

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Nome completo do autor

Título do Trabalho de Conclusão de Curso

Uberlândia, Brasil

Ano de defesa

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Nome completo do autor

Título do Trabalho de Conclusão de Curso

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Nome completo do orientador

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

Ano de defesa

Nome completo do autor

Título do Trabalho de Conclusão de Curso

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Ciência da Computação.

Trabalho aprovado. Uberlândia, Brasil, 24 de novembro de 2012:

Nome completo do orientador
Orientador

Professor

Professor

Uberlândia, Brasil
Ano de defesa

Insere o texto de dedicatória ...

Agradecimentos

Inserir o texto de agradecimentos ...

Insere alguma citação que ache conveniente caso queira ...

Resumo

Segundo a [ABNT \(2003, 3.1-3.2\)](#), o resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (...) As palavras-chave devem figurar logo abaixo do resumo, antecidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto.

Palavras-chave: Até, cinco, palavras-chave, separadas, por, vírgulas.

Lista de ilustrações

Figura 1 – Visão geral das principais etapas de classificação de lesões da cavidade bucal.	16
Figura 2 – Exemplo de tecido histológico da cavidade bucal: (a) saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa.	17
Figura 3 – Exemplo das imagens originais de tecido histopatológico da cavidade oral: (a) saudável, (b) carcinoma escamoso da cavidade oral.	18
Figura 4 – Exemplo das imagens de tecido histopatológico da cavidade oral: (a) saudável, (b) carcinoma escamoso da cavidade oral, redimensionadas a partir do centro das imagens da Figura 3.	18
Figura 5 – As imagens (a) e (b) são amostras de tecidos com displasia leve, (c) imagem construída por meio da combinação de (a) e (b).	19
Figura 6 – Exemplo das etapas da criação de imagens sintéticas utilizando <i>Mixup</i> : (a) Imagens obtidas aleatoriamente, (b) Vetorização das imagens, (c) Matrizes proporcionais de x_i e x_j , (d) Matriz \tilde{x} resultante da soma das matrizes proporcionais e (e) Imagem equivalente a \tilde{x}	20
Figura 7 – Exemplo de transformação dos dados de treinamento para cada época. Fonte (Matlab documentation, 2022)	21
Figura 8 – Arquitetura da rede neural <i>Resnet50</i> . M1, M2, M3, M4 representam os módulos compostos por blocos residuais. Adaptado de (SILVA, 2019).	22
Figura 9 – Blocos residuais presentes na arquitetura <i>ResNet50</i> : (a) bloco residual sem convolução 1x1 e (b) bloco residual com convolução 1x1.	23
Figura 10 – Arquitetura da rede neural <i>DenseNet201</i> . B1, B2, B3, B4 representam os blocos densos. T1, T2, T3 representam as camadas de transição.	24
Figura 11 – Representação de um bloco denso, composto por 5 blocos de convolução, onde cada bloco de convolução tem como entrada a saída dos blocos anteriores. Fonte (HUANG et al., 2017)	24
Figura 12 – Exemplo de uma matriz de confusão para análise das imagens histológicas	26
Figura 13 – Imagens de carcinoma oral de células escamosas. As imagens das colunas (a) e (b) são imagens escolhidas aleatoriamente, a coluna (c) representa a combinação resultante de (a) e (b).	28
Figura 14 – Imagens de displasia oral. As imagens das colunas (a) e (b) são imagens escolhidas aleatoriamente, a coluna (c) representa a combinação resultante de (a) e (b).	29

Figura 15 – Acurácia média obtida para o conjunto de dados de displasia da cavidade oral, onde (a) utilizando somente a técnica <i>Mixup</i> , (b) utilizando a combinação das técnicas <i>Mixup</i> e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.	31
Figura 16 – Acurácia média obtida para o conjunto de dados de carcinoma escamoso da cavidade oral, onde (a) utilizando somente a técnica <i>Mixup</i> , (b) utilizando a combinação das técnicas <i>Mixup</i> e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.	32
Figura 17 – F1-Score médio obtido para o conjunto de dados de carcinoma escamoso da cavidade oral, onde (a) utilizando somente a técnica <i>Mixup</i> , (b) utilizando a combinação das técnicas <i>Mixup</i> e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.	32
Figura 18 – Taxa de perda para as diferentes estratégias. Conjunto de dados de displasia bucal, arquitetura ResNet50, onde (a) utilizando a combinação das técnicas <i>Mixup</i> e aumento de dados geométricos, (b) utilizando somente a técnica <i>Mixup</i> , (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.	34

Lista de tabelas

Tabela 1 – Exemplo de tabela.	15
Tabela 2 – Acurácia média obtida para o conjunto de dados de displasia da cavidade oral.	30
Tabela 3 – Acurácia média obtida para o conjunto de dados de carcinoma escamoso da cavidade oral.	30
Tabela 4 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral. Arquitetura <i>ResNet50</i>	33
Tabela 5 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral. Arquitetura <i>DenseNet201</i>	33

Lista de abreviaturas e siglas

AHP	Analytical Hierarchy Process
IA	Inteligência Artificial
UML	Unified Modeling Language

Lista de símbolos

\in	Pertence
F	Valor falso do tipo de dado primitivo Booleano
Γ	Letra grega Gama
Λ	Lambda
V	Valor verdadeiro do tipo de dado primitivo Booleano
ζ	Letra grega minúscula zeta

Sumário

1	INTRODUÇÃO	14
1.1	Objetivos	14
1.2	Método	14
1.3	Resultados	14
1.4	Organização do Trabalho	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Conceitos Básicos	15
2.1.1	Exemplos	15
2.2	Trabalhos Relacionados	15
3	METODOLOGIA	16
3.1	Visão Geral	16
3.2	Banco de Imagens	16
3.2.1	Lesões displásicas da cavidade oral	17
3.2.2	Carcinoma escamoso da cavidade oral	17
3.3	Aumento dados utilizando a técnica <i>Mixup</i>	19
3.4	Aumento de dados por transformações geométricas	20
3.5	Redes Neurais Convolucionais	20
3.5.1	ResNet50	22
3.5.2	DenseNet201	23
3.6	Treinamento das Redes	25
3.7	Métricas de Avaliação	25
4	RESULTADOS	28
4.1	Avaliação das imagens geradas utilizando a técnica <i>Mixup</i>	28
4.2	Avaliação comparativa das arquitetura utilizadas	29
4.3	Avaliação comparativa das estratégias adotadas	31
5	CONCLUSÃO	35
	REFERÊNCIAS	36

APÊNDICES 37

APÊNDICE A – COISAS QUE FIZ E QUE ACHEI INTERESSANTE MAS NÃO TANTO PARA ENTRAR NO CORPO DO TEXTO	38
---	-----------

ANEXOS 39

ANEXO A – COISAS QUE EU NÃO FIZ MAS QUE ACHEI INTE- RESSANTE O SUFICIENTE PARA COLOCAR AQUI	40
--	-----------

1 Introdução

Contextualização, problema, hipótese, ... ([TECHAMERICA, 2011](#)).

1.1 Objetivos

Descrever o objetivo geral e os objetivos específicos.

1.2 Método

Descrever o método utilizado.

1.3 Resultados

Descrever os resultados de seu trabalho.

1.4 Organização do Trabalho

Descrever a organização do trabalho.

2 Fundamentação Teórica

Neste capítulo é apresentado os conceitos básicos necessários para compreensão deste trabalho bem como os trabalhos relacionados.

2.1 Conceitos Básicos

Esta seção apresenta os conceitos teóricos relacionados ...

2.1.1 Exemplos

Exemplo tabela (Figura 1).

Tabela 1 – Exemplo de tabela.

Algoritmo	Resultado	Tempo
Alg 1	10,5	0,2
Alg 2
...

Exemplo de equação (Equação (2.1)).

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (2.1)$$

Além disso, quando formos escrever números ou símbolos matemáticos ao longo do texto, usamos a seguinte notação: $x = 2$, $y^2 = \log_2(x)$, etc.

Exemplo de algoritmo, usando o pacote `algorithm2e` (Algoritmo 1)

Algoritmo 1: Exemplo de algoritmo.

Entrada: x e y inteiros.

Saída: z inteiro.

se $x > y$ **então**

$z = x - y$;

fim

2.2 Trabalhos Relacionados

3 Metodologia

3.1 Visão Geral

Conjunto de dados grandes e diversificados frequentemente aumentam a precisão dos algoritmos em inteligência artificial. No entanto, a obtenção de dados reais pode ser algo demorado, limitado e muito caro em diversas aplicações como na área médica. Uma estratégia é o uso de dados aumentados, o que geralmente não é considerado informações sintéticas. O presente trabalho traz uma investigação de algoritmos de classificação baseados em redes neurais convolucionais sobre abordagens de aumento de dados. O sistema proposto aborda a geração de dados utilizando a técnica *Mixup*, que consiste na combinação de duas ou mais imagens para criação de uma nova representação. Também são investigados abordagens tradicionais de aumento de dados baseadas em representações geométricas. Na Figura 1 é apresentado um diagrama das principais etapas do método proposto e investigadas neste trabalho.

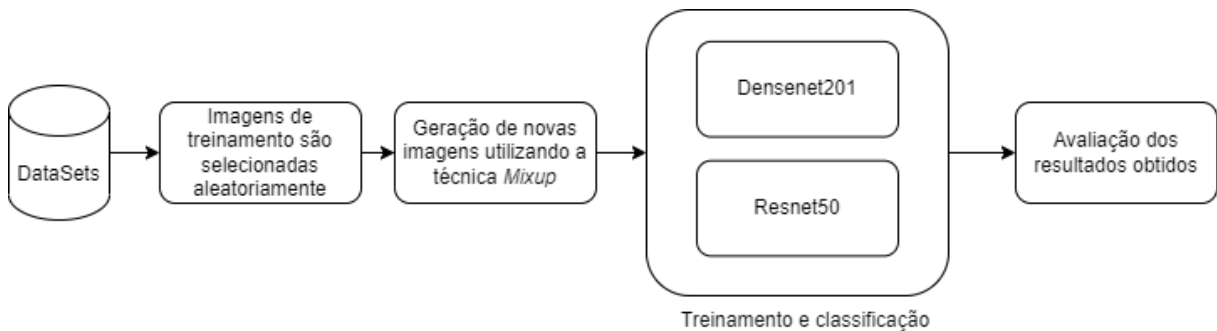


Figura 1 – Visão geral das principais etapas de classificação de lesões da cavidade bucal.

O trabalho foi desenvolvido utilizando a linguagem de programação MatLab® e todo o processamento feito em um computador com processador Intel™ Core® i7 7500U com 8GB de memória RAM.

3.2 Banco de Imagens

Neste estudo foram empregados dois *datasets* para avaliar e gerar imagens, sendo um contendo lesões displásicas e outro contendo imagens de carcinoma escamoso da cavidade oral.

3.2.1 Lesões displásicas da cavidade oral

Esse conjunto é composto de amostras de lesões induzidas na língua de 30 camundongos com o uso do carcinógeno 4-nitroquinolina 1-óxido (4NQO). O experimento *in vivo*, realizado no período de 2009 a 2010, foi aprovado pelo Comitê de Ética na Utilização de Animais sob número de protocolo 038/09, sendo utilizado pelo pesquisador Adriano B. Silva (SILVA, 2019) no estudo proposto em seu trabalho de Mestrado.

Após desenvolverem a lesão, os camundongos foram sacrificados, suas línguas removidas e colocadas em uma solução de formalina tamponada a 4%. Após 24 horas, foram processadas e inseridas em parafina para confecção dos blocos. Uma vez executado esses procedimentos, as línguas foram recortadas e coradas por H&E para estudo histopatológico.

Usando um microscópio óptico Leica DM500, em magnificação de 400x, foram digitalizadas um total de 43 lâminas, com resolução de 2048×1536 pixels, utilizando o modelo de cores RGB e armazenadas em formato TIFF. Com o auxílio de um patologista, as imagens obtidas foram categorizadas em saudáveis, displasia leve, moderada e severa e então recortadas em regiões de interesse (do inglês, *region of interest* - ROI) com tamanho de 450×250 pixels, sendo 74 ROIs para cada classe de tecido, totalizando 296 ROIs.

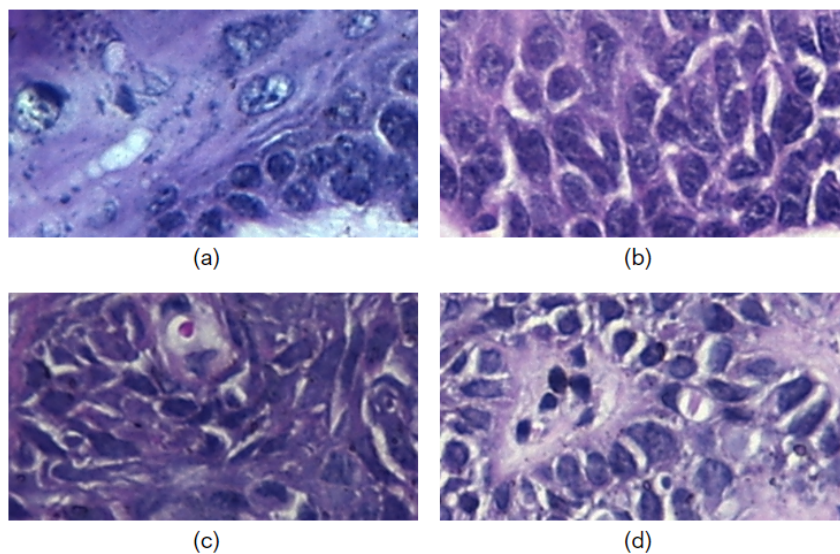


Figura 2 – Exemplo de tecido histológico da cavidade bucal: (a) saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa.

3.2.2 Carcinoma escamoso da cavidade oral

As imagens que integram o segundo conjunto foram obtidas a partir de um banco de dados de imagens histopatológicas para análise de câncer oral (RAHMAN, 2019).

O repositório é composto de 1.224 imagens, divididas em dois conjuntos com diferentes ampliações. O primeiro, com magnificação de 100x, contém 89 imagens histopatológicas com o epitélio normal da cavidade oral e 439 imagens de carcinoma escamoso da cavidade oral (do inglês, *oral squamous cell carcinoma* - OSCC). O segundo, com magnificação de 400x, contém 201 imagens histopatológicas com o epitélio normal da cavidade oral e 495 imagens de carcinoma escamoso da cavidade oral.

As imagens foram capturadas usando um microscópio Leica DM 750 com câmera modelo ICC50 HD, com resolução de 2048×1536 pixels. Ao todo, 230 pacientes foram recomendados para teste de Biópsia Oral entre outubro de 2016 e novembro de 2017. As lâminas de biópsia foram coletadas por duas instituições de serviços de saúde de renome, Ayursundra Healthcare Pvt. Ltd e Dr. B. Borooah Cancer Institute, sendo preparadas e catalogadas por especialistas médicos como tecidos saudáveis e tecidos com carcinoma.

Neste estudo, foram utilizadas as imagens com magnificação de 400x (ver Figura 3), mas devido às limitações do hardware utilizado, foi necessário realizar um pré-processamento das imagens. As imagens foram redimensionadas em 50% e ocorreu um recorte a partir do centro da imagem de tamanho 250×450 pixels (ver Figura 4).

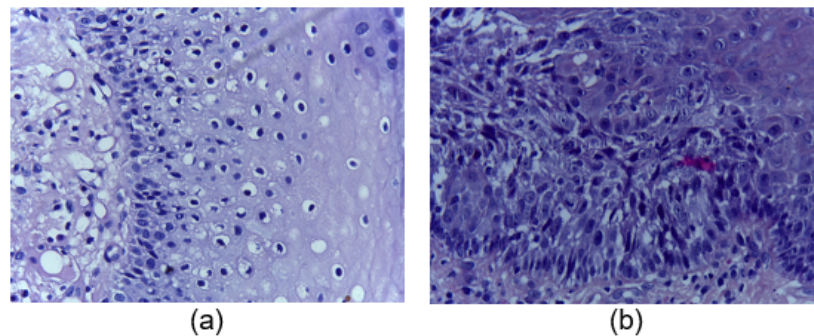


Figura 3 – Exemplo das imagens originais de tecido histopatológico da cavidade oral: (a) saudável, (b) carcinoma escamoso da cavidade oral.

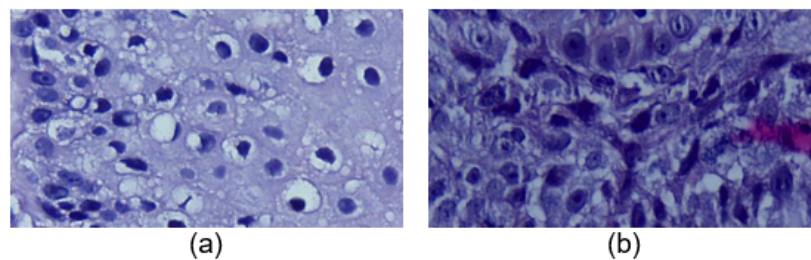


Figura 4 – Exemplo das imagens de tecido histopatológico da cavidade oral: (a) saudável, (b) carcinoma escamoso da cavidade oral, redimensionadas a partir do centro das imagens da Figura 3.

3.3 Aumento dados utilizando a técnica *Mixup*

A técnica *mixup* pode ser considerada uma abordagem de aumento de dados que consiste em misturar duas ou mais imagens com o objetivo de construir uma nova imagem representativa. De acordo com os autores em (ZHANG et al., 2017), o *Mixup* pode ser definido por:

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j, & \text{onde } x_i \text{ e } x_j \text{ representam as imagens de entrada} \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j, & \text{onde } y_i \text{ e } y_j \text{ representam os rótulos das imagens}\end{aligned}$$

(x_i, y_i) e (x_j, y_j) são dois exemplos extraídos aleatoriamente do conjunto de dados de treinamento e $\lambda \in [0,1]$.

O primeiro passo dessa abordagem consiste em definir de forma aleatória um valor para o parâmetro λ entre intervalos de 0 até 1. Esse termo representa a proporção das imagens de entrada que irão compor a nova amostra construída de forma sintética. Para aplicação, duas imagens foram selecionadas de forma aleatória. A primeira imagem é vetorizada e cada elemento da matriz resultante é multiplicado pelo valor de λ . A segunda imagem também é vetorizada e cada elemento da matriz resultante é multiplicado pelo valor de $1-\lambda$. Com isso, se uma imagem, denominada imagem 1, é multiplicada por um valor λ igual a 0,3, a segunda, denominada imagem 2, é obrigatoriamente multiplicada por um valor λ igual a 0,7. As duas matrizes foram somadas, obtendo a matriz equivalente da nova imagem. Isso representa uma nova imagem sintética composta de 30% da imagem 1 e 70% da imagem 2. A Figura 6 apresenta o processo de criação de imagens utilizando *Mixup* entre lesões displásicas leves.

Neste trabalho, o conjunto de dados utilizado foi dividido em subconjuntos de acordo com a rotulação atribuída previamente por um especialista. O processo de *Mixup* foi aplicado em cada um dos subconjuntos isoladamente, ou seja, apenas nas imagens de cada classe investigada (ver Figura 5).

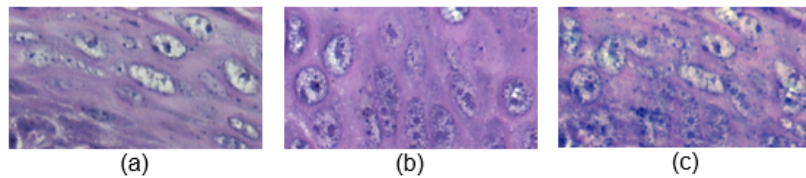


Figura 5 – As imagens (a) e (b) são amostras de tecidos com displasia leve, (c) imagem construída por meio da combinação de (a) e (b).

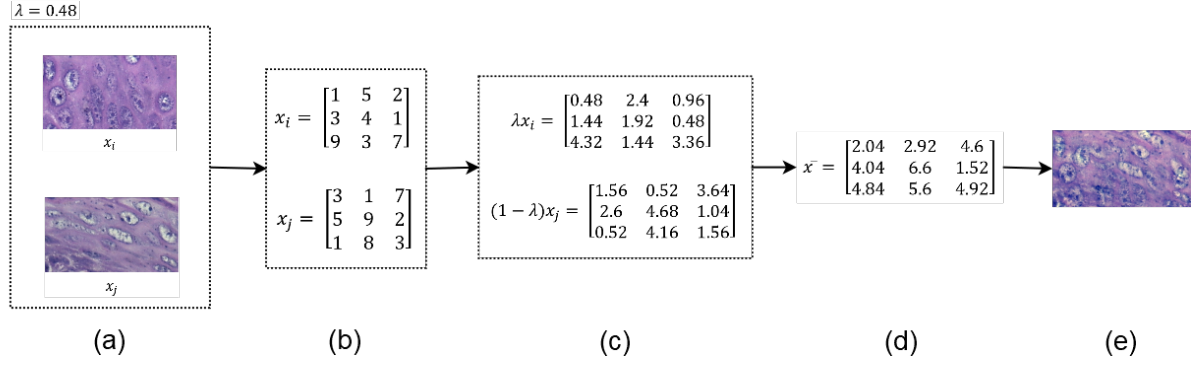


Figura 6 – Exemplo das etapas da criação de imagens sintéticas utilizando *Mixup*: (a) Imagens obtidas aleatoriamente, (b) Vetorização das imagens, (c) Matrizes proporcionais de x_i e x_j , (d) Matriz \tilde{x} resultante da soma das matrizes proporcionais e (e) Imagem equivalente a \tilde{x} .

3.4 Aumento de dados por transformações geométricas

O aumento de dados (do inglês, *data augmentation* - DA) consiste no processo de aumentar a diversidade e/ou quantidade dos dados de um *dataset* aplicando diversas transformações às imagens existentes, tais como rotação, reflexão, redimensionamento, recorte, entre outras coisas. Neste trabalho, as seguintes transformações foram empregadas: rotação aleatória, reflexão aleatória, cisalhamento horizontal e redimensionamento aleatório.

No caso do programa *Matlab*, a função denominada *augmentedImageDatastore* foi empregada para realizar essas operações. Na Figura 7 é possível observar as etapas do processo de construção das imagens transformadas pelo emprego da função *augmentedImageDatastore*. Para essa etapa, primeiramente, o conjunto de treinamento foi dividido em subconjuntos denominados mini-lotes (do inglês, *mini-batch*). O tamanho do mini-lote representa a quantidade de dados processados em cada época. Isso é feito para que o processo de treinamento seja otimizado, uma vez que utilizar todo o conjunto de treinamento em cada uma das épocas tornaria o processo muito lento, demandando muita memória e espaço de armazenamento para processar os dados. Então, em cada época ocorreu o processo de iterações, em que para cada iteração de treinamento, uma combinação aleatória de transformações foi aplicada às imagens do mini-lote de treinamento possibilitando executar cada época com um conjunto de dados ligeiramente diferente.

3.5 Redes Neurais Convolucionais

Uma rede neural convolucional (do inglês, *convolutional neural network* - CNN) é uma arquitetura de rede utilizada no reconhecimento de imagens, semelhante às redes neurais artificiais, sendo sua principal diferença possuir uma ou mais camadas convolu-

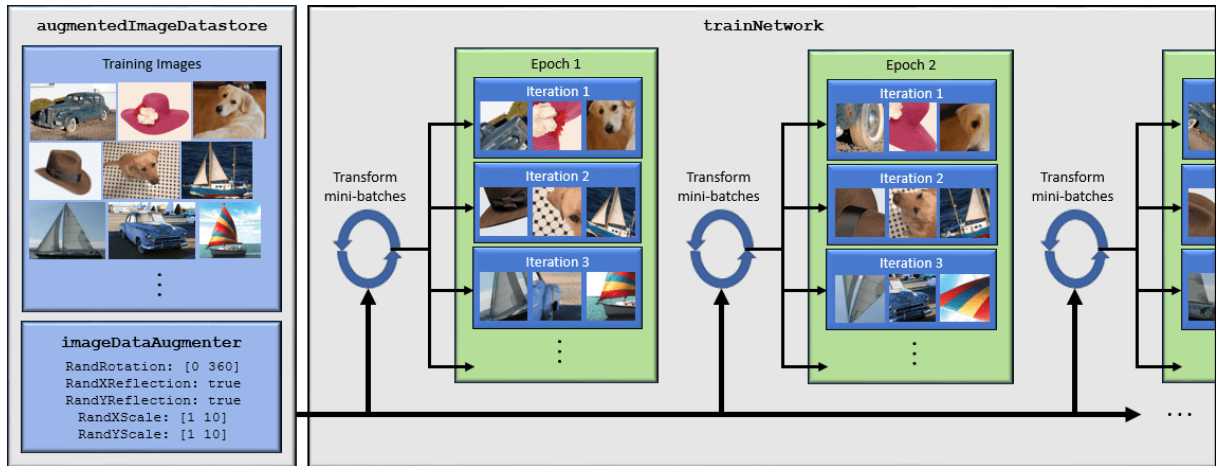


Figura 7 – Exemplo de transformação dos dados de treinamento para cada época. Fonte (Matlab documentation, 2022)

cionais. Essas camadas convolucionais são responsáveis por extrair as principais características das imagens, por meio da aplicação de filtros ou *kernels*, reduzindo o tamanho dos dados sem que se perca os recursos necessários para uma classificação, o que a torna escalável para conjuntos de dados massivos. Os autores em Simonyan e Zisserman (2014) evidenciaram que a profundidade da rede representa um fator crucial para melhorar a precisão da classificação, no entanto, conforme mostrado pelos autores em He et al. (2016), um maior número de camadas pode resultar em uma saturação da precisão e, conseqüentemente, uma rápida degradação. Isso ocorre devido a um problema comum no aprendizado profundo: a dissipação/explosão do gradiente (do inglês, *vanishing/exploding gradient*) detalhado em Bengio, Simard e Frasconi (1994). Destaca-se que a dissipação do gradiente ocorre quando os gradientes retropropagados assumem valores tão baixos que os ajustes dos pesos são praticamente irrelevantes nas camadas iniciais da rede. Isso faz com que o treinamento se torne muito custoso, sendo necessário uma quantidade grande de iterações para um ajuste pouco significativo. Já a explosão do gradiente acontece quando os gradientes retropropagados assumem valores tão altos que os pesos sofrem grandes atualizações, tornando o modelo instável e incapaz de aprender.

Neste trabalho, foram utilizados os modelos de CNNs *Resnet50* (He et al. (2016)), com 50 camadas de profundidade e *Densenet201* (Huang et al. (2017)), com 201 camadas de profundidade. Essas redes foram definidas nesse trabalho devido ao fato que essas arquiteturas permitem que um maior número de camadas sejam adicionadas às redes neurais convolucionais ao mesmo tempo que tratam o problema de instabilidade do gradiente.

3.5.1 ResNet50

O modelo *ResNet50* empregado nesse trabalho é constituído por uma camada convolucional de entrada, 4 módulos denominados M1, M2, M3, M4 e uma camada de classificação (ver Figura 8).

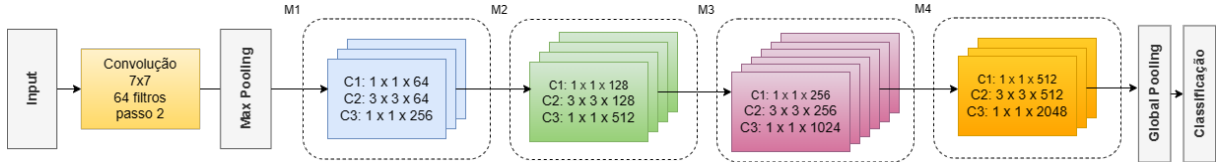


Figura 8 – Arquitetura da rede neural *Resnet50*. M1, M2, M3, M4 representam os módulos compostos por blocos residuais. Adaptado de (SILVA, 2019).

A camada convolucional de entrada possui 64 filtros de tamanho 7×7 *pixels* que atuam sobre a imagem original utilizando uma janela deslizante, também conhecida como passada (do inglês, *stride*), com tamanho de deslocamento de 2 *pixels*. Ela é seguida por uma camada de *max pooling* de tamanho 3×3 *pixels* com uma passada de 2 *pixels*.

O módulo M1 é composto por três blocos residuais. Cada bloco residual possui três camadas convolucionais, sendo a primeira camada formada por 64 filtros de tamanho 1×1 *pixel*, a segunda formada por 64 filtros de tamanho 3×3 *pixels* e a terceira com 256 filtros de tamanho 1×1 *pixel*.

O módulo M2 possui quatro blocos residuais, que por sua vez possuem três camadas convolucionais cada. A primeira camada contém 128 filtros convolucionais de tamanho 1×1 *pixel*. A segunda camada possui 128 filtros de tamanho 3×3 *pixels* e a terceira compreende 512 filtros de tamanho 1×1 *pixels*.

O módulo M3 contém seis blocos residuais com três camadas convolucionais cada. A primeira camada convolucional possui 256 filtros com tamanho de 1×1 *pixel*. A segunda 256 filtros com tamanho de 3×3 *pixels* e a terceira 1024 filtros de tamanho 1×1 *pixel*.

O módulo M4 é formado por três blocos residuais, cada um contendo três camadas convolucionais, sendo a primeira camada composta por 512 filtros de tamanho 1×1 *pixel*, a segunda contendo 512 filtros com tamanho de 3×3 *pixels* e a terceira com 2048 filtros de tamanho 1×1 *pixel*.

A saída de cada bloco residual é somada elemento a elemento e canal a canal à entrada do próprio bloco e então transmitidas ao bloco seguinte. Para que essa soma aconteça é necessário que entrada e saída tenham tamanhos e quantidade de canais compatíveis, entretanto ao longo da rede convolucional, o tamanho da imagem de entrada é reduzido ao passo que o número de canais aumenta. Por esse motivo, foi realizado um ajuste no primeiro bloco residual de cada módulo, introduzindo uma convolução 1×1 no

caminho da conexão de salto, reduzindo a amostra a um tamanho apropriado e aumentando a quantidade de canais, possibilitando a soma da entrada x com a saída $f(x)$ (ver Figura 9).

Por fim, a saída do módulo M4 é submetida a um pooling médio global que irá transformar os dados em um vetor que será empregado à camada responsável pela classificação.

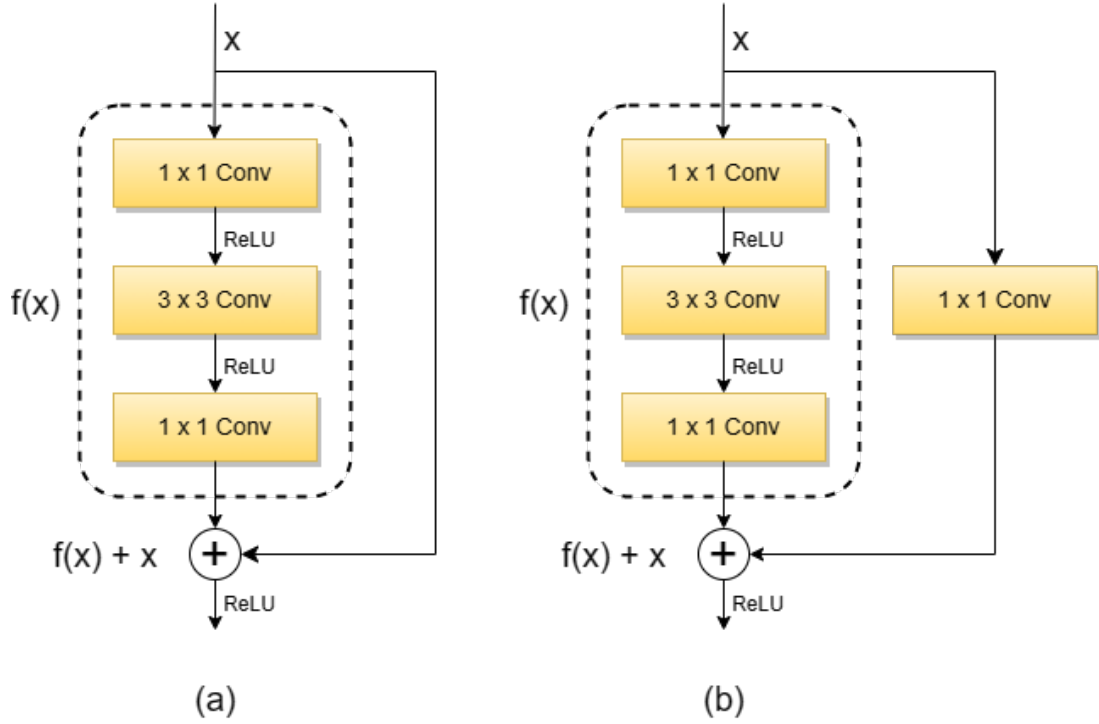


Figura 9 – Blocos residuais presentes na arquitetura *ResNet50*: (a) bloco residual sem convolução 1x1 e (b) bloco residual com convolução 1x1.

3.5.2 DenseNet201

A arquitetura *DenseNet201* (HUANG et al., 2017) utilizada neste trabalho é composta por uma camada convolucional de entrada, quatro blocos densos denominados B1, B2, B3, B4, intercalados por camadas de transição denominadas T1, T2, T3 e uma camada de classificação (ver Figura 10). Semelhante a *ResNet50*, a camada de entrada possui 64 filtros convolucionais de tamanho 7×7 pixels, com uma passada (do inglês, *stride*) de 2 pixels seguida de uma camada de *max pooling* de tamanho 3×3 pixels e uma passada de 2 pixels.

O bloco denso é formado por diversos blocos de convolução, sendo B1 formado por seis blocos de convolução, B2 por 12 blocos de convolução, B3 por quarenta e oito blocos de convolução e B4 por trinta e dois blocos de convolução. Cada bloco de convolução é constituído de duas camadas convolucionais, sendo a primeira camada com 128 filtros

convolucionais de tamanho 1×1 *pixel* e a segunda com 32 filtros de tamanho 3×3 *pixels*. As entradas de cada bloco convolucional são concatenadas às saídas e então transmitidas aos blocos seguintes conforme demonstrado na Figura 11.

Uma vez que o bloco denso irá aumentar a quantidade de canais, para evitar que o modelo se torne excessivamente complexo, entre cada bloco é inserido uma camada de transição constituída de uma camada de convolução com 128 filtros convolucionais de tamanho 1×1 *pixel*, seguida de uma camada de pooling médio com uma passada de 2 *pixels*, fazendo com que o número de canais seja reduzido. Por fim, uma camada de pooling médio global e uma camada convolucional foram conectadas ao bloco B4 produzindo a saída. Assim como no modelo *ResNet50*, a função de ativação empregada foi a *ReLU*.

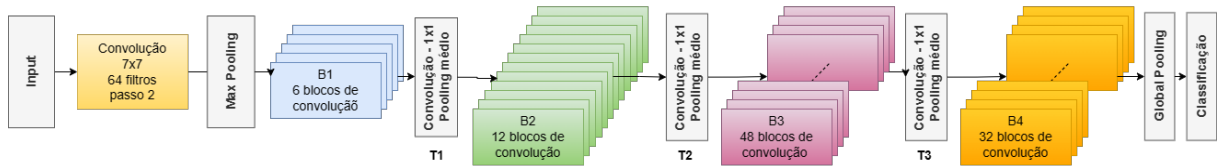


Figura 10 – Arquitetura da rede neural *DenseNet201*. B1, B2, B3, B4 representam os blocos densos. T1, T2, T3 representam as camadas de transição.

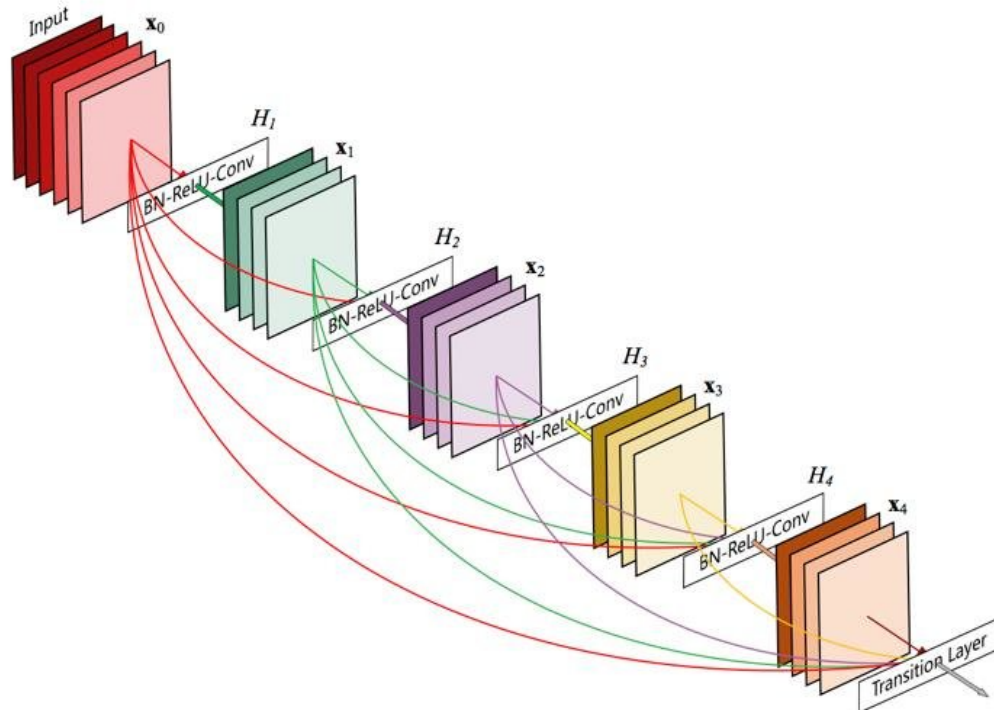


Figura 11 – Representação de um bloco denso, composto por 5 blocos de convolução, onde cada bloco de convolução tem como entrada a saída dos blocos anteriores. Fonte (HUANG et al., 2017)

3.6 Treinamento das Redes

Quatro estratégias foram definidas para a realização dos experimentos neste estudo: utilização da técnica de aumento de dados *Mixup*, denominada estratégia (a), utilização de aumento de dados por transformações geométricas, denominada estratégia (b), uma combinação das técnicas de aumento de dados por transformação geométrica e *Mixup*, denominada estratégia (c) e, finalmente, a não utilização das técnicas de aumento de dados, denominada estratégia (d). As imagens utilizadas nas estratégias foram previamente geradas e divididas em 80% para treinamento, 10% para validação e 10% para testes. Isso foi realizado de forma que os mesmos conjuntos de dados submetidos ao modelo *ResNet50* foram mantidos para o modelo *DenseNet201*.

Devido ao tamanho reduzido dos conjuntos de carcinoma e displasia, foi empregado a técnica de transferência de aprendizado (do inglês, *transfer learning*). Essa técnica consiste em aplicar o conhecimento obtido na resolução de um determinado problema em um problema similar. Ambos os modelos utilizados foram obtidos de versões pré-treinadas em mais de um milhão de imagens do banco de dados *ImageNet*.

As imagens de entrada foram redimensionadas para 125 x 225 *pixels* de resolução para realização do treinamento. Com isso, tanto a camada de entrada, quanto a primeira camada convolucional dos modelos foram substituídas para corresponder ao tamanho das imagens de entrada. A camada totalmente conectada também foi substituída por uma camada convolucional e conectada a uma camada de *dropout*. Essa camada é responsável por remover aleatoriamente alguns neurônios/pesos em cada iteração baseado em uma probabilidade. Neste estudo, a probabilidade foi definida em 40%, ou seja, a cada iteração 40% dos neurônios são desativados temporariamente. Isso ajuda a evitar que a rede se ajuste demais aos dados de treinamento (HINTON et al., 2012). Por fim, a camada de classificação foi substituída por uma nova camada de classificação para suportar as classes de saída (saúdável, leve, moderada e severa no conjunto de displasia da cavidade oral e saúdável e não saúdável para o conjunto de carcinoma escamoso da cavidade oral).

Os modelos foram treinados utilizando 50 épocas, com taxa inicial de aprendizagem de 0,0003, minilote de tamanho 16 e otimizador Adam. Um número maior de épocas foi testado, no entanto, foi observado que após 50 épocas, a rede se estabilizou e não apresentando melhora significativa no desempenho.

3.7 Métricas de Avaliação

As métricas de avaliação são índices utilizados para quantificar o desempenho dos modelos de uma rede neural. A escolha dessas métricas é determinante para garantir a qualidade de um modelo. O cálculo é realizado utilizando parâmetros retirados da matriz

de confusão, em que indica a quantidade de exemplos classificados como verdadeiros positivos (do inglês, *true positives* - tp), falsos positivos (do inglês, *false positives* - fp), verdadeiros negativos (do inglês, *true negatives* - tn) e falsos negativos (do inglês, *false negatives* - fn). Na Figura 12 é apresentado uma matriz com as informações de cada tipo.

Matriz de confusão

Classe de saída	Não saudável	<div>14</div> <div>20.0%</div> <div>(tn)</div>	<div>10</div> <div>14.3%</div> <div>(fn)</div>	<div>58.3%</div> <div>41.7%</div>
	Saudável	<div>6</div> <div>8.6%</div> <div>(fp)</div>	<div>40</div> <div>57.1%</div> <div>(tp)</div>	<div>87.0%</div> <div>13.0%</div>
		<div>70.0%</div> <div>30.0%</div>	<div>80.0%</div> <div>20.0%</div>	<div>77.1%</div> <div>22.9%</div>
		Saudável	Não saudável	
		Classe esperada		

Figura 12 – Exemplo de uma matriz de confusão para análise das imagens histológicas

Para esse estudo foram empregadas as métricas de acurácia, precisão, sensibilidade, especificidade e *F-Score*. A métrica acurácia, definida pela equação 3.1, representa a quantidade de exemplos classificados corretamente. A principal desvantagem dessa métrica é que em conjuntos desbalanceados, um valor elevado não representa necessariamente um bom desempenho do modelo.

$$Acurácia = \frac{vp + vn}{vp + vn + fp + fn}. \quad (3.1)$$

A precisão, representada pela equação 3.2, é definida pela razão de valores classificados corretamente como positivos sobre todos os valores classificados como positivo. Essa medida possibilita identificar dentre todos os exemplos classificados como positivos,

quais realmente eram positivos.

$$Precisão = \frac{vp}{vp + fp}. \quad (3.2)$$

A sensibilidade permite avaliar a capacidade do modelo de classificar como positivos os exemplos que são positivos. Essa métrica é representada pela equação 3.3.

$$Sensibilidade = \frac{vp}{vp + fn}. \quad (3.3)$$

Por outro lado, a especificidade, representada pela equação 3.4, permite avaliar a capacidade do modelo de classificar como negativos os exemplos que realmente são negativos.

$$Especificidade = \frac{vn}{vn + fp}. \quad (3.4)$$

A *F-Score* é definida pela média harmônica entre a precisão e a sensibilidade. A vantagem dessa métrica é que, caso a precisão ou a sensibilidade sejam zero ou um valor muito próximo de zero, a métrica *F-Score* será baixa. Desse modo, para que *F-Score* apresente um valor alto, precisão e sensibilidade também terão que apresentar valores altos. Logo, essa métrica tende a resumir melhor a qualidade do modelo, uma vez que evidencia que o mesmo não só possui uma boa precisão, como é capaz de acertar as previsões da classe de interesse.

$$F - Score = 2 * \frac{precisão * sensibilidade}{precisão + sensibilidade}. \quad (3.5)$$

4 Resultados

Neste capítulo são apresentados os resultados obtidos ao utilizar as diferentes técnicas de aumento de dados em arquiteturas distintas. Os resultados foram comparados a fim de avaliar a viabilidade da utilização ou não de determinada técnica ou combinação de técnicas.

4.1 Avaliação das imagens geradas utilizando a técnica *Mixup*

Nas Figuras 13 e 14 são mostrados os resultados da aplicação da técnica de aumento de dados *Mixup*. As colunas (a) e (b) representam as imagens que foram combinadas e a coluna (c) a imagem resultante dessa combinação. A escolha das imagens ocorre de forma aleatória, porém dentro de uma mesma classe, ou seja, a imagem resultante sempre será a composição de duas outras imagens de uma mesma classe. É possível observar nas imagens resultantes (Figuras 13(c) e 14(c)), características das imagens utilizadas em sua composição (Figuras 13(a), 13(b) e 14(a), 14(b)). Entretanto, também é possível observar a perda de nitidez de tais características, problema que se agrava à medida que a proporção que cada imagem representa na composição se torna mais desbalanceada.

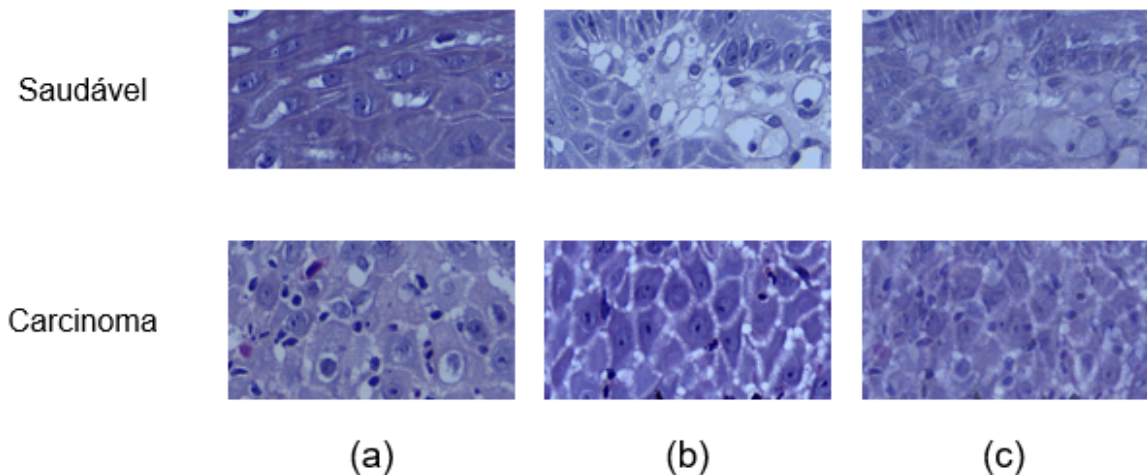


Figura 13 – Imagens de carcinoma oral de células escamosas. As imagens das colunas (a) e (b) são imagens escolhidas aleatoriamente, a coluna (c) representa a combinação resultante de (a) e (b).

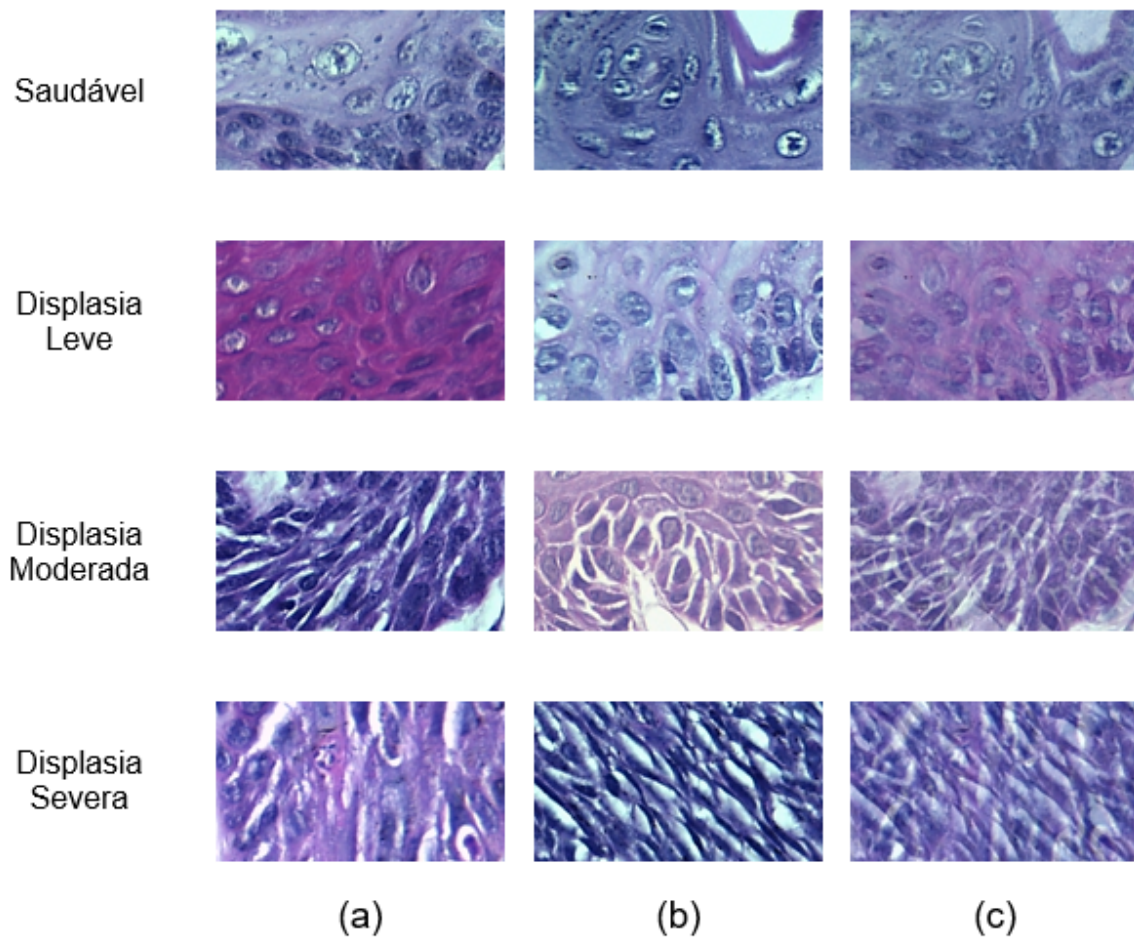


Figura 14 – Imagens de displasia oral. As imagens das colunas (a) e (b) são imagens escolhidas aleatoriamente, a coluna (c) representa a combinação resultante de (a) e (b).

4.2 Avaliação comparativa das arquitetura utilizadas

As figuras 15, 16 e 17 apresentam informações sobre o desempenho das arquiteturas *ResNet50* e *DenseNet201* nas etapas de testes para os diferentes conjuntos de dados. O eixo X representa as estratégias adotadas (ver seção 3.6) e o eixo Y o desempenho obtido de acordo com a métrica utilizada. Uma avaliação empírica foi realizada para determinar a quantidade de épocas a ser utilizado. Foram realizados treinamentos com 400 épocas para diferentes estratégias e então foi constatado que após aproximadamente 50 épocas a precisão não apresentava melhoras significativas. Dessa forma, definiu-se que todos os treinamentos seriam realizados utilizando 50 épocas.

Inicialmente, todos os treinamentos deveriam ser realizados sob as mesmas condições, no entanto a escolha do mini-lote utilizado em cada época ocorria de forma aleatória, influenciando o resultado final. Além disso, nas estratégias adotadas em que eram utiliza-

dos aumento de dados geométricos, não era possível garantir que sua aplicação ocorreria igualmente em todas as execuções. Dessa forma, o estudo comparativo das arquiteturas foi realizado levando em conta o desempenho médio obtido ao realizar cinco treinamentos consecutivos para cada uma das estratégias adotadas em cada conjunto de dados. As tabelas 2 e 3 apresentam os valores de acurácia obtidos em cada uma das execuções, por arquitetura e conjunto de dados.

Tabela 2 – Acurácia média obtida para o conjunto de dados de displasia da cavidade oral.

Displasia da cavidade oral							
Modelos	Estratégias	Execuções - AC (%)					
		1	2	3	4	5	Média
ResNet50	<i>Mixup</i>	81,2	84,4	75,0	81,2	84,4	81,2 \pm 0,03
	<i>Mixup</i> e aumento de dados	84,4	90,6	87,5	87,5	75,0	85,0 \pm 0,05
	Aumento de dados	78,1	75,0	84,4	81,2	81,2	80,0 \pm 0,03
	Sem aumento de dados ou <i>Mixup</i>	81,2	78,1	78,1	84,4	84,4	81,2 \pm 0,03
DenseNet201	<i>Mixup</i>	81,2	90,6	81,2	81,2	84,4	83,7 \pm 0,04
	<i>Mixup</i> e aumento de dados	84,4	84,4	90,6	90,6	78,1	85,6 \pm 0,05
	Aumento de dados	81,2	81,2	81,2	81,2	84,4	81,8 \pm 0,01
	Sem aumento de dados ou <i>Mixup</i>	81,2	81,2	87,5	81,2	87,5	83,7 \pm 0,03

Tabela 3 – Acurácia média obtida para o conjunto de dados de carcinoma escamoso da cavidade oral.

Carcinoma escamoso da cavidade oral							
Modelos	Estratégias	Execuções - AC (%)					
		1	2	3	4	5	Média
ResNet50	<i>Mixup</i>	77,1	81,4	74,3	75,7	80,0	77,7 \pm 0,03
	<i>Mixup</i> e aumento de dados	87,1	87,1	80,0	88,6	82,9	85,1 \pm 0,03
	Aumento de dados	82,9	81,4	81,4	80,0	84,3	82,0 \pm 0,01
	Sem aumento de dados ou <i>Mixup</i>	81,4	81,4	78,6	78,6	78,6	79,7 \pm 0,01
DenseNet201	<i>Mixup</i>	82,9	80,0	84,3	75,7	75,7	79,7 \pm 0,04
	<i>Mixup</i> e aumento de dados	85,7	87,1	88,6	88,6	92,9	88,6 \pm 0,02
	Aumento de dados	90,0	85,7	82,9	85,7	87,1	86,3 \pm 0,02
	Sem aumento de dados ou <i>Mixup</i>	84,3	82,9	81,4	84,3	81,4	82,9 \pm 0,01

Na figura 17 são apresentados os resultados da classificação dos dados de carcinoma utilizando a métrica F1-Score. O conjunto de carcinoma possui dados desbalanceados e por esse motivo a acurácia pode não resumir a performance dos modelos precisamente.

Entretanto, é possível observar que as curvas obtidas são semelhantes às apresentadas na figura 16 o que indica que os valores de acurácia obtidos não sofreram influência de uma determinada classe em detrimento da outra.

A partir dos resultados mostrados nas figuras 15, 16 e 17 fica evidente que a arquitetura *DenseNet201* apresentou um desempenho superior à arquitetura *ResNet50* em todas as estratégias adotadas.

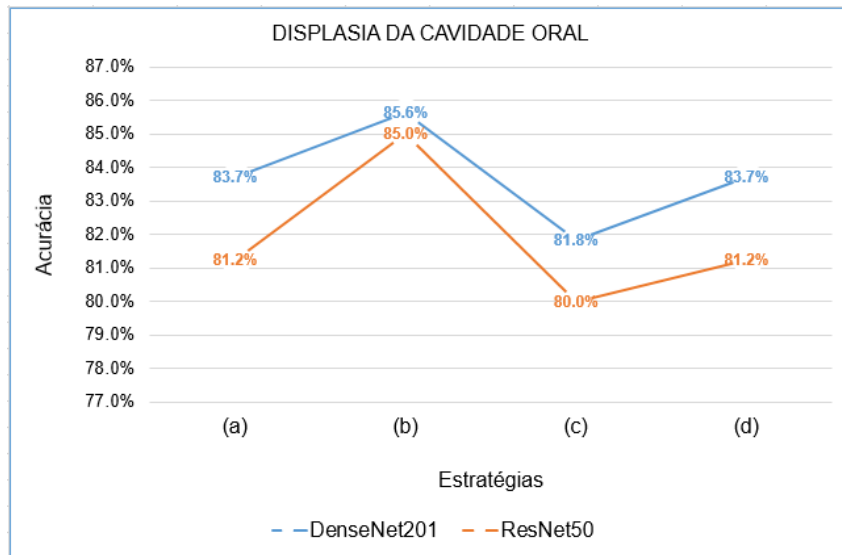


Figura 15 – Acurácia média obtida para o conjunto de dados de displasia da cavidade oral, onde (a) utilizando somente a técnica *Mixup*, (b) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

4.3 Avaliação comparativa das estratégias adotadas

Nesta etapa foram comparados os melhores resultados obtidos em cada uma das estratégias adotadas. A escolha foi feita levando em conta os testes que obtiveram o maior valor de acurácia. Foram utilizadas 32 imagens para avaliação dos modelos treinados, divididas igualmente entre quatro classes: tecido saudável, displasia leve, displasia moderada e displasia severa. Estas imagens não foram utilizadas em nenhuma etapa de treinamento ou validação. O objetivo foi avaliar se houve ganho ao utilizar a técnica de aumento de dados *Mixup*.

Nas tabelas 4 e 5 são demonstrados as métricas obtidas na etapa de testes para o conjunto de dados de displasia da cavidade oral. Os dados obtidos evidenciaram o ganho ao utilizar *Mixup* para aumentar os dados. Observa-se que para o modelo *DenseNet201*, o uso da técnica *Mixup* garantiu os maiores valores de acurácia, sendo ela combinada a outras técnicas de aumento de dados ou não. Para o modelo *ResNet50*, é possível notar que a estratégia adotada que empregou a técnica *Mixup* combinada com aumento de dados

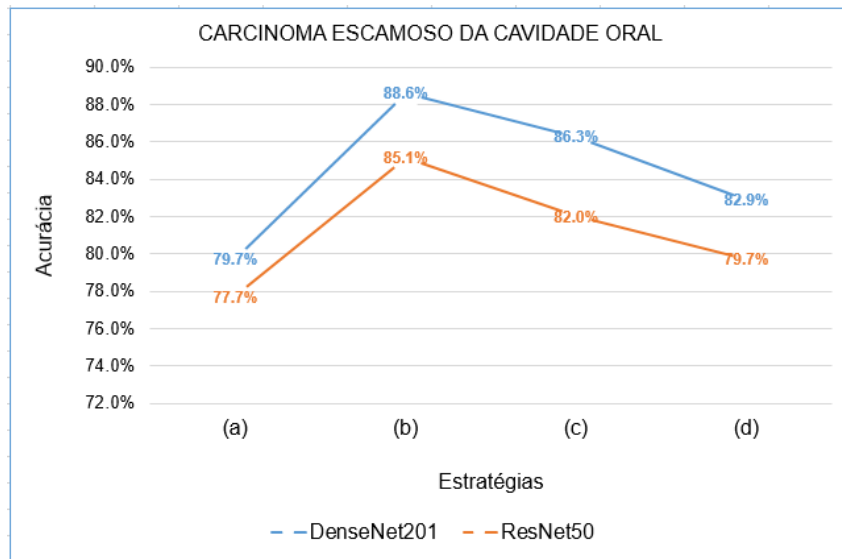


Figura 16 – Acurácia média obtida para o conjunto de dados de carcinoma escamoso da cavidade oral, onde (a) utilizando somente a técnica *Mixup*, (b) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

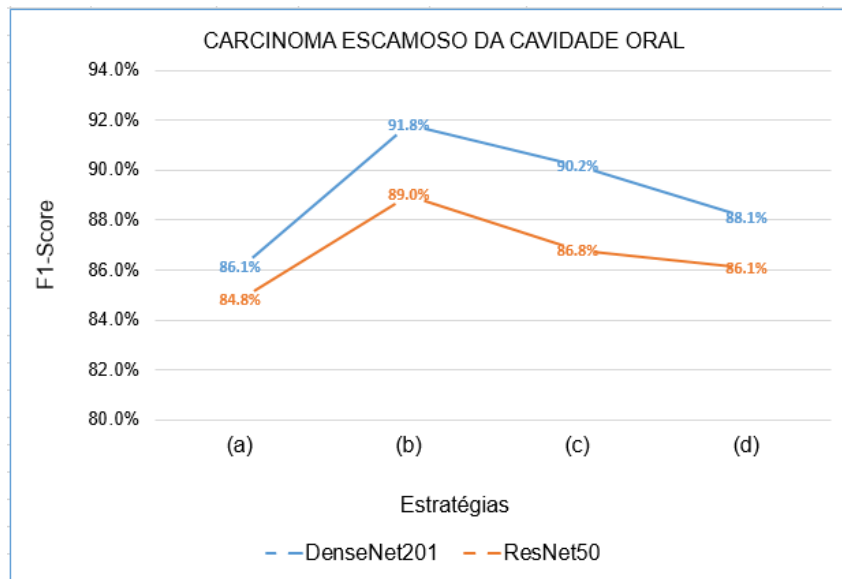


Figura 17 – F1-Score médio obtido para o conjunto de dados de carcinoma escamoso da cavidade oral, onde (a) utilizando somente a técnica *Mixup*, (b) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

geométricos apresentou o melhor resultado, no entanto o uso da técnica de forma isolada apresentou resultados tão bons quanto às demais estratégias (utilizando somente aumento de dados geométricos e utilizando somente as imagens originais).

As figuras 18 e x apresentam a taxa de perda dos modelos para as diferentes

Tabela 4 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral. Arquitetura *ResNet50*.

Displasia da cavidade oral - <i>ResNet50</i> (%)					
Estratégias	Classes	Especificidade	Precisão	Sensibilidade	Acurácia
<i>Mixup</i> e aumento de dados	Saudável	95,7	87,5	87,5	90,6
	Leve	91,7	77,8	87,5	
	Moderada	100,0	100,0	100,0	
	Severa	100,0	100,0	87,5	
<i>Mixup</i>	Saudável	100,0	100,0	100,0	84,4
	Leve	91,3	75,0	75,0	
	Moderada	90,9	77,8	87,5	
	Severa	95,5	85,7	75,0	
Aumento de dados geométricos	Saudável	100,0	100,0	87,5	84,4
	Leve	83,3	63,6	87,5	
	Moderada	95,5	85,7	75,0	
	Severa	100,0	100,0	87,5	
Sem <i>Mixup</i> ou aumento de dados geométricos	Saudável	100,0	100,0	100,0	84,4
	Leve	95,5	85,7	75,0	
	Moderada	91,3	75,0	75,0	
	Severa	90,9	77,8	87,5	

Tabela 5 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral. Arquitetura *DenseNet201*.

Displasia da cavidade oral - <i>DenseNet201</i> (%)					
Estratégias	Classes	Especificidade	Precisão	Sensibilidade	Acurácia
<i>Mixup</i> e aumento de dados	Saudável	100,0	100,0	100,0	90,6
	Leve	87,5	72,7	100,0	
	Moderada	100,0	100,0	75,0	
	Severa	100,0	100,0	87,5	
<i>Mixup</i>	Saudável	100,0	100,0	75,0	90,6
	Leve	95,7	87,5	87,5	
	Moderada	95,5	88,9	100,0	
	Severa	95,5	88,9	100,0	
Aumento de dados geométricos	Saudável	100,0	100,0	100,0	84,4
	Leve	83,3	63,6	87,5	
	Moderada	95,5	85,7	75,0	
	Severa	100,0	100,0	75,0	
Sem <i>Mixup</i> ou aumento de dados geométricos	Saudável	100,0	100,0	87,5	87,5
	Leve	91,3	77,8	87,5	
	Moderada	95,5	87,5	87,5	
	Severa	95,5	87,5	87,5	

estratégias nas fases de treinamento e validação. O eixo X representa a quantidade de iterações e o eixo Y a taxa de perda. Nota-se que as estratégias no qual foram combinadas as técnicas *Mixup* e aumento de dados geométricos apresentaram taxas de perdas mais estáveis, evidenciando que o modelo foi capaz de generalizar, ou seja, capaz de classificar

dados desconhecidos corretamente.

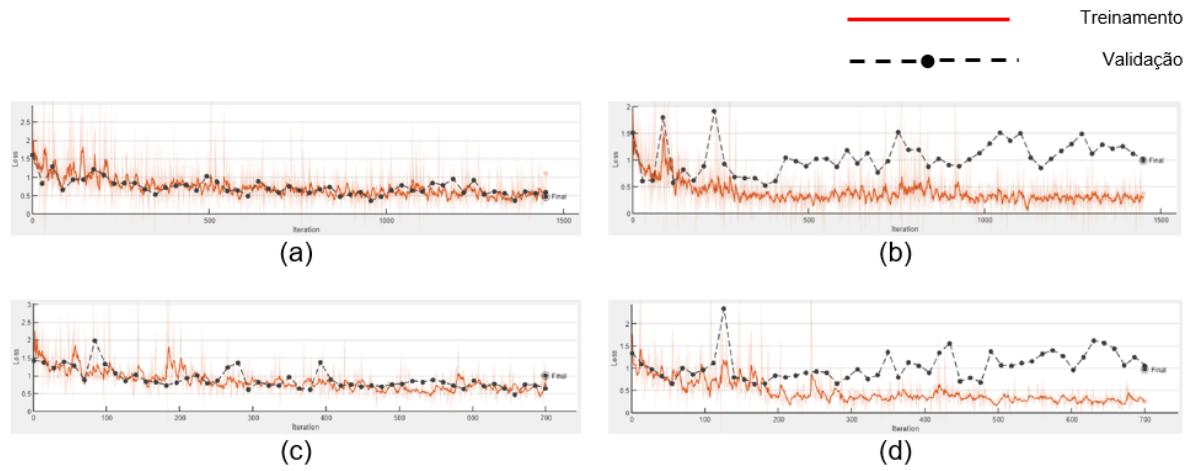


Figura 18 – Taxa de perda para as diferentes estratégias. Conjunto de dados de displasia bucal, arquitetura ResNet50, onde (a) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (b) utilizando somente a técnica *Mixup*, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

5 Conclusão

Descrever aqui as conclusões e/ou considerações finais.

Destacar as contribuições originais do trabalho.

Propor trabalhos futuros em continuidade ao trabalho realizado.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 6028*: Resumo - apresentação. Rio de Janeiro, 2003. 2 p. Citado na página 6.

BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, IEEE, v. 5, n. 2, p. 157–166, 1994. Citado na página 21.

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 21.

HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012. Citado na página 25.

HUANG, G. et al. *Densely Connected Convolutional Networks*. arXiv, 2017. Disponível em: <<https://arxiv.org/abs/1608.06993>>. Citado 4 vezes nas páginas 7, 21, 23 e 24.

Matlab documentation. *Preprocess Images for Deep Learning*. 2022. Disponível em: <<https://www.mathworks.com/help/deeplearning/ug/preprocess-images-for-deep-learning.html>>. Acesso em: 19 de dezembro 2022. Citado 2 vezes nas páginas 7 e 21.

RAHMAN, T. Y. *A histopathological image repository of normal epithelium of Oral Cavity and Oral Squamous Cell Carcinoma*. Mendeley, 2019. Disponível em: <<https://data.mendeley.com/datasets/ftmp4cvtmb/1>>. Citado na página 17.

SILVA, A. B. *Métodos Computacionais para Análise e Classificação de Displasias em Imagens da Cavidade Bucal*. Dissertação (Mestrado) — Universidade Federal de Uberlândia - Faculdade de Computação, 2019. Citado 3 vezes nas páginas 7, 17 e 22.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 21.

TECHAMERICA. *ANSI/EIA 649-B: Configuration Management Standard*. EUA, 2011. Citado na página 14.

ZHANG, H. et al. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. Citado na página 19.

Apêndices

APÊNDICE A – Coisas que fiz e que achei interessante mas não tanto para entrar no corpo do texto

Quisque facilisis auctor sapien. Pellentesque gravida hendrerit lectus. Mauris rutrum sodales sapien. Fusce hendrerit sem vel lorem. Integer pellentesque massa vel augue. Integer elit tortor, feugiat quis, sagittis et, ornare non, lacus. Vestibulum posuere pellentesque eros. Quisque venenatis ipsum dictum nulla. Aliquam quis quam non metus eleifend interdum. Nam eget sapien ac mauris malesuada adipiscing. Etiam eleifend neque sed quam. Nulla facilisi. Proin a ligula. Sed id dui eu nibh egestas tincidunt. Suspendisse arcu.

Anexos

ANEXO A – Coisas que eu não fiz mas que achei interessante o suficiente para colocar aqui

Sed consequat tellus et tortor. Ut tempor laoreet quam. Nullam id wisi a libero tristique semper. Nullam nisl massa, rutrum ut, egestas semper, mollis id, leo. Nulla ac massa eu risus blandit mattis. Mauris ut nunc. In hac habitasse platea dictumst. Aliquam eget tortor. Quisque dapibus pede in erat. Nunc enim. In dui nulla, commodo at, consectetur nec, malesuada nec, elit. Aliquam ornare tellus eu urna. Sed nec metus. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.