

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
PROJETO DE GRADUAÇÃO**



**AIURY SANT'ANNA JUREWSKI**

**TÉCNICAS DE DATA AUGMENTATION APLICADAS A  
MODELOS CONVOLUCIONAIS ORIENTADOS À  
SEGMENTAÇÃO DE IMAGENS DE LESÕES DE PELE**

VITÓRIA-ES

2022

Aiury Sant'Anna Jureswski

# **TÉCNICAS DE DATA AUGMENTATION APLICADAS A MODELOS CONVOLUCIONAIS ORIENTADOS À SEGMENTAÇÃO DE IMAGENS DE LESÕES DE PELE**

Parte manuscrita do Projeto de Graduação do aluno Aiury Sant'Anna Jureswski, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

VITÓRIA-ES

2022


Aiury Sant'Anna Jureswski

# **TÉCNICAS DE DATA AUGMENTATION APLICADAS A MODELOS CONVOLUCIONAIS ORIENTADOS À SEGMENTAÇÃO DE IMAGENS DE LESÕES DE PELE**

Parte manuscrita do Projeto de Graduação do aluno Aiury Sant'Anna Jureswski, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Aprovado em 24 de março de 2022.

## **COMISSÃO EXAMINADORA:**



---

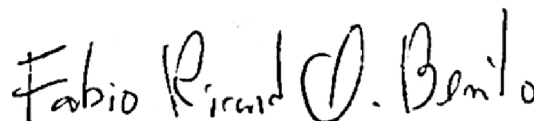
**Prof. Dr. Jorge Leonid Aching  
Samatelo**

Universidade Federal do Espírito Santo  
Orientador



---

**Prof. Dr. Patrick Marques Ciarelli**  
Universidade Federal do Espírito Santo  
Examinador



---

**Prof. Msc. Fabio Ricardo Oliveira  
Bento**

Instituto Federal do Espírito Santo  
Examinador

VITÓRIA-ES

2022

*Em homenagem a todos e todas que lutam por um mundo melhor para todas e todos.*

## AGRADECIMENTOS

Gostaria de agradecer a todas as pessoas que apareceram e estiveram comigo durante minha vida. Carrego comigo um pedacinho de cada um de vocês.

Aos meus pais, Patricia e Walteir, que sempre estiveram presentes, me apoiaram e tornaram tudo isso possível já que nunca mediram esforços para assegurar que eu e meus irmãos tivéssemos acesso ao conhecimento.

Aos meus irmãos, Aisha e Walteir, que sempre serviram como exemplos e inspiração no que quero me tornar e estiveram sempre presentes em tudo o que eu precisava.

À turma de engenharia elétrica 2017/1, também conhecida como a melhor turma de engenharia que já pisou na UFES, que tive o prazer de fazer parte, e, em especial, aos meus amigos Gustavo Rossi, Ramon Cesconetto e Pedro Henrique Ribeiro, que sempre me ajudaram e estiveram presentes durante toda essa minha jornada na UFES.

Aos meus professores da UFES por todo conhecimento passado, em especial ao meu orientador, Prof. Dr. Jorge Samatelo, que aceitou em embarcar nesse projeto comigo, que sempre foi muito solícito, atencioso e extremamente paciente comigo.

Ao Programa de Educação Tutorial Engenharia Elétrica, que durante minha permanência por quase dois anos, foi fonte de conhecimento e aprendizado, não só sobre engenharia, mas também sobre discussões que englobam a nossa sociedade. Em especial aos tutores Prof. Dr. Paulo José Menegáz e Prof. Dra. Tiara Smarssaro de Freitas.

Ao Prof. Dr. Nilson Marcellos, cujo sua orientação e experiência foi essencial nessa reta final da minha graduação.

À banca examinadora pelo tempo investido para a leitura e avaliação desse trabalho.

Por fim, agradeço à Universidade Federal do Espírito Santo pela minha formação. Que suas portas sempre estejam abertas com educação de qualidade e gratuita para todos que nela desejam estudar.

*"Faça ou não faça. Tentativa não há."*  
(Mestre Yoda)

## RESUMO

O câncer é um dos principais problemas de saúde pública no mundo. Um dos tipos de câncer mais comuns no Brasil é o câncer de pele. Para atingir uma maior chance de tratamento, a detecção precoce desse tipo de câncer é uma excelente estratégia, já que, possibilita um tratamento no estágio inicial do tumor. Nesse contexto, desenvolver ferramentas computacionais para suporte ao diagnóstico é essencial. Entre elas está a segmentação de imagens dermatoscópicas que objetiva particionar a região que contém a lesão principal do fundo da imagem. Supondo que tal tarefa seja efetuada por um modelo orientado a dados, será necessário contar com um grande conjunto de dados para treinamento do mesmo. Sendo assim, aqui é proposto usar como modelo uma rede neural orientada ao problema de segmentação semântica, e como metodologia para contornar o problema do volume de dados, técnicas e algoritmos de *Data Augmentation*(DA), efetuando uma análise comparativa das diversas configurações a ser consideradas. Foi possível implementar um abordagem que buscava o melhor grupo de técnicas de *Data Augmentation* para determinado modelo. Na etapa de avaliação dos resultados, a melhor estratégia dessa abordagem atingiu um resultado de 0,8040 no índice de Jaccard quando testado no banco de dados do desafio ISIC 2016, situando-se em 7<sup>o</sup> lugar nos resultados deste desafio.

**Palavras-chave:** *Deep learning*; Redes neurais convolucionais; *Data Augmentation*; Análise de lesões de pele; Segmentação semântica; U-net.

## ABSTRACT

Cancer is one of the main public health problems in the world. One of the most common types of cancer in Brazil is skin cancer. To achieve a higher chance of treatment, the early detection of this type of cancer is an excellent strategy allowing treatment in the early stages of the development of the tumor. In this context, develop computational tools to support the diagnosis is essential. Among them is the segmentation of dermoscopic images that aim to segment the region lesion of the image background. Assuming that a data-driven model performs this task, a large dataset for the training step will be necessary. Therefore, it's proposed the use of a semantic segmentation task-oriented neural network and, as a methodology to overcome the data volume problem, techniques and algorithms of data augmentation(DA), performing a comparative analysis of the different configurations. It was possible to obtain an algorithm of who search for the best group of data augmentation techniques for a determinate model. In the evaluation stage, the best strategy of this approach reached a result of 0,8040 with the Jaccard index when tested in the ISIC 2016 challenge dataset, ranking seventh place in the same challenge.

**Keywords:** Deep learning; Convolutional neural network; Data augmentation; Skin lesion analysis; Semantic Segmentation; U-net.



## LISTA DE FIGURAS

Figura 1 – Exemplo de segmentação semântica onde os pulmões estão em verde, o coração em vermelho e as clavículas em azul. . . . .	19
Figura 2 – Exemplo da segmentação de uma imagem dermastoscópica. . . . .	20
Figura 3 – Arquitetura da U-net . . . . .	21
Figura 4 – Aplicação da técnica de inversão imagem sendo: a) imagem original, b) com inversão horizontal e c) com inversão vertical. . . . .	22
Figura 5 – Aplicação da técnica de translação imagem sendo: a) imagem original, b) com translação horizontal de 10% e preenchimento e c) com translação vertical de 10% e preenchimento. . . . .	23
Figura 6 – Aplicação da técnica de rotação imagem sendo: a) imagem original, b) com rotação de $-10^\circ$ e c) com rotação de $+10^\circ$ . . . . .	23
Figura 7 – Representação de espaço de cores. . . . .	24
Figura 8 – Aplicação de técnicas de redistribuição em imagem sendo: a) imagem original, b) com equalização de histograma e c) com balanceamento de branco. . . . .	25
Figura 9 – Aplicação da transformação de brilho e contraste em uma imagem, sendo: a) imagem original, b) com 20% mais de brilho e c) com 20% mais de contraste. . . . .	25
Figura 10 – Aplicação de tonalização de cor em uma imagem, sendo: a) imagem original, b) com tonalização para verde e c) com tonalização para azul. . . . .	26
Figura 11 – Aplicação de solarização e inversão em uma imagem, sendo: a) imagem original, b) solarização aplicada com limiar de 127 e c) o negativo da imagem original. . . . .	26
Figura 12 – Aplicação de filtros em uma imagem, sendo: a) imagem original, b) suavização com um filtro de média de tamanho $3 \times 3$ , c) Suavização com um filtro Gaussiano de tamanho $3 \times 3$ e d) Aguçamento com um filtro laplaciano de tamanho $3 \times 3$ . . . . .	28
Figura 13 – Injeção de ruído em imagem, sendo: a) imagem original, b) imagem com ruído gaussiano e c) imagem com ruído do tipo sal e pimenta. . . . .	28
Figura 14 – Aplicação de <i>Random Erasing</i> a) imagem original, b) imagem com <i>Random Erasing</i> com <i>pathes</i> de $300 \times 100$ , c) imagem com <i>Random Erasing</i> com <i>pathes</i> de $100 \times 200$ . . . . .	29
Figura 15 – Exemplo de <i>MixUp augmentation</i> a) imagem 1, b) imagem 2, c) resultado transformação por <i>MixUp</i> entre a imagem 1 e 2 com $\lambda = 0,5$ . . . . .	30

Figura 16 – Exemplos de Equalização de histograma específico a) imagem original, b) imagem com $\lambda = 0,75$ , c) imagem com $\lambda = 0,5$ , d) imagem com $\lambda = 0,0$ . . . . .	31
Figura 17 – Fluxograma do funcionamento da busca bayesiana. . . . .	34
Figura 18 – Fluxograma do algoritmo de busca de melhor grupo de técnicas de DA. . . . .	36
Figura 19 – Exemplos de imagens divididas em categorias pela especificidade a) imagem normal, b) imagem com baixo contraste, c) imagem com marcações circulares, d) imagem com baixo contraste e marcações circulares. . . . .	46
Figura 20 – Exemplos de imagens divididas em categorias pelo tamanho da região da lesão a) imagem muito grande, b) imagem grande, c) imagem pequena, d) imagem muito pequena. . . . .	47
Figura 21 – Imagens no geral: a) imagem original, b) máscara de referência, c) máscara prevista (IJ = 0,0000), d) imagem original, e) máscara de referência, f) máscara prevista (IJ = 0,9720). . . . .	50

## LISTA DE TABELAS

Tabela 1 – Diretrizes principais das estratégias . . . . .	40
Tabela 2 – Número de imagens em cada categoria no conjunto de treinamento . .	46
Tabela 3 – Número de imagens em cada categoria no conjunto de teste . . . . .	47
Tabela 4 – Hiperparâmetros de treinamento do modelo base . . . . .	49
Tabela 5 – Resultados do modelo base . . . . .	50
Tabela 6 – Hiperparâmetros de treinamento das técnicas de DA individuais . . . .	51
Tabela 7 – Resultados para as técnicas de DA geométricas . . . . .	52
Tabela 8 – Resultados para as técnicas de DA no espaço de cor . . . . .	53
Tabela 9 – Resultados para as técnicas de DA com filtros . . . . .	53
Tabela 10 – Resultados para as técnicas de DA com algoritmos . . . . .	53
Tabela 11 – Hiperparâmetros de treinamento para grupos de técnicas estudadas . . .	53
Tabela 12 – Resultados obtidos para os grupos das técnicas de DA estudadas . . .	54
Tabela 13 – Melhores valores encontrados para cada uma das técnicas de DA . . .	55
Tabela 14 – Hiperparâmetros encontrados no primeiro experimento para cada estratégia . . . . .	57
Tabela 15 – Resultados encontrados no primeiro experimento para cada estratégia .	57
Tabela 16 – Hiperparâmetros encontrados no segundo experimento para cada estratégia	58
Tabela 17 – Resultados encontrados no segundo experimento para cada estratégia .	58
Tabela 18 – Hiperparâmetros encontrados no terceiro experimento para cada estratégia	59
Tabela 19 – Resultados encontrados no terceiro experimento para cada estratégia .	59
Tabela 20 – Hiperparâmetros encontrados no quarto experimento para cada estratégia	60
Tabela 21 – Resultados encontrados no quarto experimento para cada estratégia . .	60
Tabela 22 – Resultados obtidos pelas diferentes avaliações realizadas . . . . .	61
Tabela 23 – Camadas do caminho de contração de dados. . . . .	70
Tabela 24 – Camadas de ligação entre os caminhos e camadas de concatenação. . .	71
Tabela 25 – Camadas do expansão simétrica dos dados. . . . .	71
Tabela 26 – Hiperparâmetros constantes utilizados nas funções de transformação .	72

## LISTA DE ABREVIATURAS E SIGLAS

CNN	<i>Convolutional Neural Network</i>
DA	<i>Data Augmentation</i>
DL	<i>Deep Learning</i>
DP	Desvio Padrão
GPU	<i>Graphics Processing Unit</i>
IJ	<i>Índice de Jaccard</i>
INCA	Instituto Nacional de Câncer
ML	<i>Machine Learning</i>
PZ	Preenchimento com zeros
PVM	Preenchimento com valor médio do fundo da imagem
UFES	Universidade Federal do Espírito Santo
VC	Valores constantes
WHO	<i>World Health Organization</i>

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
1.1	Apresentação	14
1.2	Trabalhos Relacionados	16
1.3	Objetivos	17
1.4	Estrutura do Texto	17
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>19</b>
2.1	Segmentação semântica	19
2.2	U-net	20
2.3	Técnicas de <i>Data Augmentation</i>	21
2.3.1	Transformações geométricas	22
2.3.2	Transformações no espaço de cor	23
2.3.3	Transformações por filtros	26
2.3.4	Outras Técnicas de Data Augmentation	27
2.4	Algoritmos de Data Augmentation	29
2.4.1	<i>Random Erasing</i>	29
2.4.2	<i>MixUp augmentation</i>	30
2.4.2.1	Equalização Específica do Histograma e MixUp	31
2.5	Busca Bayesiana de Hiperparâmetros	32
<b>3</b>	<b>PROPOSTA</b>	<b>35</b>
3.1	Introdução	35
3.2	Busca de Técnicas de DA	35
<b>4</b>	<b>RESULTADOS</b>	<b>44</b>
4.1	Recursos computacionais	44
4.2	Conjunto de treinamento e testes	44
4.3	Métrica de avaliação	48
4.4	Modelo Base	48
4.5	Avaliação individual de cada técnica de DA selecionadas manualmente	50
4.6	Avaliação de grupos de técnicas de DA selecionados manualmente	52
4.7	Avaliação individual de cada técnica de DA selecionadas automaticamente	54
4.8	Avaliação do grupo de técnicas de DA selecionadas automaticamente	55
4.8.1	Resultados do primeiro experimento	57
4.8.2	Resultados do segundo experimento	57

4.8.3	Resultados do terceiro experimento . . . . .	58
4.8.4	Resultados do quarto experimento . . . . .	59
<b>4.9</b>	<b>Comparação dos resultados . . . . .</b>	<b>60</b>
<b>5</b>	<b>CONCLUSÕES E PROJETOS FUTUROS . . . . .</b>	<b>63</b>
<b>5.1</b>	<b>Conclusões . . . . .</b>	<b>63</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>64</b>
	<b>ANEXOS . . . . .</b>	<b>67</b>
	<b>ANEXO A – OPTUNA . . . . .</b>	<b>68</b>
	<b>ANEXO B – IMPLEMENTAÇÃO DA U-NET . . . . .</b>	<b>70</b>
	<b>ANEXO C – TABELA DE HIPERPARÂMETROS CONSTANTES . . . . .</b>	<b>72</b>

# 1 INTRODUÇÃO

## 1.1 Apresentação

Câncer é um termo que categoriza e abrange tipos diferentes de doenças que tem em comum o crescimento desordenado e anormal de células. Com o processo de crescimento e divisão rápida, estas células tendem a invadir tecidos adjacentes ou órgãos a distância. O câncer é um dos principais problemas de saúde pública no mundo e, de acordo com a World Health Organization (WHO, 2020), após os grupos de doença cardiovasculares, o câncer foi a principal causa das mortes no mundo entre 2000 e 2019.

Os diferentes tipos de câncer correspondem a tipos diferentes de células quando são atingidas. Quando é um tumor que atinge a pele, ele pode ser um câncer melanoma, que atinge as células produtoras de melanina, substância que determina a cor da pele, ou então câncer não melanoma. De acordo com o Instituto Nacional de Câncer (INCA, 2021), cerca de 30% de todos os tumores malignos registrados no país são de câncer de pele não melanoma. O câncer de pele melanoma representa somente cerca de 3% das neoplasias, entretanto, o câncer de pele melanoma é o tipo mais grave devido à sua alta possibilidade de provocar metástase.

De acordo com o Atlas de Mortalidade do INCA (2020), em 2019 cerca de 2647 óbitos foram registrados em decorrência ao câncer de pele e 1159 foram por causa do câncer de pele melanoma. Estes números representam 0,36% e 0,16% do total de óbitos registrados no Brasil, respectivamente. A estimativa para o ano de 2020 é que sejam registrados mais de 185 mil casos de câncer de pele, onde cerca de 8 mil sejam casos de câncer melanoma de acordo com o INCA (2019).

A detecção precoce do câncer é uma excelente estratégia para encontrar o tumor em uma fase inicial e, assim, possibilitar maior chance de tratamento bem sucedido. Há uma estimativa de que a taxa de sobrevivência de uma melanoma de 5 anos cai de 99% caso seja detectada nos estágios iniciais para 14% se detectada nos estágios mais avançados (ESTEVA et al., 2017). O diagnóstico do câncer de pele é feito pelo dermatologista por meio de um exame clínico, onde o mesmo analisa manchas pruriginosas, sinais ou pintas e feridas não cicatrizantes. Em determinadas situações, é necessário realizar um exame chamado dermatoscopia, que envolve o uso de um aparelho chamado dermatoscópio, sendo este um aparelho que permite ao profissional de saúde visualizar camadas da pele não vistas a olho nu.

A dermatoscopia, também conhecida como microscopia de epiluminescência ou microscopia

de superfície ampliada, é um método para permitir visualizar tecidos abaixo da camada mais externa da epiderme que normalmente surgem como o principal indicador no diagnóstico de lesões pigmentadas da pele (REZZE; SÁ; NEVES, 2006). De maneira generalizada, o fluxograma para o diagnóstico automatizado é a aquisição da imagem da lesão de pele, em seguida, a segmentação da região da pele onde a lesão se encontra, a extração de características da mancha da lesão segmentada e, por fim, a classificação das características obtidas.

O processo de segmentação é a separação da região da lesão das demais regiões. Esse processo pode ser dado por: (i) técnicas de limiarização de histograma, onde envolve a determinação de um valor de intensidade que serve como um limiar para dividir a imagem em duas regiões, o objeto de interesse e o fundo da imagem; (ii) técnicas de análise de agrupamento de dados, também conhecido como *Clustering*, onde envolve fazer um agrupamento de dados seguindo um grau de semelhança, nesse caso é analisado o segmento de pele e ele é separado do fundo devido a sua semelhança com atributos já pré-definidos (GONZALEZ; WOODS, 2008).

Uma outra abordagem que está presente nos estudos recentes é a segmentação e a classificação das imagens da lesão por *Machine Learning*. O ML (tradução livre, aprendizado de máquinas) é uma área de estudos onde é fornecido aos computadores uma habilidade de aprender a realizar uma tarefa sem serem explicitamente programados para realizar a mesma. Esse aprendizado pode ser por modelos supervisionados, como modelos preditivos de classificação ou regressão, ou então modelos não supervisionados, como modelos descritivos de sumarização, associação ou *clustering* (MITCHELL et al., 1997). Uma das subáreas de ML é a denominada *Deep Learning* (tradução livre, aprendizado profundo). Essa abordagem vem da ideia de aprender através da experiência e entender os dados de entrada em termos de hierarquia de conceitos, onde cada um desses conceitos é uma relação mais simples de outros conceitos. Essa estrutura hierárquica permite a uma rede neural entender conceitos complicados a partir de conceitos mais simples, como por exemplo, entender o conceito de uma pessoa através de bordas, cantos e contornos (GOODFELLOW; BENGIO; COURVILLE, 2016).

Uma etapa importante para uma rede neural é o treinamento. Normalmente os dados disponíveis são divididos em dois subconjuntos, o primeiro, é o conjunto de treinamento, que é utilizado para o treinamento do modelo em si, o segundo, é o conjunto de teste, que é utilizado para verificar o desempenho do modelo. Um dos problemas mais comuns é a necessidade de um conjunto de treino grande, necessário para obter um modelo ajustado com um desempenho admissível no conjunto de teste. Uma das técnicas mais utilizadas para incrementar o número de amostras de treino é *Data Augmentation* - DA (tradução



livre, aumento de dados), que consiste em adicionar cópias das amostras de treinamento originais com certas modificações. Essas novas modificações auxiliam em reduzir o erro de generalização (GOODFELLOW; BENGIO; COURVILLE, 2016). Com isto em mente, o propósito desse trabalho é implementar uma rede neural orientada ao problema de segmentação de imagens dermatoscópicas e, através de técnicas e algoritmos de DA, fazer uma análise comparativa do desempenho do modelo, objetivando determinar a configuração de DA que gere o maior desempenho.

## 1.2 Trabalhos Relacionados

Alguns estudos sobre DA relevantes para este trabalho são apresentados a seguir.

Em *Data augmentation for improving deep learning in image classification problem* (MIKOŁAJCZYK; GROCHOWSKI, 2018), os autores discutem métodos de DA desde os mais simples e clássicos, como transformações geométricas até técnicas mais recentes, como transferência de textura. O ponto importante da contribuição para este trabalho é que ele é focado nas técnicas de DA em imagens de lesões de pele. Ele traz algumas discussões importantes sobre que técnicas podem ser usadas e quais devem ser evitadas afim de evitar a diminuição do desempenho invés da melhoria.

Em *A survey on Image Data Augmentation for Deep Learning* (SHORTEN; KHOSHGOFTAAR, 2019), os autores discutem as técnicas clássicas de DA até as técnicas mais recentes do estado da arte para melhorar o desempenho dos modelos neurais. O artigo é bem detalhado sobre as técnicas abordadas trazendo discussões sobre o impacto de cada uma delas no treinamento da rede.

Em *Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network* (LI; SHEN, 2018), os autores descrevem o desenvolvimento, a implementação e os resultados de uma rede chamada *Lesion Feature Network* (Tradução direta, rede de características de lesões), sendo essa baseada em uma rede CNN. Este artigo foi desenvolvido para o desafio ISIC 2017 e descreve todo o processo de criação, como o tratamento de dados do conjunto de treinamento, as técnicas de DA utilizadas, a arquitetura da rede proposta e as métricas utilizadas para medir o desempenho do modelo.

Em *Data Augmentation for Skin Lesion Analysis* (PEREZ et al., 2018), os autores discutem técnicas de DA para o problema de análise de imagens de lesões de pele utilizando CNN e o banco de dados ISIC 2017. Esse artigo traz discussões sobre diferentes técnicas de DA aplicadas, como técnicas clássicas e alguns algoritmos de DA como *Random Erasing* e

*Mixup Augmentation*, assim como os métodos de treinamento e o impacto do tamanho do conjunto de dados no desempenho final da rede.

### 1.3 Objetivos

#### Objetivo Geral

- Implementar um modelo neural orientado à segmentação de imagens de câncer de pele e uma abordagem para a busca das melhores técnicas de DA.
- Realizar uma análise comparativa do desempenho do modelo entre a abordagem proposta e a metodologia clássica.

#### Objetivos Específicos

- Estudar e implementar a rede U-net;
- Estudar e implementar as técnicas clássicas de DA;
- Estudar e implementar algoritmos modernos de DA;
- Estudar e implementar uma abordagem para a busca de técnicas de DA;
- Na tarefa alvo, avaliar o desempenho da rede U-net considerando a abordagem proposta e efetuar uma análise comparativa dos resultados obtidos.

### 1.4 Estrutura do Texto

O presente trabalho está estruturado da seguinte maneira:

- **Introdução:** este capítulo inicial tem como objetivo contextualizar o problema estudado, apresentar os itens mais pertinentes as soluções atuais e a proposta de estudo;
- **Referencial teórico:** este capítulo tem como objetivo explicar com mais detalhes os assuntos pertinentes para este trabalho;

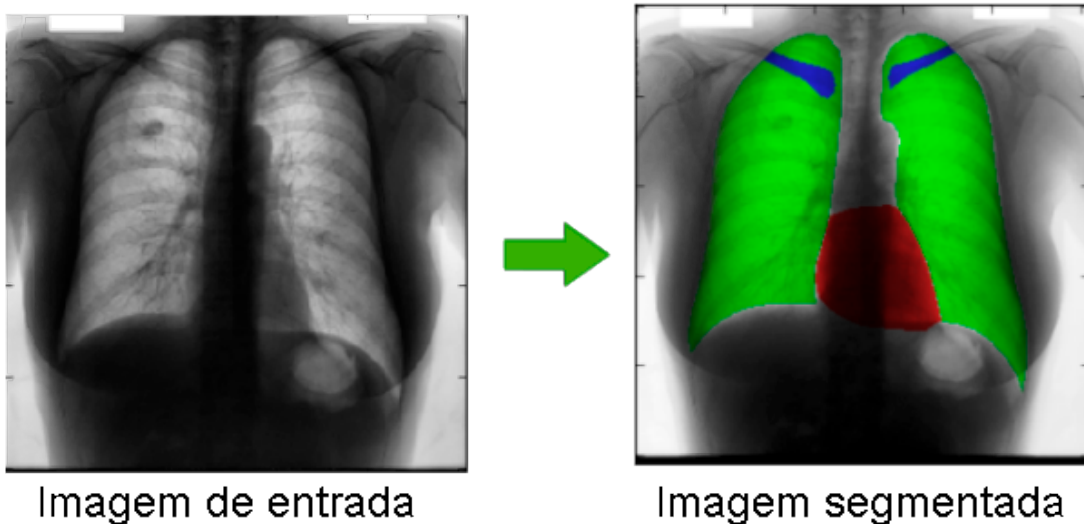
- **Proposta:** neste capítulo é apresentada a abordagem proposta para a determinação do melhor grupo de técnicas de DA objetivando um maior desempenho do modelo;
- **Resultados:** neste capítulo são apresentados os resultados dos experimentos, assim como uma análise comparativa dos mesmos;
- **Conclusão:** no capítulo final deste trabalho são apresentadas as discussões sobre os resultados encontrados, assim como a aplicabilidade e propostas de melhorias para estudos futuros.

## 2 REFERENCIAL TEÓRICO

### 2.1 Segmentação semântica

A segmentação semântica tem como objetivo particionar as regiões de uma determinada entrada em regiões semanticamente significativas e classificar cada uma dessas regiões em classes previamente determinadas. Esse processo é diferente da segmentação comum, já que, quando a entrada é particionada, é feita uma tentativa de entender o contexto e o que essas partições significam dentro da entrada. Pode-se observar na Figura 1 um exemplo de segmentação semântica onde, a partir de uma radiografia do tórax, é possível segmentar os órgãos do paciente.

Figura 1 – Exemplo de segmentação semântica onde os pulmões estão em verde, o coração em vermelho e as clavículas em azul.

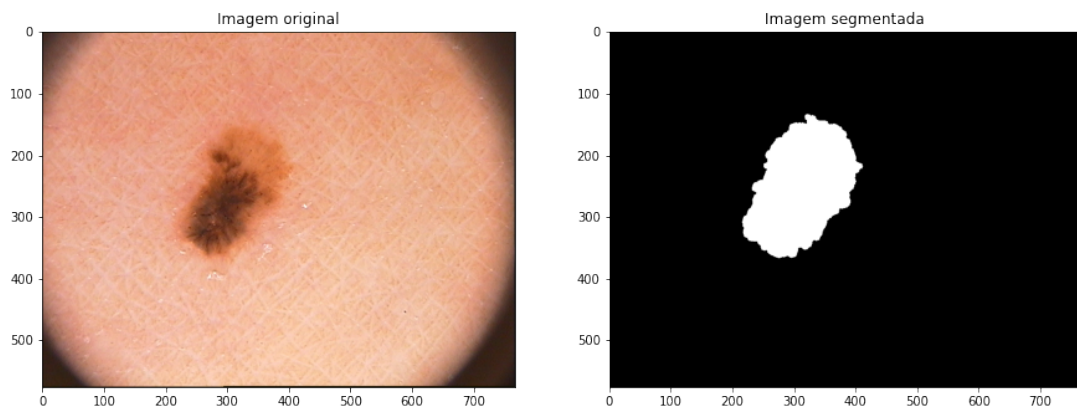


Fonte: Imagem traduzida e adaptada de (NOVIKOV et al., 2018).

Na Figura 2 é mostrada a segmentação semântica, em duas regiões, de uma imagem dermascópica: na primeira região, está a lesão de pele; na segunda região, está o fundo da imagem.

Existem diversas técnicas desenvolvidas na área de processamento digital de imagens orientadas para a tarefa de segmentação semântica. Entretanto, nos últimos anos, arquiteturas neurais profundas tem apresentado um desempenho superior em relação às técnicas clássicas (GHOSH et al., 2019). No entanto, devido à complexidade do problema, estes modelos ainda apresentam um desempenho inferior em relação à visão humana (PLANCHE; ANDRES, 2019), portanto, ainda é necessário trabalhar em cima das arquiteturas assim como os procedimentos de treinamento, para que sejam orientadas a ter um melhor desempenho.

Figura 2 – Exemplo da segmentação de uma imagem dermastoscópica.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

## 2.2 U-net

Atualmente, seja na tarefa de classificação ou segmentação, as redes baseadas em *Convolutional Neural Network* (tradução livre, redes neurais convolucionais) são as mais utilizadas. Uma CNN é um modelo, constituído principalmente por camadas convolucionais, as quais efetuam a operação de convolução entre a entrada e os filtros de uma camada convolucional.

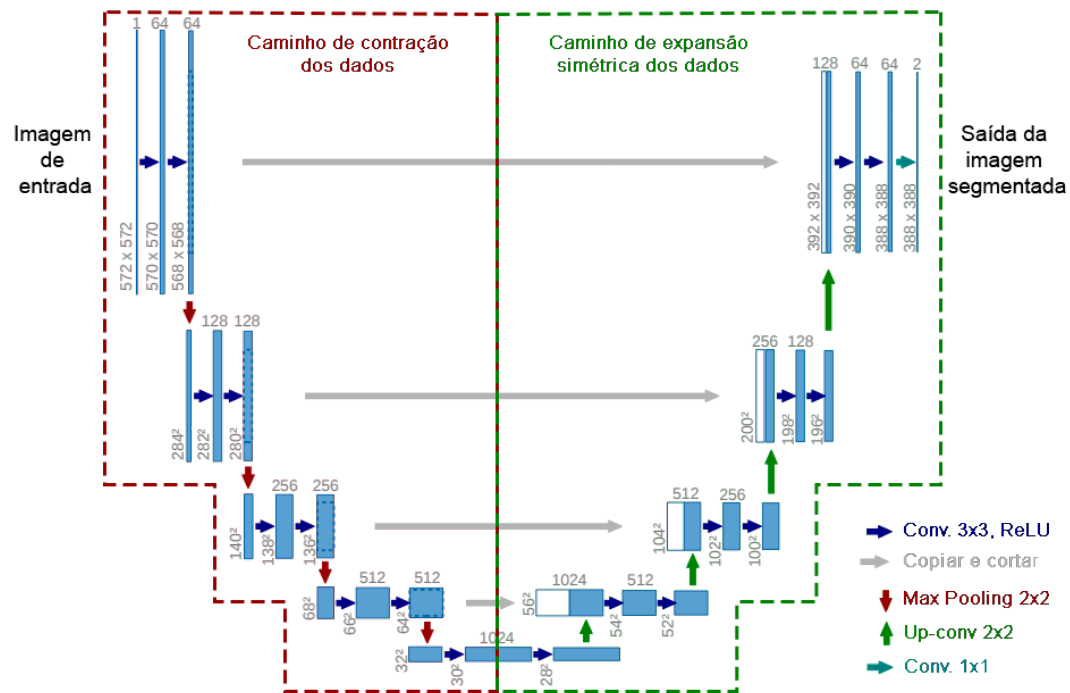
O primeiro modelo neural orientado à tarefa de segmentação semântica de imagens, foi a arquitetura *Fully Convolutional Network* (FCN) proposta por Long, Shelhamer e Darrell (2015). A FCN popularizou o uso de arquiteturas CNN para o problema de segmentação. Uma arquitetura que também usa camadas convolucionais é a U-net, proposta por Ronneberger, Fischer e Brox (2015), sendo aplicada na tarefa de segmentação semântica de imagens biomédicas. A U-net está dividida em dois blocos principais:

- **Encoder** (tradução livre, codificador). Tem como função extrair as características da imagem de entrada e está constituído por blocos de camadas convolucionais. Em cada bloco são realizadas duas convoluções  $3 \times 3$  cada uma seguida por uma função de ativação ReLU e uma operação de *Max Pooling*  $2 \times 2$  com *stride* igual a 2 que efetua uma redução do tamanho dos mapas de características gerado por cada camada convolucional. Cabe indicar que, o *Encoder* também é chamado de “caminho de contração dos dados” pois a cada passagem por um bloco, o tamanho dos correspondentes mapas de características diminui, mas seu número dobra.
- **Decoder** (tradução livre, decodificador). A função do *Decoder* é localizar precisamente as características extraídas. Tem blocos semelhantes ao *Encoder*, entretanto, cada bloco inicia com um incremento do tamanho dos mapas de características de en-

trada via uma operação de *upsampling* seguida por uma convolução  $2 \times 2$  que diminui pela metade o número de canais, depois é efetuada uma concatenação com o mapa de características do bloco correspondentemente do *Encoder*, e finalmente, são efetuadas duas convoluções  $3 \times 3$ , cada uma seguida por uma função de ativação ReLU. Cabe indicar que, o *Decoder* também é chamado de “caminho de expansão simétrica dos dados” pois a cada passagem por um bloco, o tamanho dos correspondentes mapas de características aumenta, mas seu número cai a metade.

A forma em que interagem o *Encoder* e o *Decoder* dão o formato em “U” para a rede. A Figura 3 ilustra a arquitetura da U-net destacando os dois caminhos e blocos de operações que constituem cada ramo.

Figura 3 – Arquitetura da U-net



Fonte: Imagem traduzida e adaptada de (RONNEBERGER; FISCHER; BROX, 2015)

### 2.3 Técnicas de *Data Augmentation*

As técnicas de DA aplicadas em imagens normalmente usam um conjunto de funções de transformação. As funções mais simples consistem em realizar rotações, inversões, deslocamentos e até aumentar e diminuir o *zoom*. Entretanto, ela pode torna-se mais complexa aplicando um conjunto diferente de máscaras e/ou então usando algoritmos orientados a aumentar o desempenho do modelo.

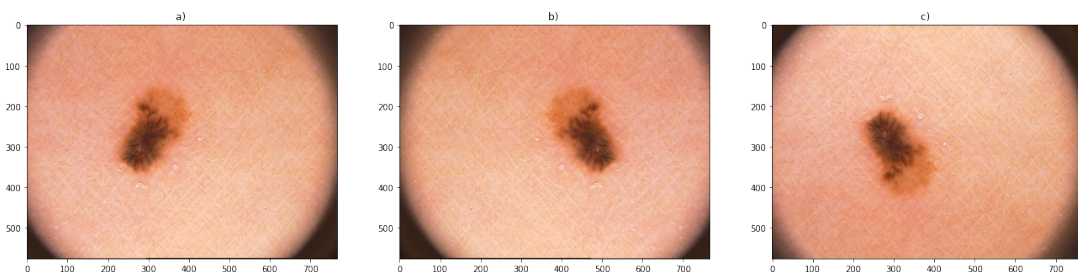
A seguir, serão explicados um conjunto de transformações comumente aplicadas para DA.

### 2.3.1 Transformações geométricas

Transformações geométricas são em geral transformações que modificam geometricamente a imagem sem a alteração do mapa de cores ou inserção de dados na imagem. Esse tipo de transformação é uma das mais fáceis de implementar, entretanto, são necessários alguns cuidados em que tipo de manipulação pode ser feita.

A inversão, também conhecida como *Flipping*, é um tipo de transformação onde a imagem é invertida em um determinado eixo, as inversões horizontais<sup>1</sup> são as mais comuns e mais aplicadas nesse caso. A Figura 4 contém um exemplo de inversão horizontal e inversão vertical a partir de uma imagem. Dependendo do caso, como por exemplo identificação de texto, essas inversões podem causar problemas com a correspondência da classificação (o número “6” e o número “9” se tornam similares depois de aplicar uma inversão vertical).

Figura 4 – Aplicação da técnica de inversão imagem sendo: a) imagem original, b) com inversão horizontal e c) com inversão vertical.

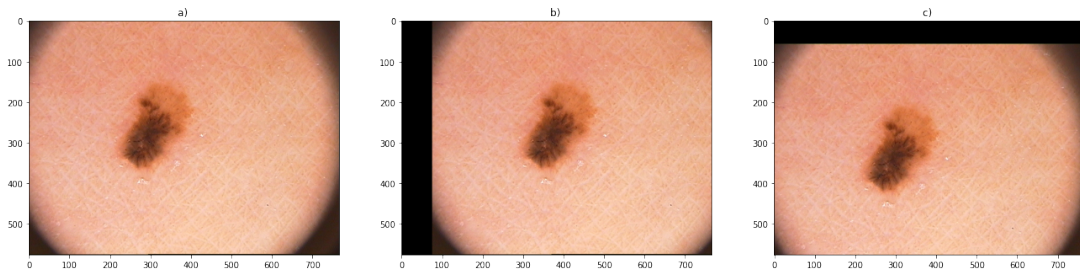


Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

A translação, é um tipo de transformação onde a imagem é deslocada em uma das direções. Essa transformação é bastante útil em DA pois é possível evitar posições enviesadas no conjunto de dados, como por exemplo o centro de imagem onde é costume ficar a maior parte das informações. O espaço restante resultante do deslocamento pode ser preenchido por valores constantes, como 0 ou 255, em imagens de 8 bits, ou então por uma média dos valores presentes na borda oposta ao sentido de deslocamento, ou ainda então por algum ruído aleatório ou gaussiano (SHORTEN; KHOSHGOFTAAR, 2019). A Figura 5 contém um exemplo de translação horizontal e vertical considerando um preenchimentos com zeros.

<sup>1</sup> Estas onde o eixo de inversão é o eixo das ordenadas tomando como um sistema de coordenadas cartesiano e o referencial como o centro da imagem.

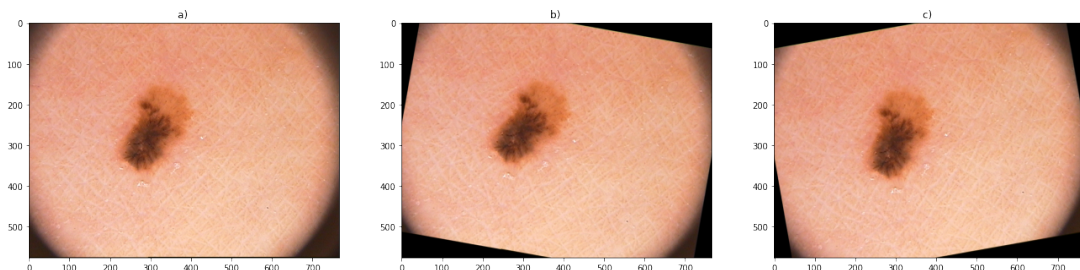
Figura 5 – Aplicação da técnica de translação imagem sendo: a) imagem original, b) com translação horizontal de 10% e preenchimento e c) com translação vertical de 10% e preenchimento.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

A rotação, é um tipo de transformação onde a imagem é rotacionada no ângulo de  $1^\circ$  a  $359^\circ$  para a esquerda ou para a direita utilizando o centro da imagem como o referencial. Para classificação de imagens, o ângulo impacta no resultado da classificação, já que, dependendo do ângulo, em texto por exemplo, o dígito “8” pode virar a ser o símbolo “∞”. A Figura 6 contém um exemplo de uma rotação no sentido horário de  $10^\circ$  e no sentido anti-horário de  $10^\circ$  aplicando preenchimentos com zeros.

Figura 6 – Aplicação da técnica de rotação imagem sendo: a) imagem original, b) com rotação de  $-10^\circ$  e c) com rotação de  $+10^\circ$ .



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

### 2.3.2 Transformações no espaço de cor

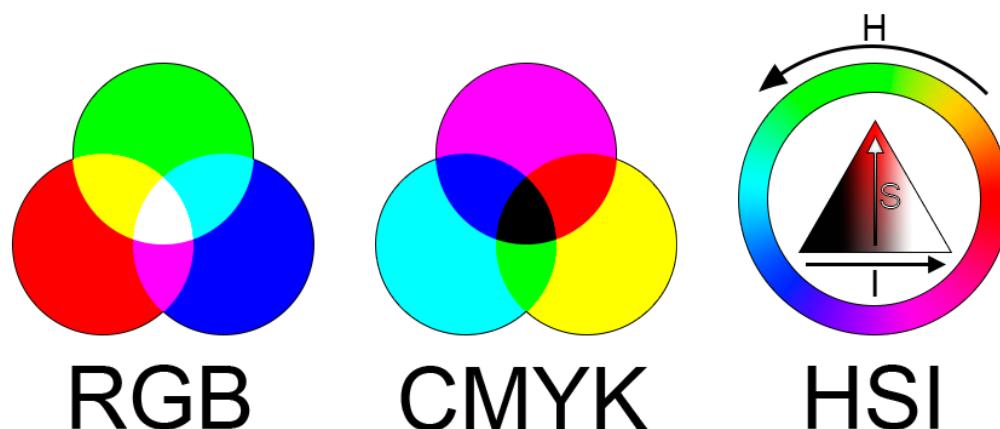
A maioria das imagens são codificadas em três matrizes diferentes, com cada uma delas com as mesmas dimensões, essas dimensões indicam a altura e a largura da imagem. Cada uma dessas matrizes representa um dos canais de cor, sendo o mais comum a representação em RGB (*Red*, *Green* e *Blue*), onde o valor de cada um dos elementos da matriz é o valor da intensidade de cada pixel.

Existem outras representações de cores que podem ser trabalhadas, como CMYK (*Cyan*, *Magenta*, *Yellow* e *Key*), HSI (*Hue*, *Saturation* e *Intensity*) ou outras e, como destaca o trabalho de Jurio et al. (2010), o espaço de CMY obtém melhor desempenho na segmentação



de imagens em relação aos outros espaços, entretanto, a transformação das imagens de RGB, ou qualquer outra representação, para CMY pode não ser vantajoso se a rede for previamente treinado por *Transfer Learning*<sup>2</sup> ou então quando a arquitetura da rede foi projetada para RGB. A Figura 7 ilustra algumas das diferentes representações de espaço de cores citadas.

Figura 7 – Representação de espaço de cores.



Fonte: Produção do próprio autor.

Outra opção seria transformar as imagens para escala de cinza, seja por média entre os três canais ou por uma soma ponderada que considera a percepção humana. A vantagem que isso traria é que resultaria em menos dados para se processar e conseqüentemente diminuiria o tempo de treinamento. Entretanto, como é pontuado pelo trabalho de Chatfield et al. (2014), isso faz com que o desempenho do modelo possa diminuir (aproximadamente 3% para classificação de imagens).

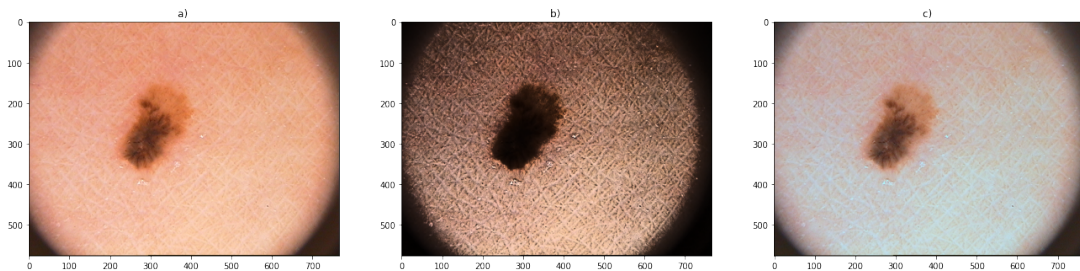
Uma desvantagem presente nas técnicas que trabalham em cima do espaço de cores é que, algumas informações importantes podem ser perdidas. Por exemplo, aumentar ou diminuir o brilho de uma imagem, pode ser que dificulte observar o objeto na imagem. Por outro lado, as transformações de cores podem ser um fator que ajude a eliminar o viés da cor, importante em tarefas de classificação de imagens (SHORTEN; KHOSHGOFTAAR, 2019).

Pode-se aplicar diversas transformações para alterar o valor do brilho, contraste ou redistribuir estas propriedades em cima de uma característica específica. Uma técnica utilizada para isso é a equalização de histograma, onde a partir do histograma normalizado

<sup>2</sup> Método onde é utilizado um modelo pré-treinado como um ponto inicial para o novo modelo em uma nova tarefa.

acumulado e uma função de transformação de redistribuição, é possível alterar os valores de ocorrência de uma imagem assim acentuando detalhes não visíveis anteriormente. Outra técnica é o balanceamento de branco onde é feita uma mudança na mistura de cores de uma imagem para fazer a correção de branco ou mudar para um conjunto qualquer de cor. A Figura 8 mostra a aplicação de técnicas de redistribuição de imagem como equalização de histograma e balanceamento de branco.

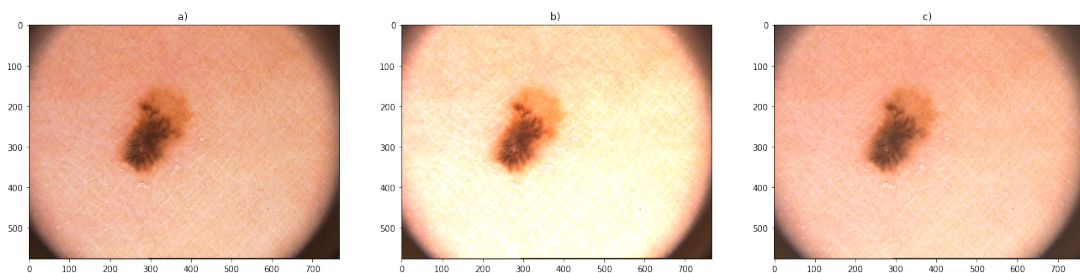
Figura 8 – Aplicação de técnicas de redistribuição em imagem sendo: a) imagem original, b) com equalização de histograma e c) com balanceamento de branco.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

Também pode ser feita a transformação direta de brilho e contraste através de uma função  $g = \alpha \cdot f + \beta$ , onde,  $\alpha$  é o valor de contraste,  $\beta$  é o valor de brilho,  $f$  é a imagem de entrada e  $g$  é a imagem de saída. A Figura 9 demonstra a aplicação da transformação de brilho e contraste.

Figura 9 – Aplicação da transformação de brilho e contraste em uma imagem, sendo: a) imagem original, b) com 20% mais de brilho e c) com 20% mais de contraste.

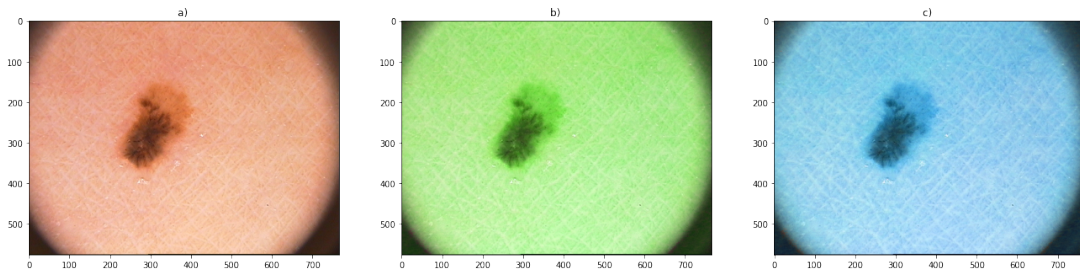


Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

Algumas transformações impactam mais na coloração direta da imagem, como por exemplo, a técnica de tonalização de cores, onde é feito um ajuste nos canais de cores da imagem para aumentar ou diminuir o impacto de uma ou mais cores. A Figura 10 demonstra a aplicação da tonalização para aumentar, respectivamente, o impacto da cor verde e da cor azul na imagem original.

Outra técnica que pode ser utilizada é a solarização, onde, a partir de um valor de limiar, todas as intensidades dos pixels são invertidas. Um processo parecido com a solarização

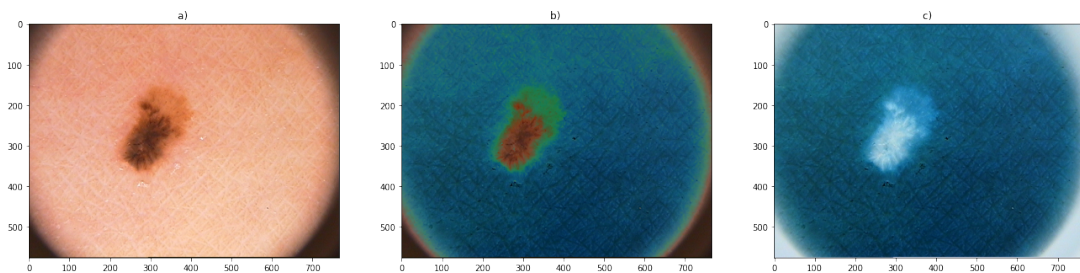
Figura 10 – Aplicação de tonalização de cor em uma imagem, sendo: a) imagem original, b) com tonalização para verde e c) com tonalização para azul.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

é a inversão, também conhecido como negativo da imagem, onde todos os valores de intensidade são invertidos. Essas técnicas são úteis para realçar detalhes não perceptíveis incorporados a certas regiões de uma imagem. A Figura 11 demonstra a aplicação da técnica de solarização e da técnica de inversão.

Figura 11 – Aplicação de solarização e inversão em uma imagem, sendo: a) imagem original, b) solarização aplicada com limiar de 127 e c) o negativo da imagem original.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

### 2.3.3 Transformações por filtros

As técnicas por transformações por filtro são populares no processamento digital de imagem para aguçar ou borrar as imagens. Essa técnica, também chamada de filtragem por *kernel*, consiste em fazer a convolução da imagem com um filtro espacial de tamanho específico. Essas técnicas normalmente utilizam uma operação de *padding*, para preencher as bordas das imagens com 0 ou 255, em caso de imagem de 8 bits, evitando a diminuição da dimensão espacial das imagens filtradas.

Existem vários filtros diferentes para suavizar uma imagem, entre eles tem-se: os filtros de média, média ponderada e Gaussiano. A Figura 12 demonstra a aplicação dos filtros de média, média ponderada e Gaussiano.

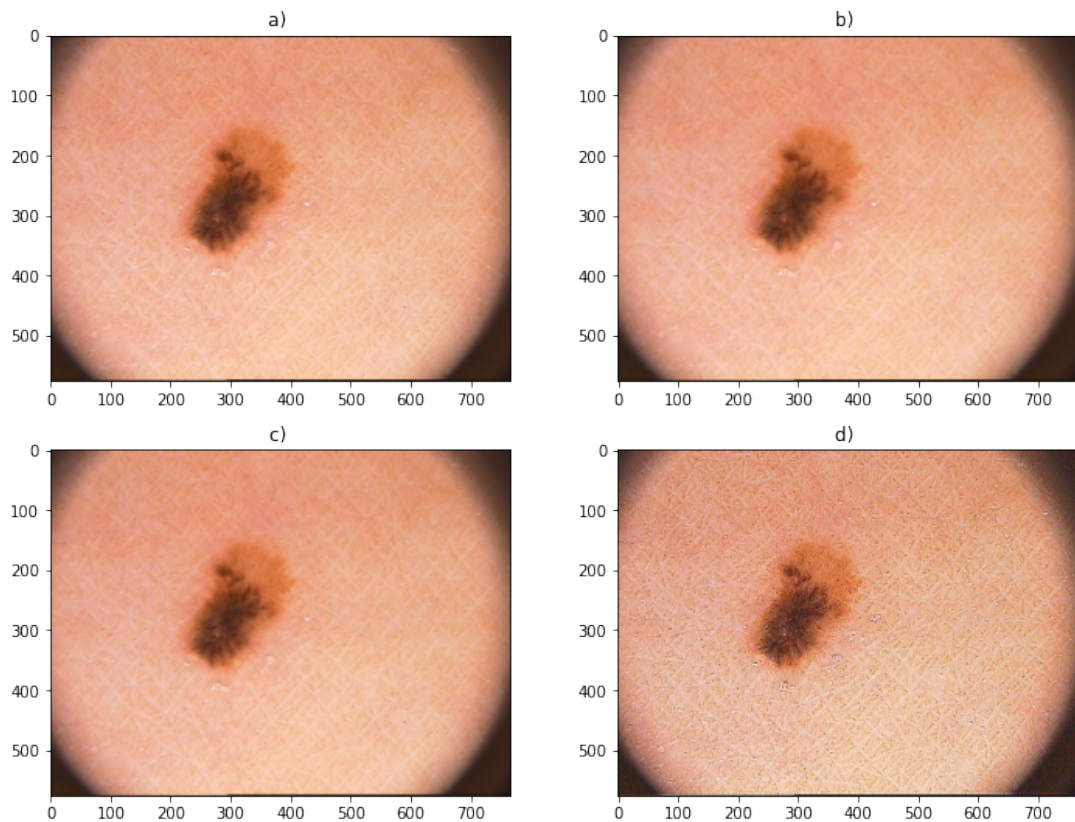
- O filtro de média calcula a média de todos os pixels contidos na vizinhança do filtro. A ideia por trás desse filtro é simples: substituir os valores pela média deles faz com que não existam mais transições abruptas de intensidade.
- O filtro de média ponderada se assemelha ao filtro de média, exceto que é calculada a média da vizinhança considerando pesos diferentes para cada um dos coeficientes do filtro, permitindo dar mais importância para alguns pixels invés de outros. O filtro de média ponderada tem uma vantagem em relação ao filtro de média pelo fato de preservar melhor as características dos contornos da imagem, já que preserva melhor os pixels centrais.
- O filtro Gaussiano é um filtro onde os coeficientes do filtro são gerados por uma distribuição gaussiana, neste caso, existe um peso maior para os pixels centrais do filtro em relação aos demais. Esses coeficientes são divididos pelo somatório de todos os coeficientes para que seja gerado um valor de média. O filtro Gaussiano é parecido com o filtro de média ponderada, exceto que esse preserva melhor os contornos, às custas de um processamento computacional maior (GONZALEZ; WOODS, 2008).

Uma forma de transformação simples para aumentar o nível de nitidez de uma imagem é através da soma da segunda derivada da imagem com a imagem original. A metodologia para isso consiste em definir uma fórmula discreta em duas dimensões para a segunda derivada de cada um dos pixels da imagem e construir um filtro com base nessa formulação. O operador derivativo mais simples nesse caso é o laplaciano, onde, devido ao fato de qualquer derivada ser operação linear, é um operador linear. Esse laplaciano então captura os detalhes da imagem que é somada a imagem original, de forma que os detalhes e a nitidez da imagem são realçados. A Figura 12 demonstra o aguçamento de uma imagem produto de somar a imagem filtrada com o laplaciano e a imagem original.

#### 2.3.4 Outras Técnicas de Data Augmentation

Existem outras técnicas para fazer o DA que são mais direcionadas a adicionar novos elementos na imagem de forma que não prejudique o desempenho na tarefa alvo, mas ajude a retirar algumas das características enviesadas existentes. A técnica de injeção de ruído consiste em inserir na imagem uma matriz de valores aleatórios. Esses valores aleatórios podem ser de uma distribuição gaussiana ou outro tipo de ruído, como Poisson ou Sal e Pimenta. Uma vantagem de aplicar ruído durante o treinamento da rede é o fato de melhorar a resposta do modelo contra ruídos fazendo ele mais robusto (JIN; DUNDAR; CULURCIELLO, 2015). A Figura 13 demonstra o resultado do algoritmo de injeção de ruído gaussiano e ruído sal e pimenta.

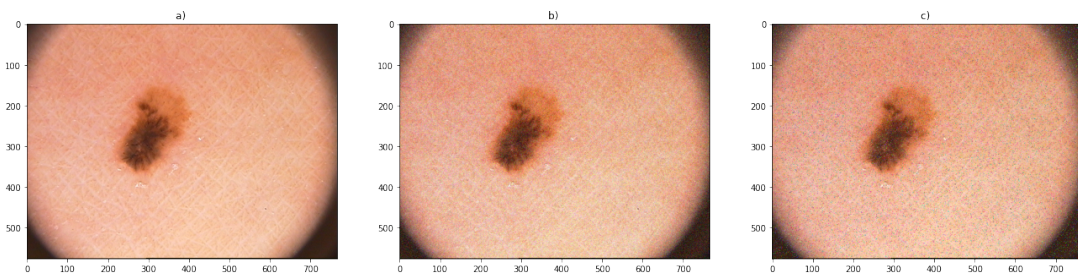
Figura 12 – Aplicação de filtros em uma imagem, sendo: a) imagem original, b) suavização com um filtro de média de tamanho  $3 \times 3$ , c) Suavização com um filtro Gaussiano de tamanho  $3 \times 3$  e d) Açuamento com um filtro laplaciano de tamanho  $3 \times 3$ .



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

Uma outra técnica seria a transferência de textura. Esta consiste em sintetizar uma textura de outra imagem. Essa técnica contém algumas dificuldades devido ao fato de aplicar uma nova textura na imagem e ao mesmo tempo manter os atributos visuais dela como o contorno, o sombreamento e outras (GATYS; ECKER; BETHGE, 2016). Algumas técnicas mais clássicas mantêm a ideia de re-amostrar a textura da imagem ao invés de transferir a textura de outra fonte (EFROS; FREEMAN, 2001).

Figura 13 – Injeção de ruído em imagem, sendo: a) imagem original, b) imagem com ruído gaussiano e c) imagem com ruído do tipo sal e pimenta.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

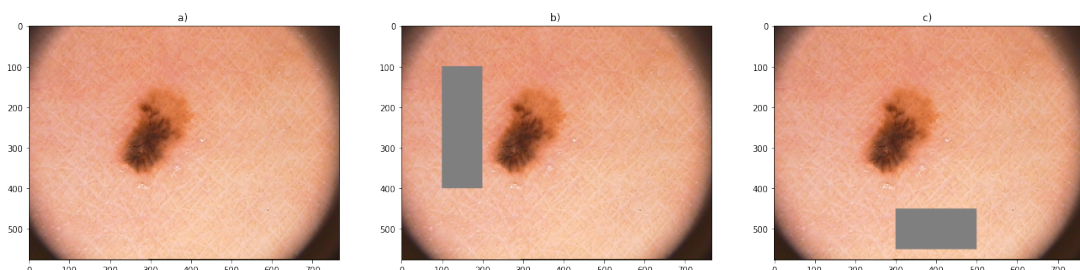
## 2.4 Algoritmos de Data Augmentation

Além das técnicas básicas de transformação de imagens para DA, alguns algoritmos mais sofisticados foram desenvolvidos ao longo do tempo para não só realizar a expansão do conjunto de dados de treinamento, mas também poder aumentar o desempenho do modelo no treinamento. Alguns desses algoritmos não só podem mas como são recomendados serem utilizados em conjunto com algumas das técnicas apresentadas anteriormente para ter melhores resultados, entretanto, alguns podem não funcionar bem em conjunto com outras técnicas.

### 2.4.1 *Random Erasing*

O algoritmo de *Random Erasing* (Tradução livre, apagar aleatoriamente) é uma abordagem desenvolvida por Zhong et al. (2020). O algoritmo funciona substituindo algum *patch* de tamanho  $n \times m$  na imagem por 0, 255, a média dos valores ou até valores aleatórios. Esse algoritmo pode ser usado em conjunto com outras técnicas vistas anteriormente, como inversão ou tonalização de cores. Quando é feita a exclusão aleatória de algum *patch* da imagem, força a rede a tentar descobrir melhor outras características que descrevem a imagem, prevenindo assim um sobreajuste devido a um fator enviesado na imagem. Em outras palavras, faz a rede tentar coletar dados da imagem toda e não apenas alguns *patches* mais comuns. A Figura 14 mostra a aplicação do algoritmo de *Random Erasing* considerando janelas de tamanhos diferentes.

Figura 14 – Aplicação de *Random Erasing* a) imagem original, b) imagem com *Random Erasing* com *pathes* de  $300 \times 100$ , c) imagem com *Random Erasing* com *pathes* de  $100 \times 200$ .



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

Uma desvantagem clara desse algoritmo é que dependendo do tamanho ou do posicionamento da janela, as imagens podem perder a região alvo da classificação. Por exemplo, dependendo da posição da janela, em textos onde um “6” pode virar um “0”. Logo, pode

ser necessário alguma intervenção manual dependendo do conjunto de dados e da tarefa (SHORTEN; KHOSHGOFTAAR, 2019).

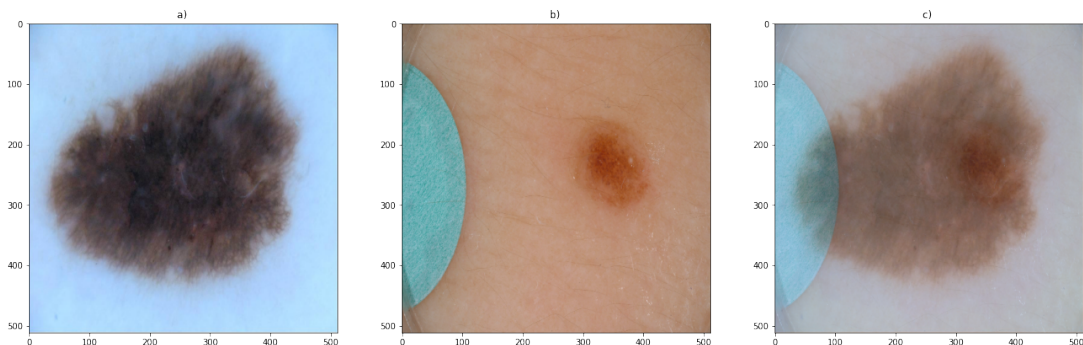
### 2.4.2 *MixUp augmentation*

O algoritmo de *MixUp augmentation* (Tradução livre, aumento por mistura) é uma técnica desenvolvida por Zhang et al. (2017) com uma abordagem bem contraintuitiva para fazer DA. O algoritmo faz uma interpolação linear com pesos de duas imagens existentes formando uma terceira. Supondo  $(x_i, y_i)$  e  $(x_j, y_j)$  serem dois pares entrada-saída do conjunto de treinamento e  $\lambda \in [0, 1]$ , então, pode-se gerar o par  $(\tilde{x}, \tilde{y})$  como:

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j.\end{aligned}$$

Do ponto de vista humano, essa é uma transformação de pouca utilidade, entretanto, do ponto de vista de um modelo neural, essa transformação permite a redução do sobreajuste do modelo, já que ocorre uma diminuição da extração de características enviesadas do conjunto de dados que não são facilmente perceptíveis para a visão humana. Outro benefício é o aumento da robustez da rede sem um gasto computacional grande, já que são necessárias apenas algumas linhas de código para fazer a interpolação. A Figura 15 ilustra exemplos de uso do algoritmo de *MixUp augmentation*. Uma desvantagem desse algoritmo é que ela é altamente dependente do conjunto de dados, ou seja, um conjunto de imagens ruins pode prejudicar o desempenho de um modelo.

Figura 15 – Exemplo de *MixUp augmentation* a) imagem 1, b) imagem 2, c) resultado transformação por *MixUp* entre a imagem 1 e 2 com  $\lambda = 0,5$



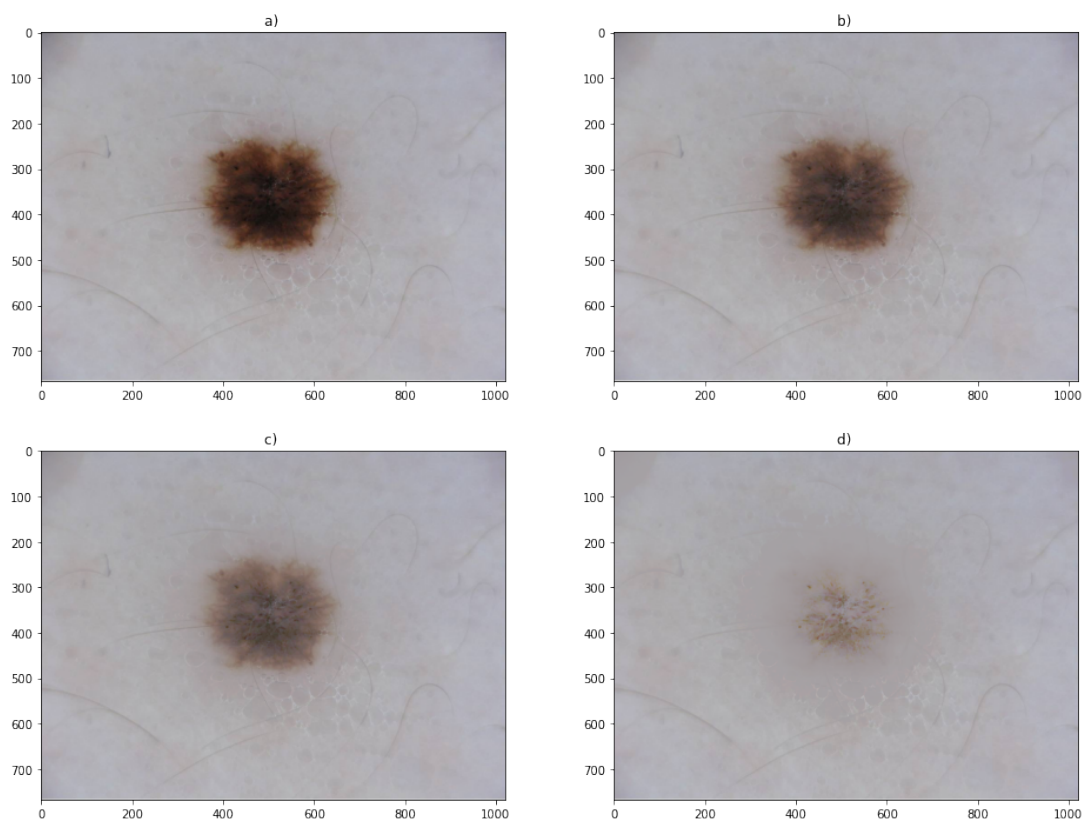
Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

### 2.4.2.1 Equalização Específica do Histograma e MixUp

A equalização específica do histograma, também conhecida como especificação do histograma, é o ajuste do histograma de uma imagem baseado em outro histograma, proveniente de outra imagem ou de uma função que define contraste de histograma. A equalização de histograma é um caso especial da equalização específica já que a função que define o contraste é uma distribuição normal.

Para a utilização em técnicas de DA, a equalização específica utiliza uma região da imagem para adequá-la a outra região ou uma função pré-determinada. Para adicionar uma variável de controle da equalização, é possível utilizar em conjunto com o algoritmo de *MixUp*, onde é feita a interpolação linear entre a imagem que teve seu histograma alterado e a imagem original. Nesse caso, a variável de controle é o  $\lambda$ . A figura 16 demonstra a aplicação de equalização de histograma específico com diferentes valores de  $\lambda$  aplicada apenas na região da lesão.

Figura 16 – Exemplos de Equalização de histograma específico a) imagem original, b) imagem com  $\lambda = 0,75$ , c) imagem com  $\lambda = 0,5$ , d) imagem com  $\lambda = 0,0$ .



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).



## 2.5 Busca Bayesiana de Hiperparâmetros

Um dos problemas mais comuns para atingir um modelo de alto desempenho dentro da área de ML é a otimização de hiperparâmetros, já que para cada problema existem valores de hiperparâmetros que melhor se adequam à solução do problema. A otimização de hiperparâmetros pode ser feita de maneira empírica, ou seja, testando manualmente, entretanto, essa metodologia pode ser inviável em modelos com numerosos hiperparâmetros devido ao tempo gasto na busca dos valores ótimos.

Outra abordagem é a utilização de algoritmos que fazem a calibração e otimização dos hiperparâmetros. Existem diversas técnicas, entre elas:

- **Busca exaustiva:** É feita a análise de todas as possíveis combinações, entretanto essa técnica não é prática em DL pela quantidade de hiperparâmetros a ser otimizados, além do espaço amostral de busca e o custo computacional;
- **Busca aleatória:** É feita uma busca de combinações aleatórias em um espaço pré-definido. É possível que seja encontrado uma solução de boa qualidade, mas não é garantido que seja encontrada a melhor;
- **Busca em grade:** É feita a análise de combinações pré-definidas para buscar o melhor conjunto discreto;
- **Busca Bayesiana:** É baseado nas informações dos conjuntos anteriores de hiperparâmetros para gerar o próximo conjunto de hiperparâmetros, quanto maior a quantidade de dados, maior a probabilidade de encontrar um conjunto ótimo de hiperparâmetros.

O propósito da utilização dessas abordagens em aprendizado de máquinas é realizar a busca de hiperparâmetros que melhor resolvem um problema de otimização. Tal problema pode ser descrito como encontrar o mínimo ponto de alguma função objetivo qualquer, onde não necessariamente é conhecida totalmente essa função, ou seja, essa função pode ser uma caixa preta. Esse problema de otimização pode ser representado como na Equação 2.1.

$$\hat{x} = \arg \min_x f(x) \quad (2.1)$$

O objetivo de realizar a otimização de hiperparâmetros é encontrar o melhor conjunto de hiperparâmetros de um dado modelo que retorne o melhor desempenho mensurado ou por um conjunto de validação ou por um conjunto de teste.

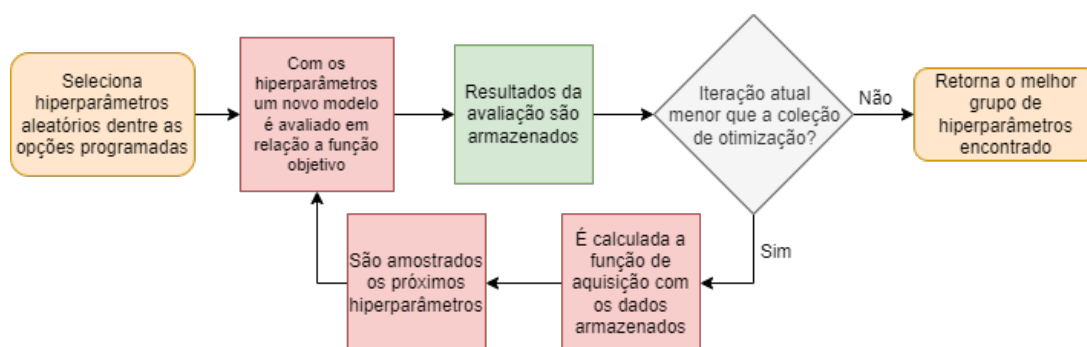
Diferente de métodos de busca aleatória ou por grade, a busca bayesiana mantém as informações das avaliações anteriores e as utiliza para formar um modelo probabilístico dos hiperparâmetros. Esse modelo é chamado de função substituta. Essa função substituta é mais fácil de ser otimizada que a função objetivo e a busca bayesiana funciona buscando o próximo conjunto de hiperparâmetros para avaliar na função objetivo através da seleção de hiperparâmetros que melhor funciona na função substituta. A ideia por trás da busca bayesiana é alcançar melhores resultados através da informação que é obtida através de cada avaliação da função objetivo. Essa informação atualiza a função substituta continuamente em um número de iterações. O número de iterações limite é chamado de coleção de otimização.

A busca bayesiana pode ser descrita através do seguinte conjunto de passos:

- Para cada iteração, é calculado o valor estimado da função substituta para uma entrada  $x$  usando um processo gaussiano, onde é formada uma estimativa da média e do desvio padrão em torno dessa estimativa. O desvio padrão serve para descrever a incerteza em cima dessa estimativa.
- É calculado o valor da função de aquisição, representada como  $\alpha(x)$ , tal função descreve o quão benéfico é amostrar o próximo ponto da faixa de valores dos hiperparâmetros.
- É achado o ponto máximo da função de aquisição e amostrado o próximo ponto para realizar a estimativa.

A vantagem de utilizar uma função de aquisição é pelo fato dessa função ser analítica e, portanto, é possível diferenciá-las e garantir que haja convergência ao menos em nível local. Para o que o processo de convergência seja global, é necessário que haja a otimização de vários pontos e, à medida que as iterações ocorram, o máximo encontrado pelo algoritmo seja próximo do global. A Figura 17 resume o Fluxograma da busca bayesiana utilizado neste trabalho.

Figura 17 – Fluxograma do funcionamento da busca bayesiana.



Fonte: Produção do próprio autor.

O *framework* de programação que será utilizado para realizar a busca Bayesiana neste trabalho é Optuna. Informações adicionais sobre o Optuna estão no anexo A.

## 3 PROPOSTA

### 3.1 Introdução

Assim como discutido anteriormente, a tarefa de segmentação semântica é rotular cada pixel de uma imagem em uma classe correspondente a partir do contexto onde a mesma está inserida. No problema de segmentação semântica de lesões de pele, a rotulação é feita em duas regiões, a lesão da pele e do fundo. É possível então definir este problema como uma função objetiva que relaciona cada imagem de entrada com sua correspondente representação binária, portanto, tomando um conjunto de treinamento, deseja-se determinar os parâmetros que melhor fazem essa classificação através da minimização de uma função de custo.

Como é possível imaginar, a determinação dos parâmetros é intimamente ligada ao conjunto de treinamento, ou seja, a presença de muitas imagens semelhantes e/ou a falta de dados prejudica o resultado final do modelo de conseguir identificar e rotular melhor cada pixel. Exatamente neste ponto que entram as técnicas de DA que permitem aumentar o conjunto de treinamento ao mesmo tempo que permite diversificar as imagens. Entretanto, existem diversas técnicas de DA, algumas podem auxiliar na determinação dos parâmetros enquanto outras podem prejudicar. Junto com cada uma dessas técnicas, muitas vezes vem a determinação de um valor ótimo, como o deslocamento em uma técnica de translação, e/ou então o número de imagens que será adicionado ao conjunto de treinamento através de determinada técnica, de forma que não prejudique através da adição de muitas imagens semelhantes. Um conjunto de treinamento desbalanceado pode atrapalhar o resultado final do modelo e realizar previsões ruins para determinadas classes, por isso é necessário a avaliação do número de imagens adicionadas para cada técnica.

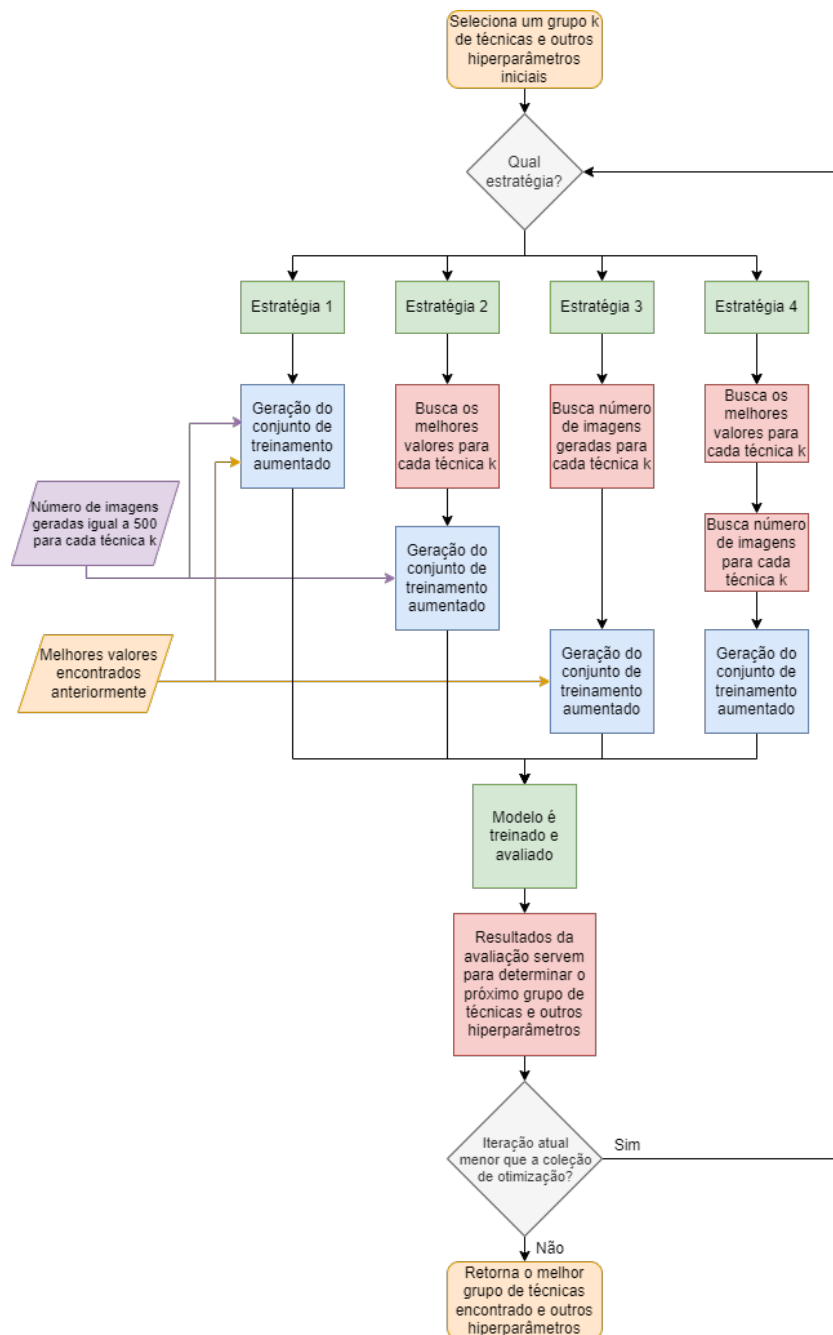
Visando solucionar estes três problemas principais, aqui é proposto usar uma busca de hiperparâmetros bayesiana para encontrar o melhor conjunto de técnicas de DA, os melhores valores e o número de imagens para ser adicionado ao conjunto de treinamento.

### 3.2 Busca de Técnicas de DA

Aqui é proposta uma abordagem de busca de técnicas de DA constituída de quatro estratégias diferentes. Cada estratégia, através da busca bayesiana, visa determinar o

melhor grupo de técnicas de DA que gerem um melhor desempenho de um modelo de segmentação neural. Entretanto, algumas dessas estratégias também visam encontrar melhores valores para cada uma das técnica selecionadas no grupo e o número de imagens para ser adicionada ao conjunto de treinamento aumentado. Todas as estratégias são independentes entre si, ou seja, não há a utilização de uma em conjunto com a outra ou o aproveitamento dos resultados encontrados por uma para a utilização da outra. A imagem 18 apresenta o Fluxograma que descreve o funcionamento do abordagem proposta.

Figura 18 – Fluxograma do algoritmo de busca de melhor grupo de técnicas de DA.



Fonte: Produção do próprio autor.

O estudo dos melhores valores encontrados para todas as técnicas pode ser resumido como: (i) uma busca bayesiana do valor de controle para a técnica específica em estudo; (ii) treinamento do modelo e avaliação do modelo; (iii) o valor que convergiu para a maior avaliação dentro do intervalo de estudo foi escolhido como melhor valor para a técnica em estudo. A metodologia completa para encontrar estes melhores valores está descrita na Seção 4.7.

Especificamente, as estratégias podem ser descritas como:

- **Estratégia 1:** Tem como objetivo buscar apenas o melhor grupo de técnicas de DA. Cada técnica deste grupo é buscada utilizando como entrada melhores valores encontrados previamente e o número de imagens para cada técnica definida como 500. A descrição algorítmica desta estratégia é mostrada no Algoritmo 1 e a metodologia completa para encontrar estes melhores valores está descrita na Seção 4.7.
- **Estratégia 2:** Tem como objetivo buscar o melhor grupo de técnicas de DA e os melhores valores para cada uma das técnicas dentro do grupo de técnicas encontradas. Para cada uma das técnicas o número de imagens geradas é igual a 500. A descrição algorítmica desta estratégia é mostrada no Algoritmo 2.
- **Estratégia 3:** Tem como objetivo buscar o melhor grupo de técnicas de DA e o melhor número de imagens para cada uma das técnicas encontradas. Cada técnica deste grupo é buscada utilizando como entrada os melhores valores encontrado previamente. A descrição algorítmica desta estratégia é mostrada no Algoritmo 3 e a metodologia completa para encontrar estes melhores valores está descrita na Seção 4.7.
- **Estratégia 4:** Tem como objetivo buscar o melhor grupo de técnicas de DA, os melhores valores e o número de imagens para cada uma das técnicas encontradas. A descrição algorítmica desta estratégia é mostrada no Algoritmo 4.

Duas destas estratégias são semi-automatizadas, já que a primeira e a terceira estratégia usam os melhores valores para encontrar o melhor grupo de técnicas com um número pré-determinado de número de imagens. As outras duas estratégias são totalmente automatizadas, já que a segunda busca o melhor grupo de técnicas e os melhores valores com o número de imagens pré-determinada e a quarta estratégia busca o melhor grupo de técnicas, os melhores valores e melhor número de imagens para cada técnica. As diretrizes principais de cada estratégia pode ser resumida através da tabela 1.

---

**Algoritmo 1:** Algoritmo Estratégia 1

---

**Data:** Conjunto de treinamento  $\mathcal{I}$ **Data:** Coleção de otimização  $CO$ **Data:** Grupo de técnicas iniciais (selecionadas aleatoriamente)  $\mathcal{T}_0$ **Result:** Grupo de técnicas  $\mathcal{T}$  $\mathcal{T} \leftarrow \mathcal{T}_0$ **for**  $i$  from 1 to  $CO$  **do**

```

  /* Seleção dos melhores valores ( $\mathcal{V}_0$ ) e do número de imagens ( $\mathcal{N}_0$ )
     para as técnicas definidas em  $\mathcal{T}$  */

```

```

   $\{\mathcal{N}_0, \mathcal{V}_0\} \leftarrow \text{MelhoresValoresPreCalculados}(\mathcal{T})$ 

```

```

  /* Geração do conjunto de treinamento aumentado */

```

```

   $\mathcal{IM} \leftarrow \text{ConjuntoDeTreinamentoAumentado}(\mathcal{I}, \mathcal{T}, \mathcal{N}_0, \mathcal{V}_0)$ 

```

```

  /* Treinamento e validação do modelo */

```

```

   $\text{modelo} \leftarrow \text{Treinamento}(\mathcal{IM})$ 

```

```

   $\text{resultados} \leftarrow \text{Validacao}(\text{modelo})$ 

```

```

  /* Busca do novo grupo de técnicas */

```

```

   $\mathcal{T} \leftarrow \text{Busca}(\text{resultados})$ 

```

**end**

---

---

**Algoritmo 2:** Algoritmo Estratégia 2

---

**Data:** Conjunto de treinamento  $\mathcal{I}$ **Data:** Coleção de otimização  $CO$ **Data:** Grupo de técnicas iniciais (selecionadas aleatoriamente)  $\mathcal{T}_0$ **Data:** Valores iniciais para as técnicas (definidos aleatoriamente)  $\mathcal{V}_0$ **Result:** Grupo de técnicas  $\mathcal{T}$ **Result:** Melhores valores para técnicas  $\mathcal{V}$  $\mathcal{T}, \mathcal{V} \leftarrow \mathcal{T}_0, \mathcal{V}_0$ **for**  $i$  from 1 to  $CO$  **do**

```

  /* Seleção do número de imagens ( $\mathcal{N}_0$ ) para as técnicas definidas
     em  $\mathcal{T}$  */

```

```

   $\{\mathcal{N}_0\} \leftarrow \text{MelhoresValoresPreCalculados}(\mathcal{T})$ 

```

```

  /* Geração do conjunto de treinamento aumentado */

```

```

   $\mathcal{IM} \leftarrow \text{ConjuntoDeTreinamentoAumentado}(\mathcal{I}, \mathcal{T}, \mathcal{V}, \mathcal{N}_0)$ 

```

```

  /* Treinamento e validação do modelo */

```

```

   $\text{modelo} \leftarrow \text{Treinamento}(\mathcal{IM})$ 

```

```

   $\text{resultados} \leftarrow \text{Validacao}(\text{modelo})$ 

```

```

  /* Busca do novo grupo de técnicas e dos novos melhores valores
     para cada técnica */

```

```

   $\{\mathcal{T}, \mathcal{V}\} \leftarrow \text{Busca}(\text{resultados})$ 

```

**end**

---

---

**Algoritmo 3:** Algoritmo Estratégia 3

---

**Data:** Conjunto de treinamento  $\mathcal{I}$ **Data:** Coleção de otimização  $CO$ **Data:** Grupo de técnicas iniciais (selecionadas aleatoriamente)  $\mathcal{T}_0$ **Data:** Número inicial de imagens por técnicas (definidas aleatoriamente)  $\mathcal{N}_0$ **Result:** Grupo de técnicas  $\mathcal{T}$ **Result:** Número de imagens por técnicas  $\mathcal{N}$  $\mathcal{T}, \mathcal{N} \leftarrow \mathcal{T}_0, \mathcal{N}_0$ **for**  $i$  from 1 to  $CO$  **do**

```

  /* Seleção dos melhores valores ( $\mathcal{V}_0$ ) para as técnicas definidas
     em  $\mathcal{T}$  */

```

```

   $\{\mathcal{V}_0\} \leftarrow \text{MelhoresValoresPreCalculados}(\mathcal{T})$ 

```

```

  /* Geração do conjunto de treinamento aumentado */

```

```

   $\mathcal{IM} \leftarrow \text{ConjuntoDeTreinamentoAumentado}(\mathcal{I}, \mathcal{T}, \mathcal{V}_0, \mathcal{N})$ 

```

```

  /* Treinamento e validação do modelo */

```

```

   $\text{modelo} \leftarrow \text{Treinamento}(\mathcal{IM})$ 

```

```

   $\text{resultados} \leftarrow \text{Validacao}(\text{modelo})$ 

```

```

  /* Busca do novo grupo de técnicas e do número de imagens para
     cada técnica */

```

```

   $\{\mathcal{T}, \mathcal{N}\} \leftarrow \text{Busca}(\text{resultados})$ 

```

**end**

---

---

**Algoritmo 4:** Algoritmo Estratégia 4

---

**Data:** Conjunto de treinamento  $\mathcal{I}$ **Data:** Coleção de otimização  $CO$ **Data:** Grupo de técnicas iniciais (selecionadas aleatoriamente)  $\mathcal{T}_0$ **Data:** Número inicial de imagens por técnicas (definidas aleatoriamente)  $\mathcal{N}_0$ **Data:** Valores iniciais para técnicas (definidos aleatoriamente)  $\mathcal{V}_0$ **Result:** Grupo de técnicas  $\mathcal{T}$ **Result:** Número de imagens por técnicas  $\mathcal{N}$ **Result:** Melhores valores para técnicas  $\mathcal{V}$  $\mathcal{T}, \mathcal{N}, \mathcal{V} \leftarrow \mathcal{T}_0, \mathcal{N}_0, \mathcal{V}_0$ **for**  $i$  from 1 to  $CO$  **do**

```

  /* Geração do conjunto de treinamento aumentado */

```

```

   $\mathcal{IM} \leftarrow \text{ConjuntoDeTreinamentoAumentado}(\mathcal{I}, \mathcal{T}, \mathcal{V}, \mathcal{N})$ 

```

```

  /* Treinamento e validação do modelo */

```

```

   $\text{modelo} \leftarrow \text{Treinamento}(\mathcal{IM})$ 

```

```

   $\text{resultados} \leftarrow \text{Validacao}(\text{modelo})$ 

```

```

  /* Busca do novo grupo de técnicas, dos novos melhores valores
     para cada técnica e do número de imagens para cada técnica */

```

```

   $\{\mathcal{T}, \mathcal{V}, \mathcal{N}\} \leftarrow \text{Busca}(\text{resultados})$ 

```

**end**

---



Tabela 1 – Diretrizes principais das estratégias

	Busca o melhor grupo de técnicas	Busca os melhores valores	Busca o número de imagens
Estratégia 1	Sim	Não	Não
Estratégia 2	Sim	Sim	Não
Estratégia 3	Sim	Não	Sim
Estratégia 4	Sim	Sim	Sim

Para todo conjunto de treinamento aumentado, sempre será utilizado o conjunto de treinamento original e o conjunto de técnicas encontrado. Todavia, algumas estratégias fazem a busca de outros hiperparâmetros, como por exemplo o número de imagens e os melhores valores para cada técnica. Importante definir que:

- o termo “**combinação**” trata-se da utilização de técnicas de DA aplicadas de maneira individual;
- o termo “**mesclagem**” trata-se da utilização de técnicas de DA aplicadas em conjunto.

Para gerar os conjuntos de treinamento aumentado utilizado pelas estratégias, considerando que será usada uma implementação *off-line* de DA, são propostas quatro Metodologias de Composição do Conjunto de Treinamento Aumentado (MCCTA):

- **MCCTA-1.** O Algoritmo 5 mostra a sequência de passos usada para a criação do conjunto de treinamento aumentado pelo MCCTA-1. De forma resumida:
  - Dado um grupo com  $k$  técnicas de DA, para cada técnica deste grupo é gerado um subconjunto aleatório de imagens a partir do conjunto de treinamento original onde as técnicas do grupo serão aplicadas. Essas imagens geradas, ou seja, imagens do subgrupo em que a combinação das técnicas de DA foram aplicadas, são adicionadas ao conjunto de treinamento aumentado.

De maneira exemplificada: Seja o grupo de técnicas de DA encontrado por uma estratégia igual a {Rotação, Translação}, e seja  $\mathcal{S}_k$  subconjuntos aleatórios de imagens do conjunto de treinamento original  $\mathcal{I}$ , o conjunto de treinamento aumentado vai ser formado por  $n_0$  Rotação( $\mathcal{S}_0$ ),  $n_1$  Translação( $\mathcal{S}_1$ ) e  $\mathcal{I}$ , onde  $n_k$  é o número de imagens geradas definida pela estratégia adotada.

- **MCCTA-2.** O Algoritmo 6 mostra a sequência de passos usada para a criação do conjunto de treinamento pelo MCCTA-2. De forma resumida:

- Dado um grupo com  $k$  técnicas de DA, é gerado um subconjunto aleatório de imagens a partir do conjunto de treinamento original. Para cada imagem do subconjunto gerado, é aplicado sucessivamente cada uma das técnicas de DA do grupo encontrado. Estas imagens geradas, ou seja, imagens do subgrupo em que as técnicas de DA foram aplicadas sucessivamente através da mesclagem, são adicionadas ao conjunto de treinamento aumentado. Não é avaliado a ordem de aplicação das técnicas.

De maneira exemplificada: Seja o grupo de técnicas de DA encontrado por uma estratégia igual a {Rotação, Brilho, Aquecimento} e seja  $\mathcal{S}$  um subconjunto aleatório de imagens do conjunto de treinamento original  $\mathcal{I}$ , o conjunto de treinamento aumentado vai ser formado por  $n$  Aquecimento(Brilho(Rotação( $\mathcal{S}$ ))), onde  $n$  é o número de imagens geradas definida pela estratégia adotada.

- **MCCTA-3.** O Algoritmo 7 mostra a sequência de passos usada para a criação do conjunto de treinamento pelo MCCTA-3. De forma resumida:

- Dado um grupo com  $k$  técnicas de DA e cada uma dessas técnicas com  $j$  técnicas de DA associadas a ela, para cada uma dessas técnicas é gerado um subconjunto aleatório de imagens a partir do conjunto de treinamento original. Em cada um destes subconjuntos, é aplicado a técnica do grupo e sucessivamente as técnicas associadas a esta. Essas imagens geradas, ou seja, imagens do subgrupo em que as técnicas de DA foram aplicadas e em seguida, as técnicas sucessivamente foram aplicadas, são adicionadas ao conjunto de treinamento aplicado. Importante destacar que os grupos de  $j$  técnicas de DA associadas não são necessariamente iguais entre si, ou seja, cada técnica  $k$  do grupo de técnicas de DA pode ter técnicas diferentes associadas a ela que serão aplicadas sucessivamente.

De maneira exemplificada: Seja o grupo de técnicas encontrada por uma estratégia igual a {Rotação, Brilho} e seja  $\mathcal{S}_k$  subconjuntos aleatórios de imagens do conjunto de treinamento original  $\mathcal{I}$ , o conjunto de treinamento aumentado vai ser formado por  $n_0$  Técnicas associadas a técnica 1(Rotação( $\mathcal{S}_0$ )),  $n_1$  Técnicas associadas a técnica 2(Brilho( $\mathcal{S}_1$ )) e  $\mathcal{I}$ , onde  $n_i$  é o número de imagens geradas definida pela a estratégia adotada.

Resumindo todas as MCCTA:

- MCCTA-1 faz a aplicação de técnicas do grupo de forma individual para cada subconjunto, ou seja, apenas a combinação das técnicas do grupo e sem mesclagem de técnicas;
- MCCTA-2 faz a aplicação de técnicas do grupo de forma mesclada, ou seja, todas são aplicadas juntas de maneira sucessiva;
- MCCTA-3 faz a aplicação de técnicas de forma individual e cada uma das técnicas desses grupo tem um grupo de técnicas associadas que são aplicadas de forma mesclada para cada técnica.

---

**Algoritmo 5:** Algoritmo MCCTA-1
 

---

**Data:** Conjunto de treinamento  $\mathcal{I}$

**Data:** Grupo de técnicas  $\mathcal{T} \leftarrow \{T_0, T_1, \dots, T_k\}$

**Data:** Valores para técnicas  $\mathcal{V} \leftarrow \{v_0, v_i, \dots, v_k\}$

**Data:** Número de imagens por técnicas  $\mathcal{N} \leftarrow \{n_0, n_1, \dots, n_k\}$

**Result:** Conjunto de treinamento aumentado  $\mathcal{IM}$

```

 $\mathcal{IM} \leftarrow \emptyset$ 
for  $k$  from 1 to  $|\mathcal{T}|$  do
  |  $\mathcal{S} \leftarrow \text{random}(\mathcal{I}, n_k)$ 
  | for  $j$  from 0 to  $|\mathcal{S}|$  do
  | |  $t \leftarrow T_k(s_j; v_k)$ 
  | |  $\mathcal{IM} \leftarrow \{t, \mathcal{IM}\}$ 
  | end
end

```

---



---

**Algoritmo 6:** Algoritmo MCCTA-2
 

---

**Data:** Conjunto de treinamento  $\mathcal{I}$

**Data:** Grupo de técnicas  $\mathcal{T} \leftarrow \{T_0, T_1, \dots, T_k\}$

**Data:** Valores para técnicas  $\mathcal{V} \leftarrow \{v_0, v_i, \dots, v_k\}$

**Data:** Número de imagens por técnicas  $n$

**Result:** Conjunto de treinamento aumentado  $\mathcal{IM}$

```

 $\mathcal{IM} \leftarrow \emptyset$ 
 $\mathcal{S} \leftarrow \text{random}(\mathcal{I}, n)$ 
for  $j$  from 0 to  $|\mathcal{S}|$  do
  |  $t \leftarrow T_0(T_1(\dots T_k(v_k, s_j); v_1); v_0)$ 
  |  $\mathcal{IM} \leftarrow \{t, \mathcal{IM}\}$ 
end

```

---

---

**Algoritmo 7:** Algoritmo MCCTA-3
 

---

**Data:** Conjunto de treinamento  $\mathcal{I}$

**Data:** Grupo de técnicas  $\mathcal{T} \leftarrow \{T_0, T_1, \dots, T_k\}$

**Data:** Grupo de técnicas associadas  $\mathcal{TA} \leftarrow \{\{T_{0,0}, \dots, T_{0,k}\}, \dots, \{T_{k,0}, \dots, T_{k,k}\}\}$

**Data:** Valores para técnicas  $\mathcal{V} \leftarrow \{v_0, v_1, \dots, v_k\}$

**Data:** Número de imagens por técnicas  $\mathcal{N} \leftarrow \{n_0, n_1, \dots, n_k\}$

**Result:** Conjunto de treinamento aumentado  $\mathcal{IM}$

$\mathcal{IM} \leftarrow \emptyset$

**for**  $k$  from 1 to  $|\mathcal{T}|$  **do**

$S \leftarrow \text{random}(\mathcal{I}, n_k)$

**for**  $j$  from 0 to  $|\mathcal{S}|$  **do**

$t \leftarrow T_{k,0}(T_{k,1}(\dots T_{k,k}(v_k; s_j); v_1); v_0)$

$\mathcal{IM} \leftarrow \{t, \mathcal{IM}\}$

**end**

**end**

---

## 4 RESULTADOS

### 4.1 Recursos computacionais

Os recursos disponíveis foram providos pelo Google Cloud, um serviço gratuito acessível pela plataforma do Google Colaboratory. Para ter um tempo de execução menor, maior ciclo de vida de execução dos notebooks e capacidade maior de RAM e VRAM, o Colab Pro foi assinado durante o período da pesquisa. A linguagem de programação para esse trabalho será o Python (3.7.12), com o framework Tensorflow (2.6.0) e com o framework Optuna (2.10.0).

O hardware da máquina virtual fornecida pelo Google Colab Pro é:

- **Processador:** Intel(R) Xeon(R) CPU @ 2.20GHz com dois núcleos físicos e duas *threads*;
- **GPU:** Tesla T4 com 16GB de VRAM ou Tesla P100 com 16GB de VRAM;
- **RAM:** Aproximadamente 25GB de RAM;
- **Armazenamento:** Aproximadamente 160 GB de acesso local.

### 4.2 Conjunto de treinamento e testes

Para a realização do treinamento dos classificadores e validação dos resultados, é necessário um conjunto de dados consolidado da literatura. Portanto, foram analisadas os conjuntos de dados de maior uso em trabalhos com objetivos semelhantes, chegando à escolha do conjunto de dados ISIC, o qual é um conjunto de dados internacional de imagens de lesões de pele. O ISIC é desenvolvido e mantido pela organização ISIC (*International Skin Imaging Collaboration*). Este conjunto de dados foi desenvolvido para facilitar a aplicação de técnicas digitais para reduzir a mortalidade por melanoma. Essas imagens digitais fornecidas podem ser utilizadas para educar profissionais ou, para desenvolver modelos automatizados relacionados às tarefas de classificação e detecção de melanomas.

O conjunto de dados ISIC 2016 contém 900 imagens para treinamento e cerca de 379 imagens para teste. As imagens do conjunto estão em formato .JPG com dimensões

diferentes, mas todas no formato RGB com 8 bits de intensidade para cada canal e 96 DPI de resolução horizontal e vertical. Os valores de referência, ou seja, as imagens segmentadas, estão todas com as mesmas dimensões de suas respectivas imagens, mas no formato .PNG em escala de cinza com 8 bits de intensidade, onde o valor de 255 indica a região da lesão e o valor 0 indica o fundo da imagem.

Para facilitar a avaliação dos resultados, as imagens dos conjuntos de treinamento e teste foram categorizados, da seguinte forma:

- Considerando suas especificidades e tamanho da região segmentada:
  - **Imagens Baixo Contraste (BC)**: São as imagens onde a diferença de contraste entre a lesão e o fundo é pequena;
  - **Imagens com Círculos (CI)**: São as imagens onde existe a presença de marcações circulares na imagem dermatoscópica;
  - **Imagens Normais (NO)**: São as imagens onde não há a presença de nenhum dos dois casos anteriores.

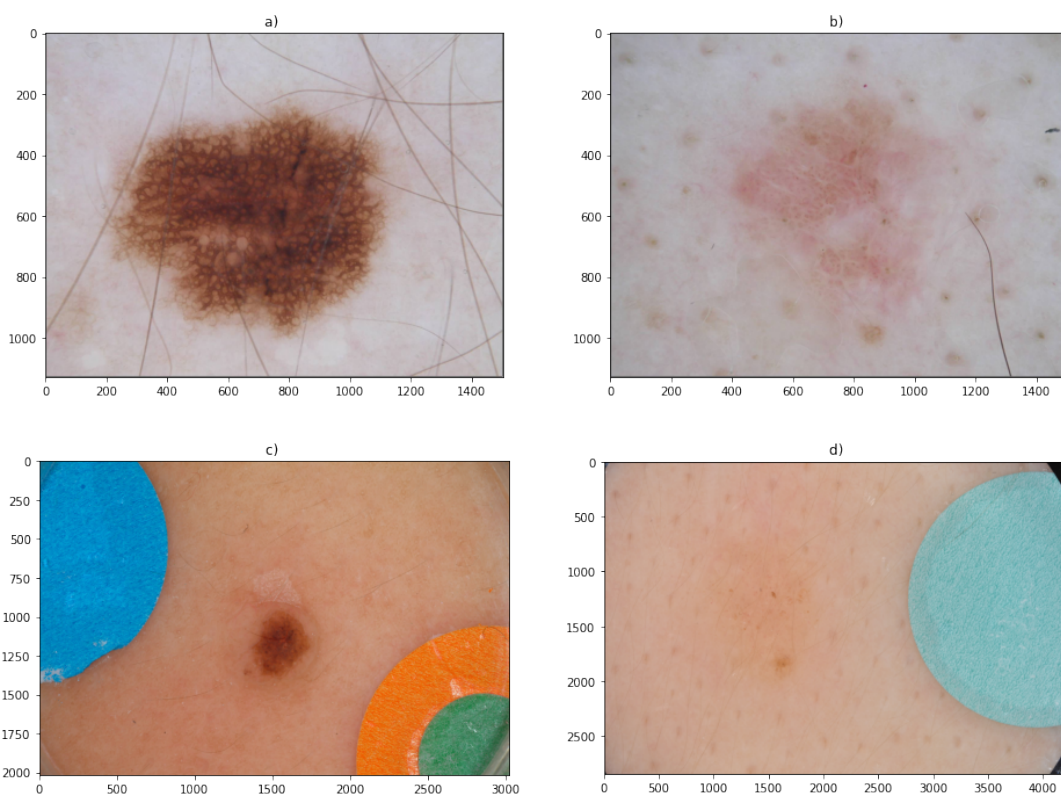
Esta categorização foi feita de forma manual via observação de cada imagem. Cabe indicar que as imagens podem pertencer as categorias BC e CI simultaneamente. Na Figura 19 são mostradas quatro imagens e suas respectivas categorias.

- Considerando o tamanho da região segmentada:
  - **Imagens Muito Grandes (MG)**: São as imagens onde a relação entre número de pixels da região segmentada e número de pixels total da imagem é maior ou igual a 0,75;
  - **Imagens Grandes (G)**: São as imagens onde esta relação é maior ou igual que 0,5, mas menor que 0,75;
  - **Imagens Pequenas (P)**: São as imagens onde esta relação é maior ou igual que 0,25, mas menor que 0,5;
  - **Imagens Muito Pequenas (MP)**: São as imagens onde esta relação é menor que 0,25.

Esta categorização foi feita de forma automática, tomando em conta as máscaras de segmentação já disponíveis no ISIC. É possível observar na Figura 20 exemplos de imagens em suas respectivas categorias de acordo com o tamanho da região da lesão.

Considerando a Tabela 2, em relação ao conjunto de treinamento, se conclui que:

Figura 19 – Exemplos de imagens divididas em categorias pela especificidade a) imagem normal, b) imagem com baixo contraste, c) imagem com marcações circulares, d) imagem com baixo contraste e marcações circulares.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

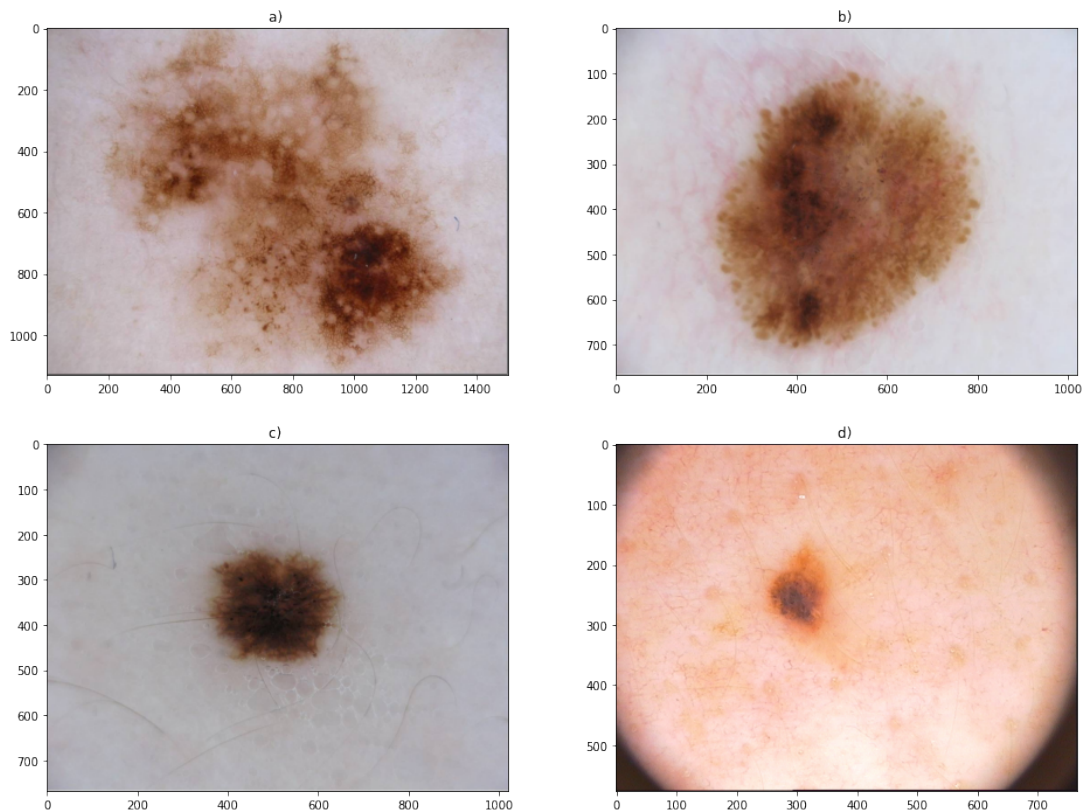
- 82, 22% são imagens classificadas como “Normais”, isso implica que o modelo pode ter resultados ruins em avaliar imagens com marcações circulares e/ou imagens com baixo contraste.
- 78, 44% das imagens tem uma região segmentada de tamanho “Grande” e “Pequena”.
- Apesar do conjunto de treinamento ter 900 imagens no total, o número total de imagens é 922 na Tabela 2, isso implica que 22 imagens estão simultaneamente nas categorias “Imagens com baixo contraste” e “Imagens com círculos”.

Tabela 2 – Número de imagens em cada categoria no conjunto de treinamento

	Muito grandes	Grandes	Pequenas	Muito pequenas	Total
Baixo contraste	7	8	33	13	61
Marcações circulares	0	1	58	62	121
Normais	84	293	313	50	740
Total	91	302	404	125	922

Considerando a Tabela 3, em relação ao conjunto de teste, se conclui que:

Figura 20 – Exemplos de imagens divididas em categorias pelo tamanho da região da lesão a) imagem muito grande, b) imagem grande, c) imagem pequena, d) imagem muito pequena.



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

- As imagens classificadas como “Normais” ocupam 81,79% e as imagens classificadas em relação a sua região segmentada como “Grande” e “Pequena” ocupam cerca de 73,61%.
- Apesar do conjunto de treinamento ter 379 imagens no total, o número total de imagens é 385 na Tabela 3, isso implica que 6 imagens estão simultaneamente nas categorias “Imagens com baixo contraste” e “Imagens com círculos”.

Tabela 3 – Número de imagens em cada categoria no conjunto de teste

	Muito grandes	Grandes	Pequenas	Muito pequenas	Total
Baixo contraste	0	9	5	9	23
Marcações circulares	0	1	14	26	41
Normais	47	117	133	24	321
Total	47	127	152	59	385

Em cada processo de treinamento, o conjunto de treinamento foi subdividido considerando a proporção 90%/10%, isto é, 90% das imagens do conjunto de treinamento permaneceram no conjunto de treinamento e 10% das imagens do conjunto de treinamento foram para o



conjunto de validação. Essa divisão implica que o conjunto de treinamento consiste em 810 imagens e o conjunto de validação em 90 imagens. Todas as imagens de entrada foram redimensionadas para  $512 \times 512 \times 3$  e as máscaras para  $512 \times 512 \times 1$ , exceto quando mencionado o contrário.

### 4.3 Métrica de avaliação

Para fazer a avaliação do modelo de segmentação, é utilizada a métrica Índice de Jaccard conhecido também como coeficiente de similaridade de Jaccard. Esta métrica é utilizada para medir a similaridade e diferença entre a máscara prevista e a máscara de referência e é definida como:

$$J(\mathbf{M}, \hat{\mathbf{M}}) = \frac{|\mathbf{M} \cap \hat{\mathbf{M}}|}{|\mathbf{M} \cup \hat{\mathbf{M}}|}. \quad (4.1)$$

onde  $\mathbf{M}$  é a máscara binária de referência e  $\hat{\mathbf{M}}$  é a máscara prevista pelo modelo. O  $|\cdot|$  é o operador de cardinalidade. Importante destacar que o Índice de Jaccard é uma métrica normalizada, já que o valor resultante está no intervalo entre 0 e 1, onde 1 indica que a máscara prevista tem a máxima similaridade com a máscara de referência (a interseção das duas máscaras é igual a união das máscaras) e 0 indica que as duas máscaras são totalmente diferentes entre si (a interseção das máscaras é um conjunto vazio).

### 4.4 Modelo Base

Para ter uma referência de como as técnicas e algoritmos de DA impactam nos resultados, é necessário ter um modelo onde não foi aplicado nenhuma técnica de DA.

O modelo de segmentação utilizado foi inteiramente baseado na arquitetura do modelo U-net, não sendo utilizado nenhuma rede CNN como *backbone*<sup>1</sup>. Outro ponto importante é que a rede foi treinado com pesos iniciais aleatórios, ou seja, não foi utilizado *transfer learning* ou alguma rede pré-treinada como ponto inicial. Informações adicionais sobre a implementação do modelo encontram-se no Anexo B.

<sup>1</sup> Backbone: É a arquitetura da rede que é usada para a extração dos mapas de características de uma tarefa alvo. Em visão computacional, são principalmente usadas em CNNs que foram desenvolvidas para lidar com tarefas de classificação de imagens. Arquiteturas de redes populares incluem AlexNet, VGG-16, ResNets, DarkNet-19, GoogLeNet, MobileNet, ShuffleNet, etc.

Importante destacar que a última camada do modelo implementado tem como função de ativação a função sigmoide, ou seja, os valores previstos pelo modelo para cada um dos pixels da imagem é um único valor entre 0 e 1, implicando na necessidade de usar um limiar,  $\tau$ , para definir o rótulo de classe do pixel. No desafio proposto pelo ISIC 2016 (GUTMAN et al., 2016), não há um valor pré-determinado para o limiar, então, foi utilizado o valor de limiar de 0,5 definido no desafio ISIC 2017 (CODELLA et al., 2017). A Equação 4.2 demonstra a aplicação do limiar  $\tau$  nos pixels previstos.

$$\hat{M}_{i,j} = \begin{cases} 0 & \text{se } \hat{Y}_{i,j} < \tau \\ 1 & \text{se } \hat{Y}_{i,j} \geq \tau \end{cases} \quad (4.2)$$

onde  $\hat{Y}$  é a saída prevista pelo modelo.

Finalmente, considerando que o problema de segmentação aqui tratado é definido como uma classificação binária a nível de pixels, aqui foi utilizada a função de perda entropia cruzada, também conhecido como *binary cross-entropy* ou *log loss*. A função calcula a diferença entre o valor rotulado corretamente e o valor previsto.

Na Tabela 4 são resumidos os valores para os hiperparâmetros usados no treinamento do modelo base como o número de épocas e a Paciência<sup>2</sup>. O modelo foi treinado usando o otimizador Adam, com uma taxa de aprendizado de 0,005.

Tabela 4 – Hiperparâmetros de treinamento do modelo base

Épocas	Paciência	Lotes por época	Tamanho do lote
60	10	50	16

A Tabela 5 mostra os resultados do modelo base. Tal modelo obteve uma média total no índice de Jaccard de 0,7725. Também é possível observar que: (i) as imagens com baixo contraste foram as que obtiveram a menor média (0,6533), seguida das imagens com círculos (0,6790); (ii) Em relação ao tamanho, as imagens muito pequenas apresentaram a menor média (0,6357). Por fim, existe pelo menos uma imagem em que o índice de Jaccard foi igual a zero, indicando que nenhum dos pixels previsto pelo modelo foi correto.

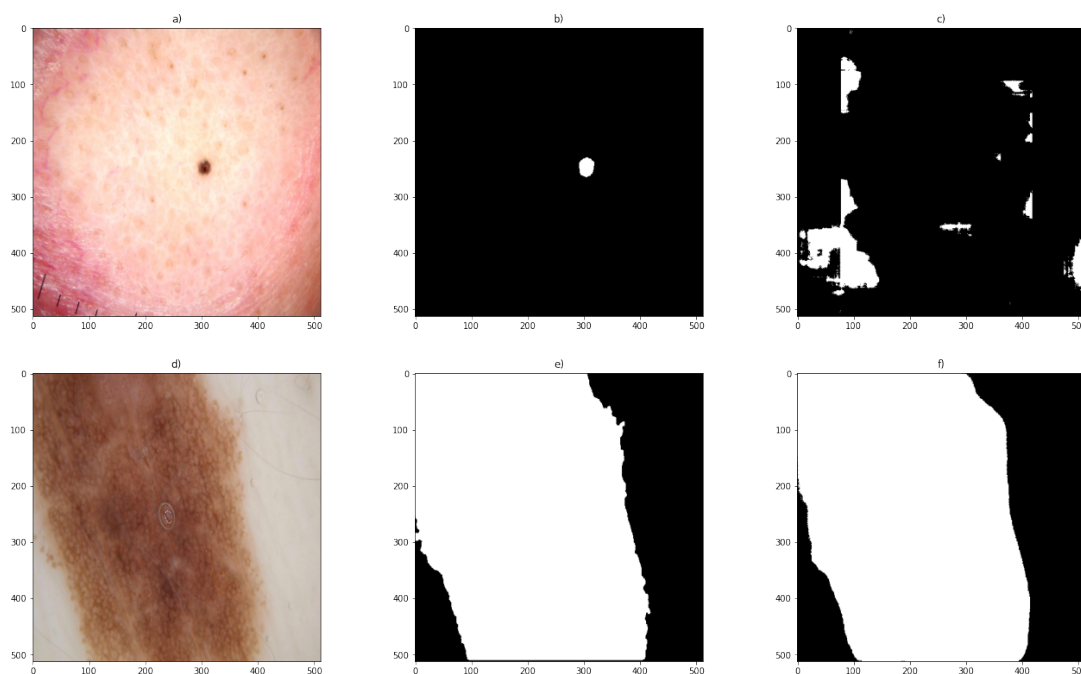
É possível observar na Figura 21 as imagens que obtiveram o melhor e o pior resultado no Índice de Jaccard. Nesta figura é possível observar que a imagem de entrada que teve o melhor resultado foi uma da categoria MG e NO e a imagem que teve o pior resultado foi uma imagem da categoria MP e NO.

<sup>2</sup> Paciência refere-se ao número de épocas a mais que a rede continuará sendo treinada sem que haja alguma melhoria. Também é conhecido como parada precoce ou *Early Stop*.

Tabela 5 – Resultados do modelo base

	Total	NO	CI	BC	MG total	G total	P total	MP total
Núm. de im.	379	321	41	23	47	127	151	54
Média	0,7725	0,7904	0,6790	0,6533	0,8207	0,8195	0,7669	0,6357
Desvio padrão	0,1893	0,1710	0,2578	0,2449	0,1276	0,1309	0,1864	0,2753
Mínimo	0,0000	0,0000	0,0351	0,0351	0,4354	0,2203	0,1632	0,0000
25 Percentil	0,7155	0,7287	0,6186	0,5617	0,7673	0,7592	0,6732	0,5008
50 Percentil	0,8355	0,8445	0,7729	0,7244	0,8495	0,8606	0,8372	0,7203
75 Percentil	0,9098	0,9156	0,8818	0,8233	0,9212	0,9182	0,9064	0,8752
Máximo	0,9720	0,9720	0,9334	0,9235	0,9720	0,9674	0,9582	0,9276

Figura 21 – Imagens no geral: a) imagem original, b) máscara de referência, c) máscara prevista (IJ = 0,0000), d) imagem original, e) máscara de referência, f) máscara prevista (IJ = 0,9720).



Fonte: Imagens adaptadas de (GUTMAN et al., 2016).

#### 4.5 Avaliação individual de cada técnica de DA selecionadas manualmente

Para ter uma visão clara de como cada técnica de DA impacta nos resultados do modelo base, foi feita uma avaliação individual aplicando a seguinte metodologia:

- Cada uma das técnicas de DA é aplicada sobre o conjunto de treino original.
- É gerado o conjunto de treinamento aumentado, de modo que ele contenha 1.800 imagens, sendo 900 das originais e 900 das geradas pela técnica de DA em estudo. Dessas 1.800, 180 das imagens originais do conjunto foram usadas para validação.

- É efetuado o treinamento do modelo base usando o conjunto de treinamento aumentado e a avaliação do mesmo usando o conjunto de teste original.

A Tabela 6 contém todos os hiperparâmetros de treinamento para cada uma das técnicas de DA. Cabe indicar que: (i) as técnicas geométricas foram testadas com preenchimentos com zeros, sendo estas com a sigla PZ, e com preenchimento com valor médio do fundo da imagem, sendo estas com a sigla PVM; (ii) as técnicas onde valores de controle eram aplicados, foram escolhidos os intervalos dos valores de acordo com as recomendações da literatura e os valores para cada uma das imagens foi escolhido de forma aleatória dentro desse intervalo. (iii) as técnicas com a sigla VC indicam valores constantes, ou seja, os hiperparâmetros constantes utilizados em cada uma das funções de transformação. A tabela contendo todos os valores constantes usados nesse trabalho encontra-se no Anexo C.

Tabela 6 – Hiperparâmetros de treinamento das técnicas de DA individuais

	Intervalo de valores	Conjunto de treinamento	Épocas	Paciência
Flipping	VC	1800	60	15
Rotação PVM	$[-15^\circ, 15^\circ]$	1800	60	15
Rotação PZ	$[-15^\circ, 15^\circ]$	1800	60	15
Translação PVM	$[-10\%, 10\%]$	1800	60	15
Translação PZ	$[-10\%, 10\%]$	1800	60	15
Alterar brilho	$[-20\%, 20\%]$	1800	60	15
Alterar contraste	$[-20\%, 20\%]$	1800	60	15
Alterar matiz	$[10, 180]$	1800	60	15
Alterar saturação	$[-20, 20]$	1800	60	15
Alterar intensidade	$[-20, 20]$	1800	60	15
Balanço de branco	VC	1800	60	15
Equalização de histograma	VC	1800	60	15
Inverter	VC	1800	60	15
Solarizar	VC	1800	60	15
Aguçamento	VC	1800	60	15
Suavização	VC	1800	60	15
Ruído gaussiano	VC	1800	60	15
Ruído sal e pimenta	VC	1800	60	15
Equalização específica	$[1]$	1800	60	15
Equalização específica 2	$[0.1, 0.5], (0.5, 0.9]$	2221	75	15
Random cutout	VC	1800	60	15
Random mix-up	$[0, 1]$	1800	60	15

As Tabelas 7, 8, 9 e 10 mostram os resultados obtidos pela aplicação de cada técnica de DA. As técnicas de DA que atingiram uma média total igual ou maior a sem DA foram marcadas de verde, nessa comparação foi desconsiderada a última casa decimal. A seguir alguns comentários sobre cada tabela:

- A Tabela 7 mostra os resultados obtidos com as técnicas de DA do tipo geométricas. É possível observar que **a técnica que obteve melhores resultados foi a rotação com preenchimento de zeros(PZ)**.
- A Tabela 8 mostra os resultados obtidos com as técnicas de DA do tipo espaço de cor. É possível observar que **as técnicas que alteraram o brilho, a intensidade e a saturação, respectivamente, obtiveram os melhores resultados**. Um ponto a destacar é que a técnica de alterar o brilho fez com que as médias na categoria CI e na MP fossem as maiores de todas as técnicas individuais testadas. Outro ponto é que, mesmo que a técnica de balanço de branco tenha obtido uma média total abaixo do modelo base, essa técnica proporcionou as maiores médias para as categorias BC, MG e G.
- A Tabela 9 mostra os resultados obtidos com as técnicas de DA do tipo filtro. É possível observar que **as técnicas que aplicaram o filtro de aguçamento, de suavização e aplicaram ruído do tipo sal e pimenta, respectivamente, obtiveram os melhores resultados**. Dessas três técnicas, o destaque é para a técnica de suavização, que permitiu ter a maior média total de todas as técnicas testadas. Outro ponto a destacar é que a técnica de aguçamento obteve a maior média nas categorias NO e P.
- A Tabela 10 mostra os resultados obtidos com as técnicas de DA do tipo algoritmo. É possível observar que **as técnicas de Random Mix-up e Equalização específica-2 foram as que tiveram os melhores resultados**. A equalização específica-2 difere-se da equalização específica-1 pelo valor de controle que permite diferenças mais sutis entre as imagens de baixo contraste, permitindo o melhor treinamento, entretanto, mesmo com esse cuidado extra os resultados não foram muito diferentes do modelo base.

Tabela 7 – Resultados para as técnicas de DA geométricas

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Flipping	0,7688	0,1975	0,7970	0,6231	0,5851	0,7945	0,8136	0,7795	0,6113
Rotação PVM	0,6931	0,2098	0,7352	0,3934	0,5490	0,8113	0,8135	0,6548	0,4139
Rotação PZ	<b>0,7857</b>	0,1784	0,8044	0,7041	0,6396	0,8022	0,8252	0,7953	0,6518
Translação PVM	0,5468	0,2775	0,6013	0,2069	0,3012	0,7374	0,6733	0,4934	0,2326
Translação PZ	0,7667	0,1851	0,7930	0,6239	0,5990	0,7911	0,8053	0,7804	0,6166

#### 4.6 Avaliação de grupos de técnicas de DA selecionados manualmente

Para avaliar o impacto dos grupos de técnicas de DA, foram testadas diferentes combinações, algumas apenas para avaliar o comportamento global de um grupo e para tentar maximizar algumas das categorias secundárias. Para isso, foi aplicada a seguinte metodologia:

Tabela 8 – Resultados para as técnicas de DA no espaço de cor

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Alterar brilho	0,7867	0,1713	0,8042	0,7115	0,6116	0,7865	0,8184	0,7963	0,6857
Alterar contraste	0,7561	0,2004	0,7892	0,5623	0,5724	0,7788	0,8029	0,7695	0,5891
Alterar matiz	0,7599	0,2021	0,7872	0,6285	0,5750	0,7954	0,8050	0,7609	0,6198
Alterar saturação	0,7818	0,1847	0,8052	0,6666	0,6174	0,8034	0,8183	0,7871	0,6619
Alterar intensidade	0,7852	0,1725	0,8016	0,7096	0,6574	0,7960	0,8292	0,7912	0,6558
Balço de branco	0,7637	0,2017	0,7890	0,5999	0,6635	0,8271	0,8368	0,7462	0,5854
Equalização de histograma	0,7277	0,2179	0,7667	0,4507	0,5933	0,7688	0,8112	0,7310	0,4861
Inverter	0,7587	0,1879	0,7902	0,5758	0,6126	0,7308	0,8222	0,7717	0,5975
Solarizar	0,7384	0,1958	0,7622	0,6399	0,5429	0,7010	0,7706	0,7643	0,6227

Tabela 9 – Resultados para as técnicas de DA com filtros

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Aguçamento	0,7822	0,1873	0,8068	0,6552	0,6219	0,7941	0,8234	0,7984	0,6294
Suavização	0,7878	0,1792	0,8061	0,6990	0,6188	0,8097	0,8200	0,7930	0,6789
Ruído gaussiano	0,7455	0,2013	0,7741	0,5948	0,5534	0,7557	0,7973	0,7596	0,5753
Ruído sal e pimenta	0,7764	0,1779	0,7939	0,7064	0,6148	0,7630	0,8057	0,7922	0,6749

Tabela 10 – Resultados para as técnicas de DA com algoritmos

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Equalização específica-1	0,7340	0,2161	0,7811	0,4307	0,4872	0,7888	0,8064	0,7598	0,4437
Equalização específica-2	0,7723	0,1936	0,7931	0,6654	0,6314	0,8215	0,8233	0,7671	0,6242
Random cutout	0,7666	0,2011	0,7896	0,6768	0,5426	0,7800	0,7944	0,7708	0,6776
Random mix-up	0,7784	0,1789	0,8007	0,6630	0,6199	0,8097	0,8292	0,7808	0,6246

- Cada uma das técnicas de DA de um grupo é selecionada de forma aleatória e aplicada sobre o conjunto de treino original, gerando assim novas imagens que serão adicionadas ao conjunto de treinamento aumentado. Todas as técnicas selecionadas de forma aleatória tem probabilidade igual de serem selecionadas. Não ocorre mesclagem entre as técnicas, ou seja, cada técnica é aplicada individualmente em cada imagem.
- O conjunto de treinamento aumentado gerado contém 1.800 imagens, sendo 900 das originais e 900 geradas pelas técnicas de DA do grupo. Dessas 1.800, 180 das imagens originais do conjunto foram usadas para validação.
- Por fim, o modelo base é retreinado usando o conjunto de treinamento aumentado e avaliado usando o conjunto de teste original.

A Tabela 11 mostra os hiperparâmetros de treinamento usados com cada grupo de técnicas de DA estudada.

Tabela 11 – Hiperparâmetros de treinamento para grupos de técnicas estudadas

	Conjunto de treinamento	Épocas	Paciência
Grupo de técnicas geométricas	1800	60	15
Grupo de técnicas no espaço de cor	1800	60	15
Grupo de técnicas por filtro	1800	60	15
Grupo de técnicas por algoritmo	1800	60	15

A Tabela 12 mostra os resultados obtidos para cada grupo estudado. É possível observar que os grupos de técnicas por filtro e por algoritmo tiveram resultados acima do modelo de base. Um ponto a destacar é que o grupo das técnicas por algoritmo obteve a maior média para imagens da categoria MP.

Tabela 12 – Resultados obtidos para os grupos das técnicas de DA estudadas

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Todas as técnicas geométricas	0,7637	0,1883	0,7898	0,6508	0,5223	0,7430	0,7871	0,7941	0,6417
Todas as técnicas no espaço de cor	0,7498	0,1970	0,7842	0,5762	0,5331	0,7458	0,7826	0,7757	0,6032
Todas as técnicas por filtro	0,7822	0,1882	0,8018	0,6900	0,6363	0,8071	0,8250	0,7899	0,6388
Todas as técnicas por algoritmo	0,7765	0,1843	0,7994	0,6839	0,5774	0,7719	0,8149	0,7860	0,6637

#### 4.7 Avaliação individual de cada técnica de DA selecionadas automaticamente

Devido ao tempo de execução para buscar os valores que entregam os melhores resultados, são necessárias algumas modificações dos hiperparâmetros utilizados anteriormente no treinamento do modelo base. Apenas durante a busca bayesiana, as imagens do conjunto de treinamento e conjunto de teste foram redimensionadas para  $128 \times 128 \times 3$  para as imagens de entrada e  $128 \times 128 \times 1$  para as máscaras respectivas. O número de épocas e a paciência para cada tentativa foi reduzido para 50 e 5, respectivamente.

Antes de realizar a busca pelo melhor grupo de técnicas de DA, é necessário encontrar os melhores valores para cada uma das técnicas de DA. Para isso, foi realizado um estudo prévio utilizando a seguinte metodologia:

- Para cada técnica de DA foi feito uma busca bayesiana, através do Optuna, do valor da variável de controle da técnica, onde era aplicado, dentro de um intervalo definido.
- Para cada imagem do conjunto de treinamento original foi gerado uma nova imagem que era adicionada ao conjunto de treinamento aumentado.
- O novo conjunto de treinamento continha então, 1800 imagens, sendo 900 das originais e 900 geradas pela técnica em estudo. Dessas 1800, 180 das imagens originais do conjunto foram usadas para validação.
- É efetuado o treinamento do modelo base usando o conjunto de treinamento e a avaliação do mesmo usando o conjunto de teste padrão.
- Através dos dados de cada conjunto treinado e das avaliações provenientes de cada modelo base treinado, o valor que convergiu para a maior média dentro do intervalo foi escolhido como melhor valor para a técnica em estudo.

A Tabela 13 explicita os intervalos de busca, os passos de cada uma das tentativas e o valor ótimo encontrado.

Tabela 13 – Melhores valores encontrados para cada uma das técnicas de DA

	Intervalo de busca	Passo	Melhor valor encontrado
Rotação	[-15, 15]	0,3	-6,0
Translação vertical	[-10, 10]	0,2	-9,2
Translação horizontal	[-10, 10]	0,2	-9,2
Flipping vertical	VC	-	-
Flipping horizontal	VC	-	-
Ajuste de Brilho	[-20, 20]	0,4	-10,4
Ajuste de Contraste	[-20, 20]	0,4	-3,6
Ajuste de Intensidade	[-20, 20]	1,4	0,8
Ajuste de Matiz	[10, 170]	2,0	30,0
Ajuste de Saturação	[-20, 20]	1,4	11,6
Balanço de branco	VC	-	-
Equalização de histograma	VC	-	-
Inversão	VC	-	-
Solarização	VC	-	-
Aguçar	VC	-	-
Suavização	VC	-	-
Ruído Gaussiano	VC	-	-
Ruído Sal e Pimenta	VC	-	-
Equalização específica	[0, 1]	0,01	0,9
Random MixUp	[0, 1]	0,01	0,6

#### 4.8 Avaliação do grupo de técnicas de DA selecionadas automaticamente

Serão efetuados quatro experimentos que abordam as estratégias propostas:

- O primeiro experimento visa encontrar o melhor grupo de técnicas de DA dentre as técnicas de DA que atingiram um resultado de média total acima do modelo base. Este experimento usa uma abordagem semiautomatizada, já que antes de buscar o melhor grupo de técnicas, é necessário fazer um estudo das técnicas individualmente e realizar um estudo dos melhores valores para cada uma das técnicas. O objetivo desse experimento é ver quais técnicas, onde cada uma delas é aplicada individualmente nas imagens do conjunto de treinamento individual, melhoram os resultados do modelo.



O conjunto de treinamento aumentado utilizado neste experimento é o MCCTA-1, ou seja, imagens onde as técnicas de DA encontradas são aplicadas individualmente.

- O segundo experimento visa encontrar o melhor grupo de técnicas de DA que, quando mescladas, melhoram os resultados do modelo. Para realizar esta mescla, o conjunto de treinamento aumentado utilizado neste experimento é o MCCTA-2. Assim como no primeiro experimento, apenas as técnicas de DA que atingiram um resultado de média total acima do modelo base foram utilizadas. Este experimento usa uma abordagem semiautomatizada, já que antes de buscar o melhor grupo de técnicas, é necessário fazer um estudo das técnicas individualmente e realizar um estudo dos melhores valores para cada uma das técnicas. Esse experimento diferencia-se do primeiro especialmente pelo ponto de haver mescla nas técnicas encontradas.
- O terceiro experimento visa encontrar o melhor grupo de técnicas de DA e, para cada técnica encontrada dentro desse grupo, um subgrupo de técnicas que será aplicado em conjunto com esta técnica encontrada. O conjunto de treinamento aumentado utilizado neste experimento é o MCCTA-3. Assim como no segundo experimento: (i) as técnicas buscadas foram apenas as técnicas de DA que atingiram um resultado de média total acima do modelo base; (ii) este experimento também usa uma abordagem semiautomatizada. O objetivo deste experimento é avaliar a possibilidade de que cada técnica do melhor grupo de técnicas detinha um subgrupo de técnicas de DA que, quando aplicado em conjunto melhorava os resultados do modelo.
- O quarto experimento visa, assim como o experimento 1, encontrar o melhor grupo de técnicas de DA. O objetivo deste experimento é avaliar a possibilidade de não ser necessário fazer uma avaliação prévia das técnicas e a possibilidade de alterar outros hiperparâmetros de treinamentos para melhorar os resultados do modelo. O conjunto de treinamento aumentado utilizado neste experimento é o MCCTA-1, ou seja, imagens onde as técnicas de DA encontradas são aplicadas individualmente. Devido às estratégias, parte deste experimento é semiautomatizado, já que ainda foi necessário realizar um estudo dos melhores valores para cada uma das técnicas. Este experimento diferencia-se dos demais por dois fatores: (i) foi utilizado todas as técnicas de DA, não apenas as que obtiveram média total acima do modelo base; (ii) a taxa de aprendizado da rede também é otimizada.

A seguir são mostrados os resultados obtidos em cada experimento orientado a encontrar o melhor grupo de técnicas de DA.

### 4.8.1 Resultados do primeiro experimento

Na Tabela 14 é possível observar o melhor grupo de técnicas de DA encontradas por cada uma das estratégias, assim como o número total de imagens adicionadas ao conjunto de treinamento.

Tabela 14 – Hiperparâmetros encontrados no primeiro experimento para cada estratégia

	Estratégia 1	Estratégia 2	Estratégia 3	Estratégia 4
Buscar valores	Não	Sim	Não	Sim
Busca número de imagens	Não	Não	Sim	Sim
Coleção de otimização	100	200	200	300
Melhor grupo	(Rotação; Translação)	(Translação)	(Equal. Esp.; Rotação; Translação)	(Brilho; Rotação; Sal e Pimenta)
Número de imagens	(500; 500)	(500)	(150; 850; 350)	(100; 600; 350)
Valores encontrados	Melhores valores	(-12; )	Melhores valores	(14,50; 9,0; -)
Conjunto de treinamento	1900	1400	2250	1950

\* **Melhores valores** indica que os valores utilizados foram idênticos aos melhores valores encontrados anteriormente.

Na Tabela 15 é possível observar os resultados do primeiro experimento. Considerando o grupo de técnicas de DA obtidas por cada estratégia, foi retreinado o modelo base considerando 100 épocas e 15 de paciência. Pode-se observar que, todas as médias estão acima do modelo base original, onde a estratégia 2 obteve o maior desempenho. Um ponto importante é o fato da Estratégia 2 ter utilizado apenas a técnica de Translação. Diferente da avaliação individual de técnicas descrita na Tabela 6, o valor do hiperparâmetro da translação fez uma diferença no resultado.

Tabela 15 – Resultados encontrados no primeiro experimento para cada estratégia

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Estratégia 1	0,7926	0,1752	0,8111	0,7019	0,6603	0,8214	0,8335	0,7937	0,6681
Estratégia 2	0,7952	0,1674	0,8139	0,7136	0,6319	0,8075	0,8293	0,8032	0,6821
Estratégia 3	0,7938	0,1703	0,8143	0,7009	0,6049	0,8008	0,8346	0,8020	0,6685
Estratégia 4	0,7839	0,1853	0,8044	0,6581	0,6523	0,8399	0,8432	0,7735	0,6251

### 4.8.2 Resultados do segundo experimento

Na Tabela 16 é possível observar o melhor grupo de técnicas de DA encontradas por cada uma das estratégias, assim como o número total de imagens adicionadas ao conjunto de treinamento.

A Tabela 17 mostra os resultados do segundo experimento. Considerando o grupo de técnicas de DA obtidas por cada estratégia, foi retreinado o modelo base considerando 100 épocas e 15 de paciência. Pode-se observar que, apenas as estratégias 2 e 3, apresentaram uma média acima do modelo base. A estratégia 3 ficou ligeiramente acima do modelo base,

Tabela 16 – Hiperparâmetros encontrados no segundo experimento para cada estratégia

	Estratégia 1	Estratégia 2	Estratégia 3	Estratégia 4
Buscar valores	Não	Sim	Não	Sim
Busca número de imagens	Não	Não	Sim	Sim
Coleção de otimização	300	300	300	300
Melhor grupo	(Contraste; Intensidade; Saturação; Aquecimento; Sal e Pimenta; Equal. Esp.; MixUp)	(Translação; Contraste; Saturação; Aquecimento; Sal e Pimenta)	(Rotação; Brilho; Contraste; Intensidade; Saturação)	(Rotação; Brilho; Intensidade; Saturação; Aquecimento)
Número de imagens	(500)	(500)	(1000)	(700)
Valores encontrados	Melhores valores	(-6,0; 7,5; 2,0; -;-)	Melhores valores	(-3,5; 13,5; 19,0; 2,5; -)
Conjunto de treinamento	1400	1400	2500	1600

\* **Melhores valores** indica que os valores utilizados foram idênticos aos melhores valores encontrados anteriormente.

enquanto a estratégia 2 teve um resultado equiparável ao melhor resultado de cada técnica e de cada grupo de técnicas avaliadas manualmente.

Tabela 17 – Resultados encontrados no segundo experimento para cada estratégia

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Estratégia 1	0,7673	0,1917	0,7914	0,6097	0,6698	0,7951	0,8306	0,7735	0,5770
Estratégia 2	0,7853	0,1728	0,8013	0,7094	0,6564	0,8195	0,8208	0,7895	0,6601
Estratégia 3	0,7768	0,1908	0,7985	0,6697	0,6189	0,8288	0,8152	0,7757	0,6443
Estratégia 4	0,7667	0,1937	0,7998	0,5779	0,5801	0,7957	0,8244	0,7830	0,5602

### 4.8.3 Resultados do terceiro experimento

Na Tabela 18 é possível observar o melhor grupo de técnicas de DA encontradas por cada uma das estratégias, assim como o número total de imagens adicionadas ao conjunto de treinamento.

A Tabela 19 mostra os resultados do terceiro experimento. Considerando o grupo de técnicas de DA obtidas por cada estratégia, foi retreinado o modelo base considerando 100 épocas e 15 de paciência. Pode-se observar que, as estratégias 2, 3 e 4 tiveram uma média total acima do modelo base original. Entretanto, apenas a estratégia 4 apresentou resultados comparáveis aos resultados do primeiro experimento. Outro ponto é que, quando comparado aos modelos do primeiro experimento, as estratégias do terceiro experimento necessitam de um conjunto de treinamento maior, quase o dobro de imagens, para ter resultados comparáveis.

Tabela 18 – Hiperparâmetros encontrados no terceiro experimento para cada estratégia

	Estratégia 1	Estratégia 2	Estratégia 3	Estratégia 4
Buscar valores	Não	Sim	Não	Sim
Busca número de imagens	Não	Não	Sim	Sim
Coleção de otimização	300	300	300	300
Melhor grupo	(Brilho/916; Equalização específica/235; Rotação/787; Saturação/596; Transla- ção/348)	(Random Mix- Up/917)	(Aguçar/837; Brilho/44; Contraste/224; Equalização específica/206; Intensida- de/867; Rota- ção/711; Trans- lação/393)	(Aguçar/165; Brilho/577; Contraste/70; Equalização específica/163; Random Mix- Up/541; Trans- lação/429)
Número de imagens	(500; 500; 500; 500; 500)	(500)	(500; 200; 400; 700; 100; 500; 400)	(200; 600; 300; 400; 700; 500)
Valores encontrados	Melhores valores	(-5,5; 4,5; -8,0; 5,0; -7,5; 0,0; -; -; -; 1,0; 0,4)	Melhores valores	(12,5; -10,0; 12,5; -5,0; -11,0; 9,5; -; -; -; 0,3; 0,6)
Conjunto de treinamento	3400	1400	3700	3600

Tabela 19 – Resultados encontrados no terceiro experimento para cada estratégia

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Estratégia 1	0,7688	0,1849	0,7905	0,7067	0,5206	0,7581	0,7893	0,7876	0,6773
Estratégia 2	0,7744	0,1912	0,8031	0,6338	0,5247	0,7678	0,8099	0,7882	0,6579
Estratégia 3	0,7832	0,1814	0,8060	0,6819	0,5991	0,7917	0,8103	0,7962	0,6756
Estratégia 4	0,7901	0,1796	0,8088	0,7203	0,5966	0,8114	0,8217	0,7917	0,6929

#### 4.8.4 Resultados do quarto experimento

Na Tabela 20 é possível observar o melhor grupo de técnicas de DA encontradas por cada uma das estratégias, assim como o número total de imagens adicionadas ao conjunto de treinamento e os valores dos hiperparâmetros encontrados para cada uma das estratégias. Cabe indicar que, o intervalo de busca para a taxa de aprendizado foi de 0,0001 até 0,0010 com passo de 0,0001.

A Tabela 21 mostra os resultados do quarto experimento. Considerando o grupo de técnicas de DA obtidas por cada estratégia, foi retreinado o modelo base considerando 200 épocas e 25 de paciência. O número de épocas e a paciência foi aumentado, já que, durante o treinamento, o modelo continuava a melhorar após atingir o número de épocas anterior. Pode-se observar que, todas as estratégias, exceto a estratégia 3, atingiram um média total acima de 0,8. Esses resultados implicam nas maiores médias encontradas, tanto na média total como nas médias em cada categoria, em todas as avaliações realizadas.

Tabela 20 – Hiperparâmetros encontrados no quarto experimento para cada estratégia

	Estratégia 1	Estratégia 2	Estratégia 3	Estratégia 4
Buscar valores	Não	Sim	Não	Sim
Busca número de imagens	Não	Não	Sim	Sim
Coleção de otimização	500	500	500	500
Melhor grupo	(Rotação; Translação Vertical; Translação horizontal; Flipping horizontal; Balanço de branco; Brilho; Contraste; Equalização de histograma; Saturação; Solarização; Suavização; Gaussiano)	(Translação vertical; Flipping vertical; Flipping horizontal; Balanço de branco; Brilho; Contraste; Equalização de histograma; Inverter; Intensidade; Matiz; Saturação; Solarização; Aguçar; Suavização)	(Rotação; Intensidade; Suavização; Translação vertical; Translação horizontal; Flipping horizontal; Balanço de branco; Equalização de histograma; Inverter)	(Rotação; Flipping vertical; Flipping horizontal; Balanço de branco; Brilho; Contraste; Equalização de histograma; Matiz; Solarização; Aguçar; Suavização)
Número de imagens	(500; 500; 500; 500; 500; 500; 500; 500; 500; 500)	(500; 500; 500; 500; 500; 500; 500; 500; 500)	(550; 600; 600; 150; 250; 250; 150; 550; 750; 500; 750; 300; 350; 150)	(800; 350; 250; 300; 700; 350; 550; 500; 350; 800; 550)
Valores encontrados	Melhores valores	(-13,0; -14,0; -; -13,0; 5,5; 7,5; -; -; -)	Melhores valores	(13,0; -14,5; -3,5; -; -16,5; -11,0; 1,0; 160,0; -; -; 1,0)
Taxa de aprendizado	0,0008	0,0008	0,0006	0,0006
Conjunto de treinamento	6900	5400	5400	6400

Tabela 21 – Resultados encontrados no quarto experimento para cada estratégia

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Estratégia 1	0,8015	0,1580	0,8140	0,7561	0,6982	0,8110	0,8365	0,8040	0,7040
Estratégia 2	0,8040	0,1604	0,8191	0,7487	0,6708	0,8089	0,8391	0,8081	0,7060
Estratégia 3	0,7951	0,1612	0,8075	0,7495	0,6701	0,8250	0,8267	0,7958	0,6929
Estratégia 4	0,8032	0,1518	0,8158	0,7673	0,6877	0,7941	0,8254	0,8152	0,7254

## 4.9 Comparação dos resultados

A Tabela 22 mostra os resultados alcançados por todas as avaliações realizadas. Os campos realçados em verde são os valores que ficaram acima do valor do modelo base e os campos realçados em azul são os maiores valores de cada uma das categorias.

Quando comparado ao modelo base, é possível chegar a resultados melhores aplicando ao menos uma técnica de DA individual manualmente. As técnicas individuais que atingiram os melhores resultados foram: **Rotação com preenchimento de zeros** (0,7857), **Alteração do brilho** (0,7867), **Filtro de Suavização** (0,7878) e **Random Mix-Up** (0,7784).

Quando é avaliado o grupo de técnicas divididas por categorias, é possível observar que apenas as técnicas por filtro e por algoritmo obtiveram uma média superior ao modelo base.

Tabela 22 – Resultados obtidos pelas diferentes avaliações realizadas

	Média	DP	Média NO	Média CI	Média BC	Média MG	Média G	Média P	Média MP
Modelo base	0,7725	0,1893	0,7904	0,6790	0,6533	0,8207	0,8195	0,7669	0,6357
Flipping	0,7688	0,1975	0,7970	0,6231	0,5851	0,7945	0,8136	0,7795	0,6113
Rotação PVM	0,6931	0,2098	0,7352	0,3934	0,5490	0,8113	0,8135	0,6548	0,4139
Rotação PZ	0,7857	0,1784	0,8044	0,7041	0,6396	0,8022	0,8252	0,7953	0,6518
Translação PVM	0,5468	0,2775	0,6013	0,2069	0,3012	0,7374	0,6733	0,4934	0,2326
Translação PZ	0,7667	0,1851	0,7930	0,6239	0,5990	0,7911	0,8053	0,7804	0,6166
Alterar brilho	0,7867	0,1713	0,8042	0,7115	0,6116	0,7865	0,8184	0,7963	0,6857
Alterar contraste	0,7561	0,2004	0,7892	0,5623	0,5724	0,7788	0,8029	0,7695	0,5891
Alterar matiz	0,7599	0,2021	0,7872	0,6285	0,5750	0,7954	0,8050	0,7609	0,6198
Alterar saturação	0,7818	0,1847	0,8052	0,6666	0,6174	0,8034	0,8183	0,7871	0,6619
Alterar intensidade	0,7852	0,1725	0,8016	0,7096	0,6574	0,7960	0,8292	0,7912	0,6558
Balanco de branco	0,7637	0,2017	0,7890	0,5999	0,6635	0,8271	0,8368	0,7462	0,5854
Equalização de histograma	0,7277	0,2179	0,7667	0,4507	0,5933	0,7688	0,8112	0,7310	0,4861
Inverter	0,7587	0,1879	0,7902	0,5758	0,6126	0,7308	0,8222	0,7717	0,5975
Solarizar	0,7384	0,1958	0,7622	0,6399	0,5429	0,7010	0,7706	0,7643	0,6227
Aguçamento	0,7822	0,1873	0,8068	0,6552	0,6219	0,7941	0,8234	0,7984	0,6294
Suavização	0,7878	0,1792	0,8061	0,6990	0,6188	0,8097	0,8200	0,7930	0,6789
Ruído gaussiano	0,7455	0,2013	0,7741	0,5948	0,5534	0,7557	0,7973	0,7596	0,5753
Ruído sal e pimenta	0,7764	0,1779	0,7939	0,7064	0,6148	0,7630	0,8057	0,7922	0,6749
Equalização específica 1	0,7340	0,2161	0,7811	0,4307	0,4872	0,7888	0,8064	0,7598	0,4437
Equalização específica 2	0,7723	0,1936	0,7931	0,6654	0,6314	0,8215	0,8233	0,7671	0,6242
Random Cutout	0,7666	0,2011	0,7896	0,6768	0,5426	0,7800	0,7944	0,7708	0,6776
Random MixUp	0,7784	0,1789	0,8007	0,6630	0,6199	0,8097	0,8292	0,7808	0,6246
Todas as técnicas geométricas	0,7637	0,1883	0,7898	0,6508	0,5223	0,7430	0,7871	0,7941	0,6417
Todas as técnicas no espaço de cor	0,7498	0,1970	0,7842	0,5762	0,5331	0,7458	0,7826	0,7757	0,6032
Todas as técnicas por filtro	0,7822	0,1882	0,8018	0,6900	0,6363	0,8071	0,8250	0,7899	0,6388
Todas as técnicas por algoritmo	0,7765	0,1843	0,7994	0,6839	0,5774	0,7719	0,8149	0,7860	0,6637
Experimento 1 - Estratégia 1	0,7926	0,1752	0,8111	0,7019	0,6603	0,8214	0,8335	0,7937	0,6681
Experimento 1 - Estratégia 2	0,7952	0,1674	0,8139	0,7136	0,6319	0,8075	0,8293	0,8032	0,6821
Experimento 1 - Estratégia 3	0,7938	0,1703	0,8143	0,7009	0,6049	0,8008	0,8346	0,8020	0,6685
Experimento 1 - Estratégia 4	0,7839	0,1853	0,8044	0,6581	0,6523	0,8399	0,8432	0,7735	0,6251
Experimento 2 - Estratégia 1	0,7673	0,1917	0,7914	0,6097	0,6698	0,7951	0,8306	0,7735	0,5770
Experimento 2 - Estratégia 2	0,7853	0,1728	0,8013	0,7094	0,6564	0,8195	0,8208	0,7895	0,6601
Experimento 2 - Estratégia 3	0,7768	0,1908	0,7985	0,6697	0,6189	0,8288	0,8152	0,7757	0,6443
Experimento 2 - Estratégia 4	0,7667	0,1937	0,7998	0,5779	0,5801	0,7957	0,8244	0,7830	0,5602
Experimento 3 - Estratégia 1	0,7688	0,1849	0,7905	0,7067	0,5206	0,7581	0,7893	0,7876	0,6773
Experimento 3 - Estratégia 2	0,7744	0,1912	0,8031	0,6338	0,5247	0,7678	0,8099	0,7882	0,6579
Experimento 3 - Estratégia 3	0,7832	0,1814	0,8060	0,6819	0,5991	0,7917	0,8103	0,7962	0,6756
Experimento 3 - Estratégia 4	0,7901	0,1796	0,8088	0,7203	0,5966	0,8114	0,8217	0,7917	0,6929
Experimento 4 - Estratégia 1	0,8015	0,1580	0,8140	0,7561	0,6982	0,8110	0,8365	0,8040	0,7040
Experimento 4 - Estratégia 2	0,8040	0,1604	0,8191	0,7487	0,6708	0,8089	0,8391	0,8081	0,7060
Experimento 4 - Estratégia 3	0,7951	0,1612	0,8075	0,7495	0,6701	0,8250	0,8267	0,7958	0,6929
Experimento 4 - Estratégia 4	0,8032	0,1518	0,8158	0,7673	0,6877	0,7941	0,8254	0,8152	0,7254

O primeiro experimento mostrou ser mais efetivo que combinações manuais de grupos de técnicas na maioria dos casos, já que todas as estratégias tiveram um aumento significativo em relação ao modelo base, entretanto, neste experimento, foi necessário fazer uma avaliação prévia de cada uma das técnicas para selecionar as que tiveram maior média total.

O segundo experimento mostrou ser menos eficaz que os resultados obtidos no primeiro experimento já que apenas duas das estratégias obtiveram média total acima do modelo base. E apenas uma dessas duas teve resultado compatível com os resultados do primeiro experimento.

O terceiro experimento mostrou ter bons resultados, apesar de ser mais complexa que o primeiro, apresentou resultados compatíveis, entretanto apresenta o mesmo problema de ser necessário fazer uma avaliação prévia das técnicas de DA de forma individual.

No quarto experimento apresentou o maior ganho nas médias em relação a todos os outros experimentos, onde as estratégias 2 e 4 obtiveram uma média de 0,8040 e 0,8032, respectivamente. Além de ambos terem tido maiores médias em categorias secundárias. Um ponto a destacar é o fato de a técnica de translação ter tido média abaixo do modelo base, entretanto, quando observado os grupos de técnicas de DA encontrados, esta técnica está constantemente presente em vários destes grupos. Isso corrobora com o fato de mesmo que haja um estudo prévio para selecionar as melhores técnicas de maneira individual, não há como ter certeza se essa busca manual de fato entregará o melhor grupo de técnicas de DA.

## 5 CONCLUSÕES E PROJETOS FUTUROS

### 5.1 Conclusões

O objetivo principal deste trabalho foi propor um abordagem que, dado diferentes técnicas de DA, encontrasse o grupo de técnicas que gere o melhor desempenho de um modelo neural de segmentação semântica orientado ao problema de lesões de peles. Tal objetivo foi alcançado através da utilização de uma busca de hiperparâmetros bayesiana aplicada ao modelo de rede neural U-Net com o conjunto de dados do desafio ISIC 2016. As estratégias adotadas por este trabalho visavam explorar combinações e diversas técnicas de DA para atingir o maior resultado possível para o índice de Jaccard, onde, a estratégia 2 no experimento 4 atingiu 0,8040, posicionando-se, caso estivesse inscrito no desafio ISIC 2016, na sétima posição.

Em relação a aplicabilidade, quando desconhecido o melhor grupo de técnicas, os algoritmos com as estratégias propostas, especialmente o experimento 4 com a estratégia 4, consegue maximizar os resultados sem que haja a necessidade de realizar buscas manuais por técnicas individualmente ou então por grupos de técnicas. Isso prova-se de utilidade quando é desenvolvido um novo modelo e deseja-se inferir o impacto da utilização de técnicas de DA nos resultados.

Com o objetivo de aprimorar os resultados alcançados é sugerido considerar a utilização de: (i) outras técnicas de DA além das utilizadas neste trabalho; (ii) outros conjuntos de dados com maior número de imagens; (iii) outros algoritmos de técnicas de DA como *RandAugment* (CUBUK et al., 2019).



## REFERÊNCIAS

- AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.: s.n.], 2019. Citado na página 68.
- CHATFIELD, K.; SIMONYAN, K.; VEDALDI, A.; ZISSERMAN, A. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531, 2014. Citado na página 24.
- CODELLA, N.; GUTMAN, D.; CELEBI, M.; HELBA, B.; MARCHETTI, M.; DUSZA, S.; KALLOO, A.; LIOPYRIS, K.; MISHRA, N.; KITTLER, H.; HALPERN, A. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi). In: International Skin Imaging Collaboration (ISIC). [S.l.: s.n.], 2017. Citado na página 49.
- CUBUK, E. D.; ZOPH, B.; SHLENS, J.; LE, Q. V. Randaugment: Practical data augmentation with no separate search. CoRR, abs/1909.13719, 2019. Disponível em: <<http://arxiv.org/abs/1909.13719>>. Citado na página 63.
- EFROS, A. A.; FREEMAN, W. T. Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. [S.l.: s.n.], 2001. p. 341–346. Citado na página 28.
- ESTEVA, A.; KUPREL, B.; NOVOA, R. A.; KO, J.; SWETTER, S. M.; BLAU, H. M.; THRUN, S. Dermatologist-level classification of skin cancer with deep neural networks. Nature, Macmillan Publishers Limited, part of Springer Nature. All rights reserved., v. 542, p. 115, jan 2017. Citado na página 14.
- GATYS, L. A.; ECKER, A. S.; BETHGE, M. Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2016. p. 2414–2423. Citado na página 28.
- GHOSH, S.; DAS, N.; DAS, I.; MAULIK, U. Understanding deep learning techniques for image segmentation. ACM Comput. Surv., Association for Computing Machinery, New York, NY, USA, v. 52, n. 4, ago. 2019. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/3329784>>. Citado na página 19.
- GONZALEZ, R. C.; WOODS, R. E. Digital image processing. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN 9780131687288 013168728X 9780135052679 013505267X. Disponível em: <<http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>>. Citado 2 vezes nas páginas 15 e 27.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado 2 vezes nas páginas 15 e 16.
- GUTMAN, D.; CODELLA, N. C.; CELEBI, E.; HELBA, B.; MARCHETTI, M.; MISHRA, N.; HALPERN, A. Skin lesion analysis toward melanoma detection: A challenge

at the international symposium on biomedical imaging (isbi) 2016. In: International Skin Imaging Collaboration (ISIC). [S.l.: s.n.], 2016. Citado 13 vezes nas páginas 20, 22, 23, 25, 26, 28, 29, 30, 31, 46, 47, 49 e 50.

INCA. Estimativa 2020 : incidência de câncer no Brasil. [S.l.]: INCA, 2019. Citado na página 14.

INCA. Atlas de Mortalidade. 2020. Disponível em: <<https://www.inca.gov.br/MortalidadeWeb/pages/Modelo01/consultar.xhtml>>. Citado na página 14.

INCA. Câncer de pele não melanoma. 2021. Disponível em: <<https://www.inca.gov.br/tipos-de-cancer/cancer-de-pele-nao-melanoma>>. Citado na página 14.

JIN, J.; DUNDAR, A.; CULURCIELLO, E. Robust convolutional neural networks under adversarial noise. arXiv preprint arXiv:1511.06306, 2015. Citado na página 27.

JURIO, A.; PAGOLA, M.; GALAR, M.; LOPEZ-MOLINA, C.; PATERNAIN, D. A comparison study of different color spaces in clustering based image segmentation. In: . [S.l.: s.n.], 2010. v. 81, p. 532–541. ISBN 978-3-642-14057-0. Citado na página 23.

LI, Y.; SHEN, L. Skin lesion analysis towards melanoma detection using deep learning network. Sensors, Multidisciplinary Digital Publishing Institute, v. 18, n. 2, p. 556, 2018. Citado na página 16.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015. p. 3431–3440. Citado na página 20.

MIKOŁAJCZYK, A.; GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In: IEEE. 2018 international interdisciplinary PhD workshop (IIPhDW). [S.l.], 2018. p. 117–122. Citado na página 16.

MITCHELL, T. M. et al. Machine learning. McGraw-hill New York, 1997. Citado na página 15.

NOVIKOV, A. A.; LENIS, D.; MAJOR, D.; HLADŮVKA, J.; WIMMER, M.; BÜHLER, K. Fully convolutional architectures for multiclass segmentation in chest radiographs. IEEE transactions on medical imaging, IEEE, v. 37, n. 8, p. 1865–1876, 2018. Citado na página 19.

PEREZ, F.; VASCONCELOS, C.; AVILA, S.; VALLE, E. Data augmentation for skin lesion analysis. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. [S.l.]: Springer, 2018. p. 303–311. Citado na página 16.

PLANCHE, B.; ANDRES, E. Hands-on Computer Vision with TensorFlow 2: Leverage Deep Learning to Create Powerful Image Processing Apps with TensorFlow 2.0 and Keras. Packt Publishing, 2019. ISBN 9781788830645. Disponível em: <<https://books.google.com.br/books?id=XkL5xQEACAAJ>>. Citado na página 19.

REZZE, G. G.; SÁ, B. C. S. d.; NEVES, R. I. Dermatoscopia: o método de análise de padrões. Anais brasileiros de dermatologia, SciELO Brasil, v. 81, p. 261–268, 2006. Citado na página 15.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. Citado 2 vezes nas páginas 20 e 21.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. Journal of Big Data, Springer, v. 6, n. 1, p. 1–48, 2019. Citado 4 vezes nas páginas 16, 22, 24 e 30.

WHO. The top 10 causes of death. World Health Organization, 2020. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>>. Citado na página 14.

ZHANG, H.; CISSE, M.; DAUPHIN, Y. N.; LOPEZ-PAZ, D. mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412, 2017. Citado na página 30.

ZHONG, Z.; ZHENG, L.; KANG, G.; LI, S.; YANG, Y. Random erasing data augmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.: s.n.], 2020. v. 34, n. 07, p. 13001–13008. Citado na página 29.

# Anexos

## ANEXO A – OPTUNA

Optuna é um *framework* de otimização de hiperparâmetros desenvolvido para a linguagem de programação Python. Este *framework* é compatível com quase todos os pacotes de ML e com uma facilidade para a implementação (AKIBA et al., 2019).

Seu funcionamento é dividido em três conceitos básicos:

- *Study*, que é a coleção de tentativas do processo da otimização;
- *Objective*, que é o objetivo da otimização do *study*;
- *Trial*, que são as tentativas de encontrar um ou mais hiperparâmetros dentro de um espaço de busca para atingir o objetivo.

Tomando em conta, os três conceitos citados, o processo de busca Bayesiana de hiperparâmetros pode ser implementado em três passos:

- **Implementação da função *objective*.** É criada uma função em cujo corpo são indicados: (i) o(s) hiperparâmetro(s) que devem ser buscados, assim como o espaço de busca para cada um deles; (ii) a função a ser avaliada no processo de otimização, como por exemplo, uma função que efetua o treinamento e validação de um modelo neural retornando o valor de uma métrica de desempenho calculada sobre o conjunto de validação. Na lista A.1 é mostrado um código de referencia de uma função *objective*.

Code Listing A.1 – Exemplo de função *objective*

```
def objective(trial = optuna.Trial, *args, **kwargs):
    # Nesta etapa seleciona a(s) sugestao(oes) de hiperparametro(s)
    # para realizar a busca
    # A funcao realiza alguma coisa, como por exemplo treinar um
    # modelo ou realizar uma
    # regressao com o(s)
    # hiperparametro(s)
    valor_da_avaliacao = funcao_que_avalua(*args, **kwargs)
    return valor_da_avaliacao
```

- **Criação e Execução do *Study*.** É criado um objeto *Study* relacionado à função *objective* implementada e é executada a otimização dos hiperparâmetros chamando o método *optimize* de objeto *Study*, aqui deve ser indicado o número de *trials* a ser

feitos. O código para criação do *study* e a execução do processo de otimização são mostrados na Lista A.2.

Code Listing A.2 – Exemplo de código com *study*

```
study = optuna.create_study(direction='maximize')
func = lambda trial: objective(trial, *args, **kwargs)
study.optimize(func, n_trials = 100)
```

- **Visualização dos resultados.** São exibidos os resultados da otimização realizada. A Lista A.3 mostra o código necessário para visualizar os valores atribuídos a cada hiperparâmetro considerado no processo de otimização.

Code Listing A.3 – Exemplo de código para visualizar os resultados do otimização

```
# Chamando diretamente o atributo
print(f'Melhores hiperparametros encontrado em {len(study.trials)}
      trials: {study.best_params}')

# Criando um banco de dados com todos os atributos
df = study.trials_dataframe(attrs=('number', 'value', 'params', '
                                state'))

df.head()
```

## ANEXO B – IMPLEMENTAÇÃO DA U-NET

A arquitetura foi totalmente implementada através dos blocos e funções disponibilizadas em TensorFlow. A Tabela 23 mostra o bloco de entrada da imagem com uma função Lambda e os quatro blocos do caminho de contração de dados. Essa tabela também descreve as saídas assim como o número de parâmetros de cada uma das camadas.

Tabela 23 – Camadas do caminho de contração de dados.

Camada	Formato de saída	Número de parâmetros
input_1 (InputLayer)	(None, 512, 512, 3)	0
lambda (Lambda)	(None, 512, 512, 3)	0
conv2d (Conv2D)	(None, 512, 512, 16)	448
dropout (Dropout)	(None, 512, 512, 16)	0
conv2d_1 (Conv2D)	(None, 512, 512, 16)	2320
max_pooling2d (MaxPooling2D)	(None, 256, 256, 16)	0
conv2d_2 (Conv2D)	(None, 256, 256, 32)	4640
dropout_1 (Dropout)	(None, 256, 256, 32)	0
conv2d_3 (Conv2D)	(None, 256, 256, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 128, 128, 32)	0
conv2d_4 (Conv2D)	(None, 128, 128, 64)	18496
dropout_2 (Dropout)	(None, 128, 128, 64)	0
conv2d_5 (Conv2D)	(None, 128, 128, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 64, 64, 64)	0
conv2d_6 (Conv2D)	(None, 64, 64, 128)	73856
dropout_3 (Dropout)	(None, 64, 64, 128)	0
conv2d_7 (Conv2D)	(None, 64, 64, 128)	147584
max_pooling2d_3 (MaxPooling2D)	(None, 32, 32, 128)	0

A Tabela 24 mostra o bloco que faz a transição do caminho de contração para o caminho de expansão simétrica dos dados. Essa tabela também descreve os quatro operadores de concatenação que fazem a ligação entre os caminhos de contração e expansão simétrica.

A Tabela 25 mostra os quatro blocos do caminho de expansão simétrica dos dados.

Tabela 24 – Camadas de ligação entre os caminhos e camadas de concatenação.

Camada	Formato de saída	Número de parâmetros
conv2d_8 (Conv2D)	(None, 32, 32, 256)	295168
dropout_4 (Dropout)	(None, 32, 32, 256)	0
conv2d_9 (Conv2D)	(None, 32, 32, 256)	590080
conv2d_transpose (Conv2DTranspose)	(None, 64, 64, 128)	131200
concatenate (Concatenate)	(None, 64, 64, 256)	0
concatenate_1 (Concatenate)	(None, 128, 128, 12)	0
concatenate_2 (Concatenate)	(None, 256, 256, 64)	0
concatenate_3 (Concatenate)	(None, 512, 512, 32)	0

Tabela 25 – Camadas do expansão simétrica dos dados.

Camada	Formato de saída	Número de parâmetros
conv2d_10 (Conv2D)	(None, 64, 64, 128)	295040
dropout_5 (Dropout)	(None, 64, 64, 128)	0
conv2d_11 (Conv2D)	(None, 64, 64, 128)	147584
conv2d_transpose_1 (Conv2DTranspose)	(None, 128, 128, 64)	32832
conv2d_12 (Conv2D)	(None, 128, 128, 64)	73792
dropout_6 (Dropout)	(None, 128, 128, 64)	0
conv2d_13 (Conv2D)	(None, 128, 128, 64)	36928
conv2d_transpose_2 (Conv2DTranspose)	(None, 256, 256, 32)	8224
conv2d_14 (Conv2D)	(None, 256, 256, 32)	18464
dropout_7 (Dropout)	(None, 256, 256, 32)	0
conv2d_15 (Conv2D)	(None, 256, 256, 32)	9248
conv2d_transpose_3 (Conv2DTranspose)	(None, 512, 512, 16)	2064
conv2d_16 (Conv2D)	(None, 512, 512, 16)	4624
dropout_8 (Dropout)	(None, 512, 512, 16)	0
conv2d_17 (Conv2D)	(None, 512, 512, 16)	2320
conv2d_18 (Conv2D)	(None, 512, 512, 1)	17



## ANEXO C – TABELA DE HIPERPARÂMETROS CONSTANTES

A Tabela 26 descreve os hiperparâmetros constantes utilizados em cada uma das funções de transformação, assim como os valores internos e descrição da operação de transformação utilizada.

Tabela 26 – Hiperparâmetros constantes utilizados nas funções de transformação

Função de transformação	Descrição e valores utilizados
Flipping horizontal	Inversão utilizando o eixo vertical que cruza o centro da imagem
Flipping vertical	Inversão utilizando o eixo horizontal que cruza o centro da imagem
Balanco de branco	Imagem convertida para o espaço de cor CIELAB, o balanceamento é feito e é convertido para o espaço RGB
Equalização de histograma	Imagem convertida para o espaço de cor HSL, a equalização é feita na camada de intensidade e é convertido para o espaço RGB
Inverter	O limiar utilizado é 0
Solarizar	O limiar utilizado é 127
Aguçamento	A máscara utilizada é do tamanho 3x3 com a utilização de preenchimento com zeros igual a 2. A máscara é um laplaciano
Suavização	A máscara utilizada é do tamanho 3x3 com a utilização de preenchimento com zeros igual a 2. A máscara é gerada através de uma distribuição gaussiana com sigma igual a 0
Ruído gaussiano	Média igual a 0. Variância igual a 0,01
Ruído sal e pimenta	Probabilidade de ocorrência igual a 0,05. Probabilidade de ser sal ou pimenta igual a 0,5
Random cutout	Região de corte do tamanho de 64x64. A região de corte é substituída por pixels de intensidade igual a 127 nos três canais