o processo de treinamento da U-Net, foi escolhida como função de perda *Jaccard Distance loss* (RAHMAN; WANG, 2016), além do otimizador Adam (KINGMA; BA, 2015), com taxa de aprendizado configurada como 10^{-4} e decaimento de peso (LOSHCHILOV; HUTTER, 2017) com taxa de 10^{-5} , usando a política de um ciclo (SMITH, 2018), usando como tamanho do lote 32 e executando por um total de 150 épocas.

Figura 4.10 – Diagrama da rede U-Net implementada.



Fonte: Do Autor (2021).

4.4 Mecanismo de Atenção e EfficientAttentionNet

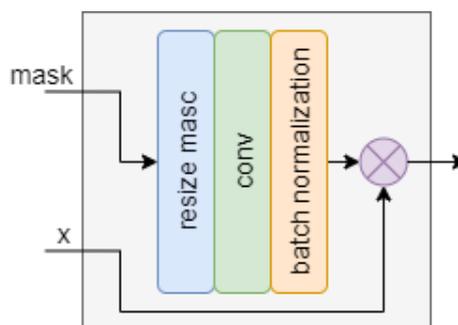
Diversas técnicas têm sido propostas para melhorar o desempenho de redes neurais implementadas para classificação e segmentação de imagens. Dentre essas técnicas, se destaca a técnica de atenção (*attention*), cujo objetivo é definir uma área de interesse de um objeto em questão na imagem de entrada e então realizar um destaque de tal objeto, fazendo com que somente a área de interesse seja considerada para a classificação.

Tal técnica é uma proposta inteligente para o problema de classificação de lesões de pele, dado que, as imagens recebidas como entrada possuem diversos elementos que podem interferir no treinamento da CNN, como pelos, marcações médicas, régua para medição do tamanho da lesão, dentre outras. Assim, a utilização de um método que consiga dar atenção somente à lesão, deixando de lado esses outros objetos, é de grande valia.

Neste trabalho, foi proposto um bloco de atenção baseado em RoI, que utilizaria como entrada a máscara da lesão de pele gerado pela rede U-Net previamente treinada visando fazer com que a rede neural convolucional focasse somente na parte com lesão, abstraindo o restante da imagem. Para isso, foi criado o bloco de atenção mostrado na Figura 4.11.

Como é possível observar pela Figura 4.11, o bloco de atenção é composto por três componentes básicos: uma camada de redimensionamento, uma camada convolucional e uma camada de *batch normalization*. O processamento do bloco de atenção proposto é o seguinte: inicialmente a máscara de entrada é redimensionada para possuir as mesmas dimensões de altura A e largura L do mapa de características x . Após isso, a máscara passa por uma camada

Figura 4.11 – Bloco de atenção proposto.



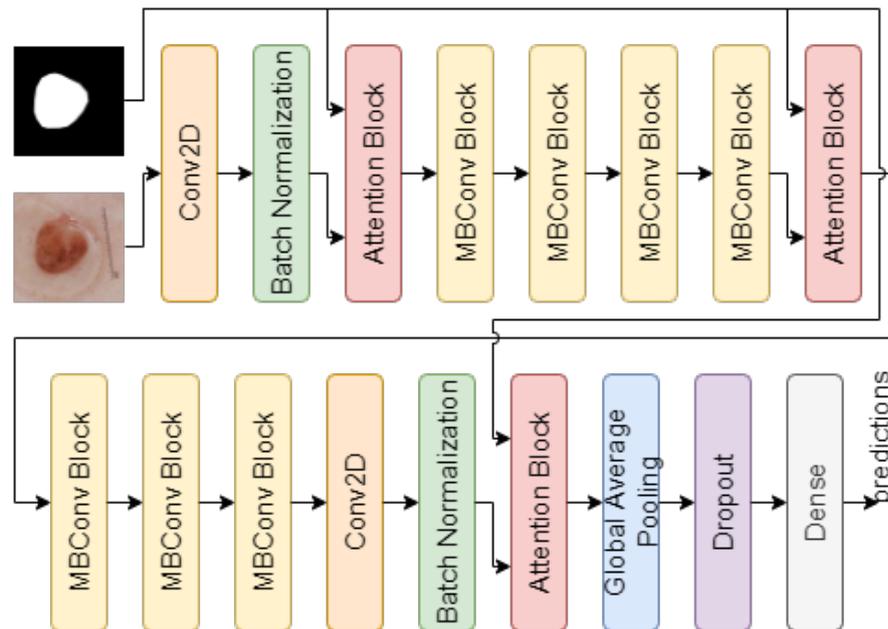
Fonte: Do Autor (2021).

de convolução com a mesma quantidade de canais que a entrada x , gerando assim um mapa de atenção com as mesmas dimensões que o mapa de característica x . Então, o mapa de atenção é passado por uma camada de *batch normalization*, visando diminuir possíveis sobreajustes do bloco de atenção. Por fim, é realizada a multiplicação do mapa de atenção com o mapa de características de entrada.

A ideia por trás do bloco de atenção proposto é fazer com que a rede neural convolucional tenha seus mapas de características filtrados, removendo objetos que não sejam de interesse para a classificação de lesão de pele, como pelos, pele, marcações médicas de referência, géis aplicados sobre a pele, entre outros objetos que não sejam a lesão de pele especificamente dito.

Para a implementação desse trabalho, foi utilizada como base a CNN EfficientNet-B0 para a classificação das lesões de pele usando o bloco de atenção proposto, dando origem assim a uma arquitetura de CNN batizada como EfficientAttentionNet. Nessa arquitetura, duas entradas são recebidas pela CNN: a imagem da lesão de pele e a sua respectiva máscara, gerada pela rede U-Net. Além disso, foram adicionados três blocos de atenção, um após a primeira camada de convolução, de modo a remover as primeiras características gerais da imagem oriundas do primeiro processamento gerado pela camada convolucional. O segundo bloco de atenção é inserido após o quarto bloco MBConv, de modo a novamente delimitar a zona de interesse da imagem que está sendo processada. Finalmente, o terceiro e último bloco de atenção é inserido após a última camada convolucional da CNN, antes da operação de *Global Average Pooling*, para trazer maior atenção às características da lesão de pele, antes de realizar a classificação da lesão. Após a camada de *Global Average Pooling*, é usada uma camada de *Dropout* configurado com taxa de 0,5, para diminuir o sobre-ajuste que possa existir no treinamento da CNN, melhorando assim sua generalização. Por fim, é usada uma camada totalmente conectada configurada com a função de ativação sigmoide, de modo a definir se a lesão é melanoma ou não melanoma. A Figura 4.12 mostra uma representação da arquitetura final da CNN proposta neste trabalho.

Figura 4.12 – Arquitetura da CNN proposta.



Fonte: Do Autor (2021).

De posse da CNN para classificação já desenvolvida, foi dado início ao processo de treinamento da rede EfficientAttentionNet para classificação de lesão de pele. De modo a obter um número ainda maior de imagens, foi utilizada a técnica de *data augmentation*.

Para este trabalho, foram utilizadas as técnicas de rotação em 90 graus, inversão da imagem, tanto na horizontal quanto na vertical, transposição da imagem, rotação aplicando deslocamento, bem como uma distorção em *grid*. Todas essas operações foram realizadas utilizando a biblioteca *albumentations*, disponibilizada por Buslaev et al. (2020). Além disso, de modo a evitar que os valores de cor de cada *pixel* fizesse com que os pesos da CNN aumentassem muito, foi realizada uma normalização da imagem de entrada em duas etapas: na primeira etapa, o valor de cor de cada *pixel* era normalizada de $0 \dots 255$ para $0 \dots 1$. Além disso, é importante frisar que, tanto a imagem como sua respectiva máscara sofre as mesmas operações com os mesmos valores, garantindo assim que a máscara condiz com a localização da lesão de pele de sua respectiva imagem.

A rede neural proposta foi treinada utilizando o otimizador Adam (KINGMA; BA, 2015). De modo a diminuir o efeito de sobre-ajuste, foi utilizado o método de decaimento de pesos sobre o otimizador Adam (LOSHCHILOV; HUTTER, 2017), usando taxa de decaimento de 10^{-5} . A busca por um valor de taxa de aprendizagem para o otimizador pode ser um desafio, uma vez que uma taxa pequena pode levar ao sobre-ajuste, enquanto uma taxa grande pode fazer que com que o treinamento seja demorado. Por isso, foi utilizado o método de taxa

de aprendizagem cíclica, proposto por (SMITH, 2017), onde o valor da taxa de aprendizagem e o momento do otimizador varia durante o treinamento. Os valores de taxa de aprendizagem podem ir de 10^{-5} a 10^{-4} seguindo a política de um ciclo. Já o momento do otimizador varia de 0,85 a 0,95. O tamanho do lote de imagens a serem processados foi de 32 imagens. Já a quantidade de épocas para o treinamento foi de 40. Tais parâmetros foram escolhidos baseados nas recomendações de Smith (2018). O *dataset* usado para seu treinamento é um conjunto de diversos *datasets*, além da adição de imagens sintéticas geradas pela GAN treinada anteriormente para o balanceamento de imagens, além das máscaras da lesão de pele definidas pela rede U-Net.

4.5 Software utilizados

Para o desenvolvimento deste projeto, diversas ferramentas de *software* foram utilizadas para a implementação e treinamento de redes neurais convolucionais. Na Subseção 4.5.1 é mostrada a linguagem de programação utilizada para o desenvolvimento do trabalho, enquanto nas Subseção 4.5.2 e Subseção 4.5.3 é tratado sobre as bibliotecas utilizadas.

4.5.1 Python

A linguagem utilizada para desenvolver o trabalho foi Python. Python começou a ser desenvolvida em 1990 por Guido van Rossum, durante o tempo que trabalhou no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda (CWI), tendo foco em usuários como físicos e engenheiros. Foi baseada em outra linguagem da época, chamada ABC (COELHO, 2007).

Segundo Borges (2010), a linguagem Python possui uma sintaxe clara e concisa, favorecendo a legibilidade do código-fonte, a tornando mais produtiva. Ainda segundo ele, possui várias estruturas de alto nível, como listas, dicionários, números complexos, e uma grande coleção de módulos prontos, além dos *frameworks* de terceiros que podem ser usados.

Python também apresenta características de outras linguagens modernas, como geradores, metaclasses e unidades de testes. Ela também é multiparadigma, suportando programação modular, funcional e também orientação a objetos. Python é uma linguagem interpretada através de *bytecode* pela máquina virtual Python, o que acaba tornando o código portátil (BORGES, 2010).

Além disso, Python é um *software* de código aberto, utilizando uma licença compatível com a GPL (*General Public License*), porém menos restritiva, permitindo seu uso em produtos proprietários, por exemplo. A especificação da linguagem é mantida pela *Python Software Foundation* (PSF). Além de seu uso como principal linguagem de programação, Python também é usada como linguagem de *script* em vários *softwares*, automatizando tarefas e adicionando novas funções (BORGES, 2010).

Coelho (2007) lista algumas características da linguagem Python que a torna interessante, como ser multiplataforma, podendo ser executada em praticamente qualquer plataforma computacional; portátil, com aplicativos Python podendo ser facilmente distribuídos para outras plataformas; ser um *software* livre, não realizando limitação à distribuição gratuita e venda de programas; extensível, podendo ser feito o uso de rotinas escritas em outras linguagens, como C e Fortran; tipagem automática, onde o tipo da variável é definido em tempo de execução, não sendo necessário explicitá-lo dentre outras. A linguagem Python foi escolhida por possuir estas características, além de ser amplamente utilizada para aplicações de aprendizagem de máquina, com diversos *frameworks* possuindo suporte à linguagem Python.

4.5.2 TensorFlow

TensorFlow é uma biblioteca escalável e multi-plataforma para a implementação e execução de algoritmos de aprendizagem de máquina, incluindo abstrações para aprendizagem profundo (RASCHKA, 2017).

Segundo Singh (2020), é uma das bibliotecas mais utilizadas para aprendizagem de máquina, criada pelo Google como parte do projeto Google Brain, porém, com diversas contribuições da comunidade de código aberto. TensorFlow foi inicialmente construída para uso interno do Google Brain, porém, em 2015 foi liberado para uso sobre uma permissão de código aberto (RASCHKA, 2017).

Para melhorar o desempenho de treinamento de modelos de aprendizagem de máquinas, TensorFlow permite a execução tanto em CPUs quanto GPUs, porém, é importante ressaltar que o melhor desempenho é encontrado utilizando GPUs (RASCHKA, 2017).

O TensorFlow atualmente oferece suporte às várias linguagens de programação, porém, a linguagem Python é a que mais se destaca, principalmente por possuir uma API (Interface de Programação de Aplicações, do inglês, *Application Programming Interface*) mais completa, o

que atrai diversos profissionais de aprendizagem de máquina e aprendizagem profundo. Outra API que se destaca é a API para C++ (RASCHKA, 2017).

Raschka (2017) diz que os cálculos do TensorFlow dependem da construção de um grafo direcionado para a representação de um fluxo de dados. A construção de tal grafo pode ser difícil, porém, TensorFlow possui APIs de alto nível, que permite a construção de modelos de aprendizagem de máquina da forma muito fácil.

4.5.3 ktrain

De modo a tornar o desenvolvimento dos modelos de aprendizagem de máquina mais simplificado, foi utilizada a biblioteca para aprendizagem de máquina ktrain. É uma biblioteca visando apresentar uma interface simples e unificada para facilitar as etapas para construção de modelos de aprendizagem de máquina. A biblioteca ktrain pode ser utilizada para a implementação de qualquer modelo usando o TensorFlow (MAIYA, 2020).

A biblioteca ktrain ainda possui outros recursos, como um localizador de taxa de aprendizagem, acesso fácil à programação da taxa de aprendizagem, como a política de um ciclo (SMITH, 2018), descida de gradiente estocástico com reinícios (SGDR, do inglês, *Stochastic Gradient Descent with Restarts*) (LOSHCHILOV; HUTTER, 2016), bem como otimizadores do estado da arte, como AdamW (LOSHCHILOV; HUTTER, 2017), além de uma API simples para previsão e validação do modelo (MAIYA, 2020).

4.5.4 Google Colab

Para o treinamento das redes neurais convolucionais foi utilizado o ambiente de desenvolvimento *online* Google Colaboratory, ou “Colab”². O Google Colab permite o desenvolvimento de código Python no próprio navegador, através da execução de um *notebook* Jupyter.

O Google Colab possui como um ponto positivo uma grande quantidade de bibliotecas já instaladas, principalmente para ciência de dados e aprendizagem de máquina. Através do Google Colab, é possível a criação de *scripts* usados para processamento e visualização de dados, usando as principais bibliotecas Python, como *numpy*, *pandas* e *matplotlib*.

Outro ponto de grande destaque do Google Colab é a disponibilização de GPUs e TPUs de alto desempenho gratuitamente para a execução de algoritmos de aprendizagem de máquina.

² Disponível em: <<https://colab.research.google.com/>>

Com a conexão automática de um *notebook* com o Google Drive, é possível realizar o carregamento de um conjunto de dados para o ambiente de desenvolvimento, realizar o treinamento de um classificador usando bibliotecas já instaladas, como TensorFlow e Torch, realizar a avaliação de desempenho, além de salvar resultados intermediários novamente para o Google Drive.

Neste trabalho, o Google Colab foi usado para treinamento das redes neurais convolucionais desenvolvidas, visto que o seu uso permite um treinamento mais rápido, dada a capacidade e desempenho das GPUs disponibilizadas gratuitamente.

4.6 Considerações Finais

Neste Capítulo, foram demonstrados os processos realizados para o desenvolvimento deste trabalho, sendo mostrado em um primeiro momento o processo de aquisição e pré-processamento do conjunto de dados utilizado. Este pré-processamento consistia em duas etapas, uma onde a imagem de origem era redimensionada para um tamanho reduzido, que seria usado tanto pela GAN quanto pela CNN, bem como a remoção de pelo.

Com o conjunto de imagens já processado, foi realizado o treinamento da rede GAN para a geração de imagens sintéticas, de modo a realizar o balanceamento do conjunto de imagens utilizado. Posteriormente, foi apresentada a rede U-Net utilizada para a geração das máscaras de recorte usadas para o método de atenção proposto, bem como o *dataset* ISIC 2017.

Então, é demonstrado o método de atenção baseado em RoI proposto, que usa o mapa das características recebido e da máscara da lesão para delimitar qual a área que terá mais foco para a classificação da imagem. Também é mostrada a CNN EfficientNet, usada de base para a construção da CNN proposta batizada de EfficientAttentionNet.

Além disso, são mostrados os *softwares* usados para o desenvolvimento deste trabalho, como a linguagem Python, o *framework* para desenvolvimento de aplicações de *deep learning* TensorFlow e a biblioteca ktrain. Além disso, é apresentado o ambiente de desenvolvimento e execução em nuvem Google Colab, usado para o treinamento das redes apresentadas neste trabalho. No próximo Capítulo são apresentados os resultados deste trabalho, sendo feita uma análise das métricas de desempenho de cada uma das etapas, bem como o impacto da inclusão de cada uma no fluxo de desenvolvimento de uma aplicação para classificação de lesão de pele.

5 RESULTADOS

Este Capítulo aborda os resultados encontrados durante o processo de desenvolvimento deste trabalho. Na Seção 5.1 é mostrada uma análise do impacto que o pré-processamento aplicado neste trabalho. Já na Seção 5.2 são mostrados os resultados alcançados no treinamento da rede generativa adversária para a geração de imagens sintéticas de lesão de pele. A Seção 5.3 mostra o desempenho alcançado pela rede U-Net implementada, mostrando as métricas alcançadas e exemplos práticos de como as máscaras são geradas. Por fim, a Seção 5.4 mostra os resultados alcançados utilizando todos os métodos já citados, bem como a aplicação do método de atenção proposto em uma CNN EfficientNetB0, mostrando a evolução de cada etapa do trabalho.

5.1 Resultados da remoção de pelo

Antes de realizar o treinamento das redes neurais artificiais do tipo GAN e a EfficientNet, foram realizados testes de modo a verificar o impacto que o algoritmo de remoção de pelo causaria no desempenho das redes, servindo como parâmetro para a continuidade do seu uso ou não.

Para isso, foi criado um conjunto de dados usando as imagens coletadas do desafio ISIC, porém, diferente dos demais conjuntos de dados, esse tinha como diferencial a quantidade de imagens pertencentes a cada classe. Como o objetivo deste experimento era verificar o impacto da remoção de pelo, o desbalanceamento do *dataset* foi removido, ou seja, a quantidade de imagens pertencentes em cada classe era a mesma.

Para isso, foi inicialmente calculado quantas imagens pertencentes à classe melanoma existiam no *dataset*. De posse disso, foi realizado o embaralhamento das imagens de não melanoma, de modo a remover qualquer viés existente, e então, a mesma quantidade de imagens era selecionada. Então, de posse desse novo conjunto, foi criada uma cópia do mesmo, porém, com o algoritmo de remoção de pelos aplicado sobre ele.

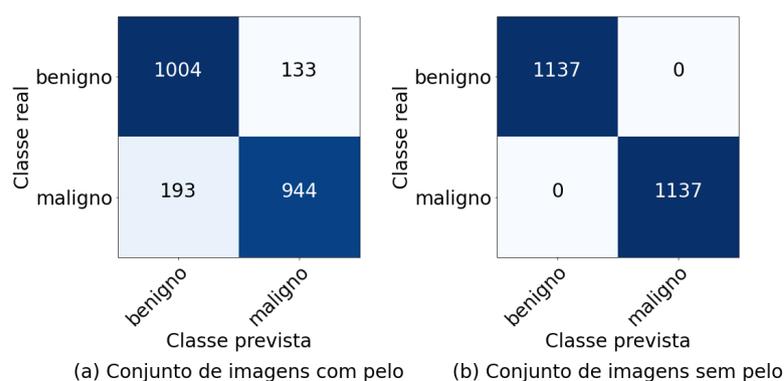
Tabela 5.1 – Comparação do desempenho da CNN com imagens com e sem pelo.

<i>Dataset</i> usado	Acurácia	Recall	Precision	ROC AUC
Com pelo	0,856640	0,834476	0,878645	0,859371
Sem pelo	1,0	0,986111	0,986111	0,986111

Fonte: Do Autor (2021).

Como pode ser observado pela Tabela 5.1, a rede EfficientNet treinada com o conjunto de imagens sem pelos, obteve resultados superiores aos alcançados com a mesma arquitetura usando imagens com pelo. Em relação ao *dataset* com pelo, a melhora nas métricas acurácia, *recall*, precisão e ROC AUC foi de 16,73%, 18,17%, 12,23% e 14,74%. A Figura 5.1 mostra as matrizes de confusão de cada um dos treinamentos. Para a criação destas matrizes, foi usado como limiar de arredondamento o valor 0,5, onde saídas provenientes da CNN maiores ou iguais a este valor, eram mapeadas para 1, ou seja, melanoma.

Figura 5.1 – Comparação entre as matrizes de confusão dos treinamentos realizados com e sem pelo no *dataset*.



Fonte: Do Autor (2021).

Uma explicação para tal resultado, é que, sem pelo, a identificação da lesão é facilitada, dado que os pelos podem ocultar partes da lesão. Além disso, outra possibilidade é que a existência de pelos na imagem possa causar uma falsa correlação. A classificação de imagens com pelo poderiam tender mais à classe não melanoma, correlação que pode não estar correta.

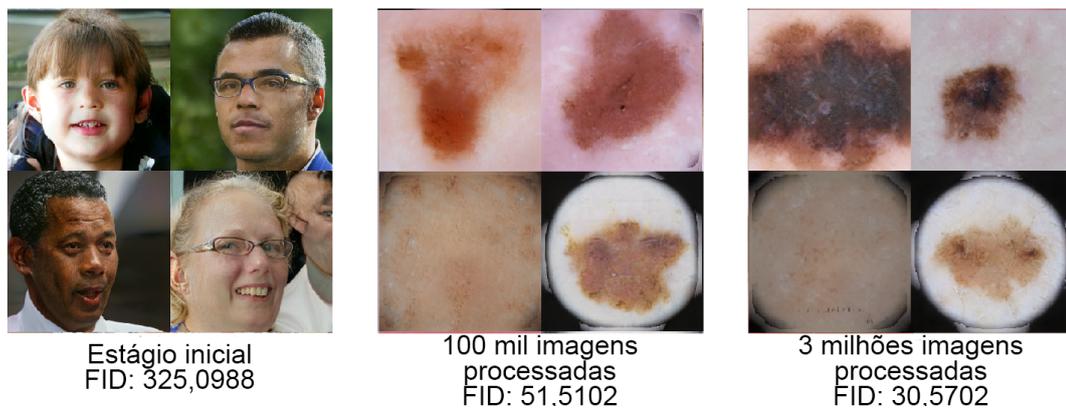
Sendo assim, os resultados apresentados nessa Seção mostram que a utilização de algoritmos para a remoção de pelos de imagens dermatoscópicas aumenta o desempenho da CNN em relação às métricas de classificação.

5.2 Resultados da GAN

Após o treinamento da rede adversária generativa, foi realizada a avaliação da qualidade e da diversidade das imagens geradas a partir das imagens de melanoma do conjunto de dados criado. A Figura 5.2 mostra a evolução das imagens geradas pela GAN. Já a Figura 5.3 mostra a evolução da GAN em relação à métrica FID.

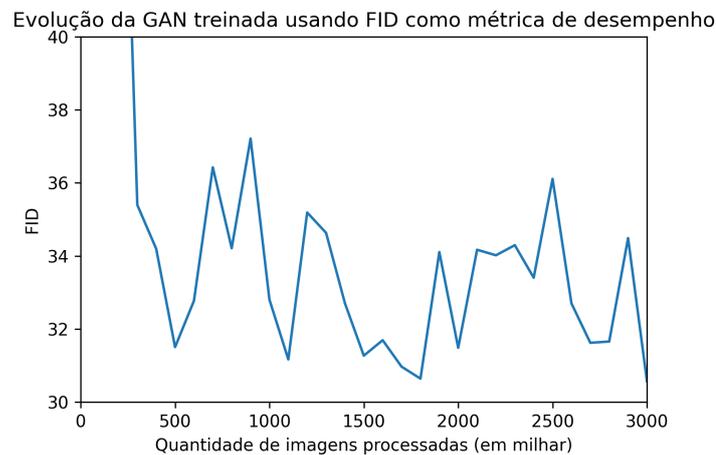
Como pode ser observado na Figura 5.2, no estágio inicial, a GAN gerava imagens de rostos humanos, dado que era um modelo pré-treinado com o *dataset* FFHQ. Porém, após o

Figura 5.2 – Exemplos de resultados da GAN com a métrica FID.



Fonte: Do Autor (2021).

Figura 5.3 – Valores da métrica FID durante o processo de treinamento da GAN.



Fonte: Do Autor (2021).

começo do processo de treinamento, a GAN começou a gerar imagens de lesões de pele. Nesse primeiro momento, alcançou FID de aproximadamente 50.

Após isso, a GAN manteve o seu FID entre 30 e 35, como visto na Figura 5.3. Tal valor de FID ter se mantido pode ser causado por dois principais motivos: (i) a quantidade de imagens usadas para o treinamento era relativamente pequena, com o *dataset* de imagens de melanoma sendo de cerca de 5 mil imagens; e (ii), a quantidade de imagens “vistas” pela GAN era relativamente pequena. Tal limitação se deve ao fato de que, o treinamento de uma GAN é um processo computacionalmente muito custoso, logo, é necessário um dispositivo massivamente paralelo, como GPUs de última geração. Entretanto, tais dispositivos não estavam disponíveis para o treinamento da GAN, impossibilitando assim um treinamento rápido.

Entretanto, ainda com tais deficiências, é possível ver que, entre as imagens geradas após 100 mil imagens executadas e 3 milhões a melhora é perceptível, com o último apresentando

além de uma melhor qualidade na imagem, as imagens possuíam uma similaridade maior com as imagens existentes no conjunto de dados criado, logo, foi dada continuidade no processo, usando tal GAN para a geração de mais imagens para o treinamento de uma CNN.

5.3 Resultados do treinamento da U-Net

Após o treinamento da rede GAN, foi realizado o treinamento da rede U-Net usando o *dataset* ISIC-2017, visando criar máscaras de segmentação de lesões de pele. Como citado anteriormente, o conjunto de imagem passou por um pré-processamento, removendo pelo da imagem, bem como redimensionando-as para o tamanho de 256×256 pixels.

De modo a identificar qual arquitetura de U-Net devia ser usada, foram realizados treinamentos com diversas arquiteturas de redes para segmentação, usando uma variedade de configurações.

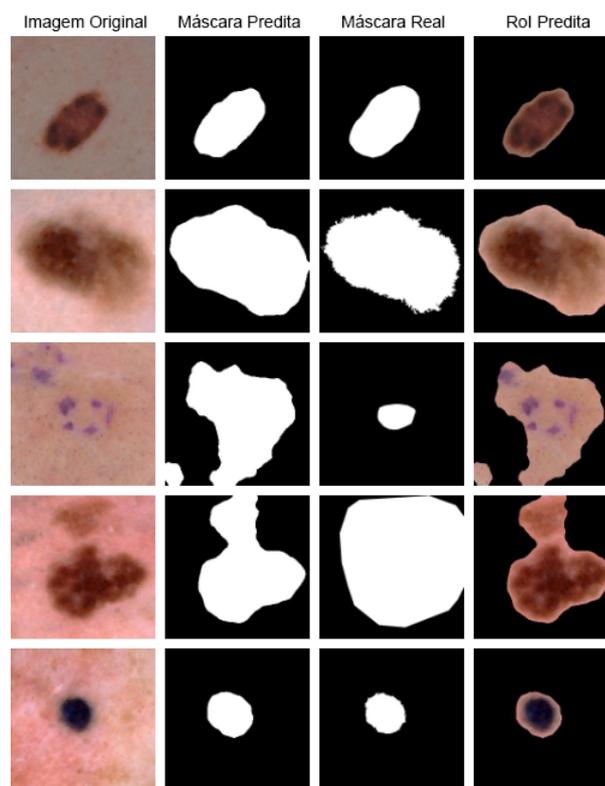
Tabela 5.2 – Resultados dos treinamentos das redes U-Net.

Modelo de U-Net	Backbone	IoU	Dice	Precisão	Recall	Acurácia
Attention U-Net	—	0,67107	0,792112	0,851064	0,758621	0,861921
TransUNET	—	0,650840	0,773635	0,836908	0,745466	0,855227
U-Net	DenseNet201	0,751079	0,851777	0,897579	0,822028	0,889160
U-Net	EfficientNetB0	0,712051	0,821449	0,849981	0,782113	0,870118
U-Net	EfficientNetB7	0,761792	0,859130	0,895648	0,836580	0,893494

Fonte: Do Autor (2021).

Como é possível observar pela Tabela 5.2, foram realizados testes usando as redes Attention U-Net (OKTAY et al., 2018), TransUNET (CHEN et al., 2021) e U-Net (RONNEBERGER; FISCHER; BROX, 2015), usando como *backbone* as CNNs DenseNet-201 (HUANG; LIU; WEINBERGER, 2016) e EfficientNet (TAN; LE, 2020), usando as configurações B0 e B7. Como é possível ver, as redes U-Net obtiveram melhores resultados no treinamento, com a rede usando como *backbone* a CNN EfficientNetB7 alcançando melhores resultados. Isso se deve ao fato de que, diferente das redes TransUNET e Attention U-Net, as redes U-Net usavam um *backbone* composto por uma CNN já treinada com o *dataset* ImageNet, logo, não seria um caso onde o treinamento seria realizado partindo de pesos aleatórios. Sendo assim, foi escolhida a rede U-Net usando a CNN EfficientNetB7 como *backbone*, dado que esta obteve os melhores resultados em quatro das cinco métricas usadas, principalmente nas métricas IoU e DC, as métricas mais importantes para o processo de segmentação.

Figura 5.4 – Exemplos de máscaras preditas pela CNN U-Net.



Fonte: Do Autor (2021).

Na Figura 5.4, é mostrada a imagem original, a máscara predita pela U-Net, a máscara real definida pelo *dataset* bem como a RoI predita, gerada pela aplicação da máscara predita sobre a imagem original. Nas linhas 1, 2 e 5 da Figura 5.4 são mostrados resultados onde a CNN conseguiu alcançar resultados bem próximos do esperado. A linha 3, por outro lado, mostra um cenário onde a U-Net não conseguiu gerar uma máscara que delimitasse de forma aceitável a lesão de pele, criando uma máscara muito maior. Um caso particular é a linha 4, onde a máscara predita não corresponde à máscara definida pelo *dataset*. Entretanto, a máscara predita pela rede U-Net contorna melhor a lesão de pele que a máscara definida pelo conjunto de dados, gerando uma RoI melhor que a do *dataset*. Com base nesses resultados, bem, como nos valores obtidos nas métricas de IoU e DC, pôde-se concluir que a U-Net treinada melhora o RoI das máscaras que serão utilizadas como entrada pelas camadas de atenção.

5.4 Resultados do treinamento da CNN EfficientAttentionNet

De posse das imagens geradas pela GAN, bem como das máscaras definidas pela U-Net, foi dado início ao processo de treinamento e validação da CNN EfficientAttentionNet proposta para a classificação de lesões de pele. Foram realizados testes para medir o impacto de cada

passo inserido no processo de modo a quantificar quais trouxeram um maior impacto para a melhora do desempenho da CNN proposta.

Sendo assim, foi inicialmente realizado o treinamento de uma rede EfficientNetB0, usando como parte classificadora a mesma composição proposta na CNN EfficientAttentionNet, de modo a criar um valor de base. Então, foi realizado o treinamento utilizando o *dataset* sem pelo, medindo o impacto da remoção de pelo em tal *dataset* ainda desbalanceado. Dito isso, foi realizado o treinamento desta EfficientNetB0 usando desta vez um *dataset* balanceado, contendo as imagens sintéticas geradas pela rede GAN. Finalmente, em um último cenário, foi realizado o treinamento da rede EfficientAttentionNet usando o *dataset* contendo as imagens sintéticas e as máscaras criadas pela rede U-Net. Os resultados de cada um desses treinamentos são mostrados na Tabela 5.3.

Tabela 5.3 – Resultados da avaliação de desempenho obtidos pelos modelos em diferentes estágios.

Modelo	Acurácia	Recall	Precisão	ROC-AUC
Modelo base	0,806948	0,800352	0,811052	0,806948
Modelo base com <i>dataset</i> sem pelo	0,889642	0,799472	0,458165	0,849361
Modelo base com imagens da GAN	0,960244	0,995534	0,924635	0,960204
EfficientAttentionNet	0,979031	0,995008	0,945401	0,976972

Fonte: Do Autor (2021).

Como pode ser observado na Tabela 5.3, existe um grande salto de desempenho da CNN quando é realizado o pré-processamento, com uma melhora de 10% de acurácia, 15% de precisão, 20% em *recall* e 25% na métrica de ROCAUC. Isso acontece porque, não existindo pelo na imagem, a CNN possui menos elementos para processar, permitindo assim que o processo de aprendizagem foque somente na lesão de pele, ao contrário de outras características, como pelo. Além disso, a inclusão de imagens sintéticas que “imitam” resultaram em um ganho de desempenho, com uma melhoria de 7,35% para precisão, 24,52% de *recall*, 101,81% para precisão e, finalmente, 13,05% para ROCAUC em relação ao modelo base.

Essa melhora se deve a dois fatores: primeiro, a inserção de mais imagens leva ao aumento do conjunto de imagens, que, por consequência, impacta na qualidade dos resultados obtidos pela CNN. Isso porque, com mais imagens disponíveis, a CNN consegue identificar melhor as características de cada classe, resultando em uma melhor acurácia da previsão. Além disso, como dito anteriormente, um *dataset* desbalanceado traz como ponto fraco a maior tendência do classificador a uma classe, sendo ela a classe dominante no *dataset*. Logo, com a adição de imagens sintéticas da classe menos dominante, neste caso, melanoma, a CNN possui um conjunto de dados balanceado para seu treinamento, eliminando este viés causado pela

baixa quantidade de imagens de uma classe. É importante destacar que no conjunto de testes usados, não existia a presença de imagens sintéticas, dado que isso poderia criar algum tipo de viés para a CNN.

Além disso, é possível ver que a aplicação de um método de atenção proposto trouxe uma melhora de desempenho, como acontece com a remoção de pelo, tal técnica traz uma atenção somente para a lesão de pele, removendo objetos que não fossem de interesse para a classificação de lesões de pele.

Tais resultados mostram que o uso de pré-processamento na imagem, como a remoção de pelos, além da inclusão de imagens sintéticas traz uma melhora significativa no desempenho da CNN, servindo como base para futuros trabalhos, mostrando os passos necessários para uma melhora nos resultados.

5.5 Considerações Finais

Após as demonstrações deste capítulo, ficou claro o impacto de cada uma das etapas da abordagem proposta para a classificação de lesão de pele. Foi demonstrado que a etapa de pré-processamento é de grande valia para a classificação de lesões de pele, dado que consiste em um algoritmo determinístico de relativa baixa complexidade, porém, capaz de melhorar consideravelmente os resultados apresentados. Além disso, a geração de imagens sintéticas é outro passo que trouxe um ganho considerável de desempenho, visto que corrige o desbalanceamento do *dataset* utilizado. No Capítulo seguinte, é feita a conclusão deste trabalho de mestrado, bem como uma alusão a possíveis trabalhos futuros.

6 CONCLUSÕES

A realização da pesquisa envolve diversas áreas da ciência e medicina. Embora um dos objetivos seja a implementação de uma abordagem para classificação de lesão de pele, é necessário o investimento de tempo de outros pesquisadores para a captura, classificação e tratamento de imagens médicas de lesões de pele.

Após o desenvolvimento de todos os passos da abordagem proposta, bem como a execução de testes de modo a comparar a melhora trazida por cada um dos passos dessa abordagem, fica claro o ganho de desempenho em algoritmos de classificação de lesões de pele. A abordagem proposta trata sobre cada uma das etapas do processo de classificação de lesões de pele. Inicialmente, é mostrado que a remoção de pelos da lesão de pele traz um ganho de resultado considerável, dado que é um processo relativamente simples de ser aplicado, não necessitando nenhuma etapa anterior, e apresentando melhora na acurácia de 10,24%. Seguido a isso, é demonstrado que a inclusão de novas imagens no conjunto de treinamento utilizando imagens sintetizadas geradas por uma GAN de modo a tornar o *dataset* balanceado gera uma melhora de 7,93% na acurácia. Por fim, foi proposto um método de atenção baseado em RoI, onde a zona da lesão de pele é delimitada através de uma rede U-Net, trazendo melhora de 1,95% na acurácia do modelo de CNN.

Todos esses passos trouxeram um ganho nas métricas usadas para a avaliação da rede CNN. Isso no mundo real pode ser refletido como um método que pode trazer melhora na avaliação gerada por sistemas de auxílio médico para classificação de lesões de pele. Com isso, o presente trabalho ajuda em um diagnóstico precoce, o que no caso dos cânceres de pele não melanoma significa a possibilidade de técnicas menos invasivas, evitando assim possíveis mutilações causadas por estas intervenções; e no caso do câncer do tipo melanoma, uma chance de cura, podendo evitar a morte do paciente.

De modo a fomentar toda a comunidade acadêmica na busca de métodos mais eficientes para a classificação de lesões de pele, o *software* desenvolvido neste trabalho está disponível gratuitamente sob Licença MIT em um repositório no site *github.com*¹.

Como sugestão para trabalhos futuros está a realização do treinamento de uma nova arquitetura de GAN, de modo a buscar melhores resultados nas métricas de FID. Também são sugeridos mais estudos na área dos métodos de atenção, buscando alternativas ou propondo

¹ Disponível em: <<https://github.com/arthurmteodoro/skin-cancer-classification-using-gan-and-roi-attention>>

melhorias para o método descrito, de modo a trazer um melhor resultado nas métricas de desempenho. Por fim, sugere-se realização de estudos com outras bases de dados, principalmente em bases que contêm mais de duas classes, visto que neste caso as informações finais do método de classificação de pele seriam mais detalhadas para os médicos, bem como isso traz um desafio maior, tanto para os métodos de classificação como geração de imagens sintéticas e segmentação.

REFERÊNCIAS

- ABDELHALIM, I. S. A.; MOHAMED, M. F.; MAHDY, Y. B. Data augmentation for skin lesion using self-attention based progressive generative adversarial network. Elsevier BV, v. 165, p. 113922, mar. 2021.
- AGGARWAL, A.; DAS, N.; SREEDEVI, I. Attention-guided deep convolutional neural networks for skin cancer classification. In: **2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)**. [S.l.: s.n.], 2019. p. 1–6.
- ALDWGERI, A.; ABUBACKER, N. F. Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images. In: **Advances in Visual Informatics**. Springer International Publishing, 2019. p. 214–226. Disponível em: <https://doi.org/10.1007/978-3-030-34032-2_20>.
- ALOM, M. Z. et al. **Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation**. 2018. Disponível em: <<https://arxiv.org/abs/1802.06955>>.
- BIBILONI, P.; GONZÁLEZ-HIDALGO, M.; MASSANET, S. Skin hair removal in dermoscopic images using soft color morphology. In: **Artificial Intelligence in Medicine**. Springer International Publishing, 2017. p. 322–326. Disponível em: <https://doi.org/10.1007/978-3-319-59758-4_37>.
- BORGES, L. E. **Python para Desenvolvedores**. 2. ed. Rio de Janeiro: Edição do Autor, 2010. ISBN 978-85-909451-1-6.
- BUCKLAND, M.; GEY, F. The relationship between recall and precision. **Journal of the American Society for Information Science**, Wiley, v. 45, n. 1, p. 12–19, jan. 1994.
- BUSLAEV, A. et al. Alumentations: Fast and flexible image augmentations. **Information**, v. 11, n. 2, 2020. ISSN 2078-2489. Disponível em: <<https://www.mdpi.com/2078-2489/11/2/125>>.
- CAO, H. et al. **Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation**. 2021. Disponível em: <<https://arxiv.org/abs/2105.05537>>.
- CHEN, J. et al. Transunet: Transformers make strong encoders for medical image segmentation. **CoRR**, abs/2102.04306, 2021. Disponível em: <<https://arxiv.org/abs/2102.04306>>.
- CODELLA, N. C. F. et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: **2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)**. [S.l.: s.n.], 2018. p. 168–172.
- CODELLA, N. C. F. et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). **CoRR**, abs/1902.03368, 2019. Disponível em: <<http://arxiv.org/abs/1902.03368>>.
- COELHO, F. C. **Computação Científica com Python: Uma introdução à programação para cientistas**. Petrópolis: Edição do Autor, 2007. ISBN 978-85-907346-0-4.

COMBALIA, M. et al. **BCN20000: Dermoscopic Lesions in the Wild**. 2019. Disponível em: <<https://arxiv.org/abs/1908.02288>>.

COMMONS, W. **A battery powered double polarized dry dermatoscope made by 3GEN called the DERMLITE**. 2009. Acessado em Março de 2020. Disponível em: <<https://commons.wikimedia.org/wiki/File:Dermatoscope.jpg>>.

DATA SCIENCE ACADEMY. **Deep Learning Book**. 2021. Acessado em 17 de outubro de 2021. Disponível em: <<https://www.deeplearningbook.com.br/>>.

DIAKOGIANNIS, F. I. et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. Elsevier BV, v. 162, p. 94–114, abr. 2020.

EICKENBERG, M. et al. Seeing it all: Convolutional network layers map the function of the human visual system. **NeuroImage**, Elsevier BV, v. 152, p. 184–194, maio 2017.

ELUYODE, O.; AKOMOLAFE, D. T. Comparative study of biological and artificial neural networks. **European Journal of Applied Engineering and Scientific Research**, v. 2, n. 1, p. 36–46, 2013.

ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, Springer Science and Business Media LLC, v. 542, n. 7639, p. 115–118, jan. 2017.

GAZZANIGA, M. S.; IVRY, R. B.; MANGUN, G. R. **Cognitive Neuroscience: The Biology of the Mind**. 5. ed. Nova Iorque: W. W. Norton & Company, 2019.

GESSERT, N. et al. Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting. **CoRR**, abs/1808.01694, 2018. Disponível em: <<http://arxiv.org/abs/1808.01694>>.

GHOSH, S. et al. Understanding deep learning techniques for image segmentation. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 52, n. 4, aug 2019. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/3329784>>.

GOODFELLOW, I. **Deep learning**. Cambridge, Massachusetts: The MIT Press, 2016. ISBN 978-0262035613.

GOODFELLOW, I. J. et al. **Generative Adversarial Networks**. 2014. Disponível em: <<https://arxiv.org/abs/1406.2661>>.

HAENSSLE, H. et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. **Annals of Oncology**, Elsevier BV, v. 29, n. 8, p. 1836–1842, ago. 2018.

HAENSSLE, H. et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. **Annals of Oncology**, Elsevier BV, v. 31, n. 1, p. 137–143, jan. 2020. Disponível em: <<https://doi.org/10.1016/j.annonc.2019.10.013>>.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. São Paulo: Bookman, 2007.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. **CoRR**, abs/1709.01507, 2017. Disponível em: <<http://arxiv.org/abs/1709.01507>>.

HUANG, G.; LIU, Z.; WEINBERGER, K. Q. Densely connected convolutional networks. **CoRR**, abs/1608.06993, 2016. Disponível em: <<http://arxiv.org/abs/1608.06993>>.

HUANG, H. et al. **UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation**. 2020. Disponível em: <<https://arxiv.org/abs/2004.08790>>.

IBM. **IBM Watson Health**. 2020. Acessado em 02 de Fevereiro de 2020. Disponível em: <<https://www.ibm.com/watson/br-pt/health/>>.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Estimativa 2020: incidência de câncer no Brasil**. Rio de Janeiro, 2019.

KARRAS, T. et al. Progressive growing of gans for improved quality, stability, and variation. **CoRR**, abs/1710.10196, 2017. Disponível em: <<http://arxiv.org/abs/1710.10196>>.

KARRAS, T.; LAINÉ, S.; AILA, T. A style-based generator architecture for generative adversarial networks. In: **2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2019. p. 4396–4405.

KHAN, S. et al. **A Guide to Convolutional Neural Networks for Computer Vision**. Morgan & Claypool Publishers LLC, 2018. Disponível em: <<https://doi.org/10.2200/s00822ed1v01y201712cov015>>.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. In: BENGIO, Y.; LECUN, Y. (Ed.). **3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings**. [s.n.], 2015. Disponível em: <<http://arxiv.org/abs/1412.6980>>.

KITTLER, H. et al. Diagnostic accuracy of dermoscopy. **The Lancet Oncology**, Elsevier BV, v. 3, n. 3, p. 159–165, mar. 2002.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. p. 3431–3440.

LOSHCHILOV, I.; HUTTER, F. SGDR: stochastic gradient descent with restarts. **CoRR**, abs/1608.03983, 2016. Disponível em: <<http://arxiv.org/abs/1608.03983>>.

LOSHCHILOV, I.; HUTTER, F. **Decoupled Weight Decay Regularization**. 2017. Disponível em: <<https://arxiv.org/abs/1711.05101>>.

LOUKAS, S. **ROC Curve explained using a COVID-19 hypothetical example: Binary & Multi-Class Classification tutorial**. 2020. Acessado em Agosto de 2020. Disponível em: <<https://towardsdatascience.com/roc-curve-explained-using-a-covid-19-hypothetical-example-binary-multi-class-classification-bab188ea869c>>.

LUGER, G. F. **Inteligência Artificial**. 6. ed. São Paulo: Pearson Education do Brasil, 2013. ISBN 978-8581435503.

MA, S.; XU, J.; ZHANG, C. Automatic identification of cited text spans: a multi-classifier approach over imbalanced dataset. **Scientometrics**, Springer Science and Business Media LLC, v. 116, n. 2, p. 1303–1330, abr. 2018. Disponível em: <<https://doi.org/10.1007/s11192-018-2754-2>>.

- MAHBOD, A. et al. Fusing fine-tuned deep features for skin lesion classification. **Computerized Medical Imaging and Graphics**, Elsevier BV, v. 71, p. 19–29, jan. 2019. Disponível em: <<https://doi.org/10.1016/j.compmedimag.2018.10.007>>.
- MAHBOD, A. et al. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. **Computer Methods and Programs in Biomedicine**, Elsevier BV, v. 193, p. 105475, set. 2020. Disponível em: <<https://doi.org/10.1016/j.cmpb.2020.105475>>.
- MAIYA, A. S. **ktrain: A Low-Code Library for Augmented Machine Learning**. 2020. Disponível em: <<https://arxiv.org/abs/2004.10703>>.
- MARCHETTI, M. A. et al. Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. **Journal of the American Academy of Dermatology**, Elsevier BV, v. 78, n. 2, p. 270–277.e1, fev. 2018.
- MAŠKA, M. et al. A benchmark for comparison of cell tracking algorithms. Oxford University Press (OUP), v. 30, n. 11, p. 1609–1617, fev. 2014.
- MAZUROWSKI, M. A. et al. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. **Neural Networks**, Elsevier BV, v. 21, n. 2-3, p. 427–436, mar. 2008. Disponível em: <<https://doi.org/10.1016/j.neunet.2007.12.031>>.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, Springer Science and Business Media LLC, v. 5, n. 4, p. 115–133, dez. 1943.
- MEDEIROS, L. F. de. **Inteligência Artificial Aplicada: uma abordagem introdutória**. 1. ed. Curitiba: InterSaber, 2018. ISBN 978-8559728002.
- MENDONÇA, T. et al. PH² - a dermoscopic image database for research and benchmarking. In: **2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. IEEE, 2013. Disponível em: <<https://doi.org/10.1109/embc.2013.6610779>>.
- MILLETARI, F.; NAVAB, N.; AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: **2016 Fourth International Conference on 3D Vision (3DV)**. [S.l.: s.n.], 2016. p. 565–571.
- MINISTÉRIO DA SAÚDE. **Câncer de pele: o que é, causas, sintomas, tratamento e prevenção**. 2018. Acessado em 25 de Março de 2020. Disponível em: <<https://saude.gov.br/saude-de-a-z/cancer-de-pele>>.
- MINSKY, M.; PAPERT, S. **Perceptrons: An Introduction to Computational Geometry**. Cambridge, MA, USA: MIT Press, 1969.
- NERSSON, R. et al. A dermoscopic skin lesion classification technique using YOLO-CNN and traditional feature model. Springer Science and Business Media LLC, v. 46, n. 10, p. 9797–9808, abr. 2021.

- OKTAY, O. et al. Attention u-net: Learning where to look for the pancreas. **CoRR**, abs/1804.03999, 2018. Disponível em: <<http://arxiv.org/abs/1804.03999>>.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Global Cancer Observatory**. 2020. Acessado em Agosto de 2020. Disponível em: <<https://gco.iarc.fr/>>.
- PATTERSON, J.; GIBSON, A. **Deep Learning: A Practitioner's Approach**. 1. ed. Sebastopol: O'Reilly, 2017.
- QIN, X. et al. U2-net: Going deeper with nested u-structure for salient object detection. **Pattern Recognition**, Elsevier BV, v. 106, p. 107404, Oct 2020. ISSN 0031-3203.
- QIN, Z. et al. A GAN-based image synthesis method for skin lesion classification. Elsevier BV, v. 195, p. 105568, out. 2020.
- RAHMAN, M. A.; WANG, Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In: **Advances in Visual Computing**. Springer International Publishing, 2016. p. 234–244. Disponível em: <https://doi.org/10.1007/978-3-319-50835-1_22>.
- RASCHKA, S. **Python machine learning : machine learning and deep learning with Python, scikit-learn, and TensorFlow**. Birmingham: Packt Publishing, 2017. ISBN 978-1-78712-593-3.
- REITER, O. et al. Artificial intelligence in skin cancer. **Current Dermatology Reports**, Springer Science and Business Media LLC, v. 8, n. 3, p. 133–140, jul. 2019.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: **Lecture Notes in Computer Science**. [S.l.]: Springer International Publishing, 2015. p. 234–241.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, p. 65–386, 1958.
- ROTEMBERG, V. et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. **Scientific Data**, Springer Science and Business Media LLC, v. 8, n. 1, jan. 2021. Disponível em: <<https://doi.org/10.1038/s41597-021-00815-z>>.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, Springer Science and Business Media LLC, v. 323, n. 6088, p. 533–536, out. 1986.
- SCIKIT-LEARN. **Confusion Matrix**. 2018. Acessado em 04 de Abril de 2020. Disponível em: <https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#sphx-glr-auto-examples-model-selection-plot-confusion-matrix-py>.
- SILVA, I. N. d.; SPATTI, D. H.; FLAUZINO, R. A. **Redes Neurais Artificiais para Engenharia e Ciências Aplicadas**. 1. ed. São Paulo: Artliber, 2010.
- SINGH, P. **Learn TensorFlow 2.0 Implement Machine Learning and Deep Learning Models with Python**. Berkeley, CA: Apress Imprint Apress, 2020. ISBN 978-1-4842-5560-5.

SKIN CANCER FOUNDATION. **Skin Cancer Facts & Statistics**. 2019. Acessado em 29 de Janeiro de 2020. Disponível em: <<https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>>.

SMITH, L. N. Cyclical learning rates for training neural networks. In: **2017 IEEE Winter Conference on Applications of Computer Vision (WACV)**. [S.l.: s.n.], 2017. p. 464–472.

SMITH, L. N. **A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay**. 2018.

SOCIEDADE BRASILEIRA DE DERMATOLOGIA. **Câncer de Pele**. 2018. Acessado em 29 de Janeiro de 2020. Disponível em: <<https://www.sbd.org.br/dermatologia/pele/doencas-e-problemas/cancer-da-pele/64/>>.

SOUZA, R. J. S. P. de et al. Estimativa do custo do tratamento de câncer de pele tipo não-melanoma no estado de são paulo - brasil. **An Bras Dermatol**, v. 82, n. 4, p. 657–662, 2011.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. 2020. Disponível em: <<https://arxiv.org/pdf/1905.11946.pdf>>.

TSCHANDL, P.; ROSENDAHL, C.; KITTLER, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. **Scientific Data**, Springer Science and Business Media LLC, v. 5, n. 1, ago. 2018. Disponível em: <<https://doi.org/10.1038/sdata.2018.161>>.

VASILEV, I. **Python deep learning : exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow**. Birmingham, UK: Packt Publishing, 2019. ISBN 978-1-78934-846-0.

VILLA-PULGARIN, J. P. et al. Optimized convolutional neural network models for skin lesion classification. *Computers, Materials and Continua (Tech Science Press)*, v. 70, n. 2, p. 2131–2148, 2022.

XIA, X.; KULIS, B. **W-Net: A Deep Model for Fully Unsupervised Image Segmentation**. 2017. Disponível em: <<https://arxiv.org/abs/1711.08506>>.

XUE, X.; KAMATA, S. ichiro; LUO, D. Skin lesion classification using weakly-supervised fine-grained method. In: . [S.l.]: IEEE, 2021.

ZHANG, N. et al. Skin cancer diagnosis based on optimized convolutional neural network. **Artificial Intelligence in Medicine**, v. 102, p. 101756, 2020. ISSN 0933-3657. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0933365719301460>>.

ZHAO, S. et al. Differentiable augmentation for data-efficient gan training. In: **Conference on Neural Information Processing Systems (NeurIPS)**. [S.l.: s.n.], 2020.

ZHONG, B. et al. Deep learning and network analysis: Classifying and visualizing accident narratives in construction. **Automation in Construction**, Elsevier BV, v. 113, p. 103089, maio 2020.

ZHOU, Z. et al. UNet++: A nested u-net architecture for medical image segmentation. In: . [S.l.]: Springer International Publishing, 2018. p. 3–11.