

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Diego Rodrigues de Jesus

**Investigação de Técnicas de Aumento de Dados
para Classificação de Lesões da Cavidade Oral**

Uberlândia, Brasil

2023

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Diego Rodrigues de Jesus

**Investigação de Técnicas de Aumento de Dados para
Classificação de Lesões da Cavidade Oral**

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 Sistemas de Informação.

Orientador: Marcelo Zanchetta do Nascimento

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2023


UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Faculdade de Computação

Av. João Naves de Ávila, nº 2121, Bloco 1A - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902

 Telefone: (34) 3239-4144 - <http://www.portal.facom.ufu.br/> facom@ufu.br

ATA DE DEFESA - GRADUAÇÃO

| | | | | | |
|--|--|-----------------|-------|-----------------------|-------|
| Curso de Graduação em: | Sistemas de Informação | | | | |
| Defesa de: | FACOM31802 – Trabalho de Conclusão de Curso II | | | | |
| Data: | 06/07/2023 | Hora de início: | 17:00 | Hora de encerramento: | 18:15 |
| Matrícula do Discente: | 11611BSI200 | | | | |
| Nome do Discente: | Diego Rodrigues de Jesus | | | | |
| Título do Trabalho: | Investigação de Técnicas de Aumento de Dados para Classificação de Lesões da Cavidade Oral | | | | |
| A carga horária curricular foi cumprida integralmente? | <input checked="" type="checkbox"/> Sim <input type="checkbox"/> Não | | | | |

Reuniu-se no na plataforma Microsoft Teams, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Curso de Graduação em Sistemas de Informação, assim composta: Prof. Dr. Paulo Henrique Ribeiro Gabriel - FACOM/UFU; Doutorando Adriano Barbosa Silva - PPGCO/UFU; Prof. Dr. Marcelo Zanchetta do Nascimento, orientador(a) do(a) candidato(a).

Iniciando os trabalhos, o(a) presidente da mesa, Dr(a). Marcelo Zanchetta do Nascimento, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao discente a palavra, para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do curso.

A seguir o(a) senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado(a) Nota [95] (Somente números inteiros)

OU

Aprovado(a) sem nota.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento, Professor(a) do Magistério Superior**, em 06/07/2023, às 18:14, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Paulo Henrique Ribeiro Gabriel, Professor(a) do Magistério Superior**, em 06/07/2023, às 18:14, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Adriano Barbosa Silva, Usuário Externo**, em 06/07/2023, às 18:18, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4628736** e o código CRC **5FC79DD9**.

Dedico este trabalho à minha mãe (in memoriam), meu maior exemplo de resiliência.

Agradecimentos

Agradeço primeiramente a Deus por me dar forças para continuar nos momentos em que pensei em desistir, aos meus pais, em especial à minha mãe pelas longas conversas e conselhos, à minha irmã por sempre acreditar em mim mais do que eu mesmo, aos meus amigos e familiares que proporcionaram um ambiente favorável para a conclusão deste trabalho, à minha esposa Angélica e minha filha Ágatha que me apoiaram incondicionalmente e que suportaram pacientemente todos os momentos de ansiedade e indecisão e finalmente ao meu orientador Marcelo Zanchetta do Nascimento por toda a paciência, educação, incentivo e aprendizado que tornaram esse trabalho possível.

Resumo

O diagnóstico precoce de câncer oral é fundamental na determinação do prognóstico do paciente, pois permite que profissionais de saúde adotem medidas preventivas, reduzindo o risco de evolução para estágios mais avançados da doença. No entanto, a avaliação de lesões da cavidade oral depende da interpretação por especialistas das características clínicas, o que pode levar a diferenças de opiniões e dificuldades no diagnóstico. O uso de sistemas de apoio ao diagnóstico pode ser aplicado como ferramenta para auxiliar e aprimorar a classificação e decisão do especialista, porém esses sistemas ainda possuem uma forte dependência de conjuntos robustos de dados para treinamento e a obtenção de imagens histológicas apresenta várias dificuldades e desafios. Com o objetivo de aprimorar o processo de diagnóstico de lesões da cavidade oral, este trabalho propõe o estudo de técnicas de aumento de dados como alternativa ao problema de escassez de imagens histológicas para etapa de classificação das lesões. São investigadas as técnicas de aumento de dados por transformações geométricas e *Mixup*, que consiste na combinação de duas ou mais amostras do conjunto de dados de treinamento para criar uma nova amostra aumentada. Para classificação foram utilizados os modelos CNNs *ResNet50* e *DenseNet201*. A combinação das técnicas de aumento de dados por transformações geométricas e *Mixup* apresentaram os melhores resultados e maior capacidade de generalização. A estratégia obteve acurácia de 90,6% para o conjunto de displasia oral nos dois modelos de aumento de dados e, para o conjunto de saudáveis e carcinoma, apresentou valores de F1-Score de 91,7% com o modelo *ResNet50* e 94,9% com o modelo *DenseNet201*. A investigação mostrou que a técnica *Mixup* não substitui técnicas convencionais de aumento de dados, no entanto pode complementá-las para o avanço na classificação.

Palavras-chave: Lesões da cavidade oral, Redes Neurais Convolucionais, Aumento de dados, Imagens Histológicas, *Mixup*, Displasia.

Lista de ilustrações

| | |
|--|----|
| Figura 1 – Alterações histopatológicas do epitélio oral. Fonte (Có, 2019) | 18 |
| Figura 2 – Visão abrangente das etapas de um sistema CAD para diagnóstico de imagens histopatológicas. Adaptado de (KAUSHAL et al., 2019) | 20 |
| Figura 3 – Representação de um neurônio biológico. Fonte (SILVA, 2019) | 22 |
| Figura 4 – Representação de um neurônio artificial. Fonte (SILVA, 2019). Adaptado de (HAYKIN, 2009) | 23 |
| Figura 5 – Exemplo de uma operação de convolução: (a) Representação de uma imagem de tamanho 7×5 <i>pixels</i> , (b) representação de um filtro de convolução de tamanho 3×3 <i>pixels</i> e (c) matriz de característica resultante. Fonte (SILVA, 2019). | 24 |
| Figura 6 – Exemplo de uma entrada passando por uma ativação ReLU. Adaptado de (ROSEBROCK, 2017). | 24 |
| Figura 7 – Exemplo das operações de <i>max pooling</i> e <i>average pooling</i> . Adptado de (FREITAS et al., 2021). | 25 |
| Figura 8 – Exemplo de aumento de dados por transformações geométricas: (a) imagem original, (b) imagem espelhada verticalmente, (c) rotação seguida de corte, (d) ampliação e (e) cisalhamento. | 26 |
| Figura 9 – Visão geral das principais etapas de classificação de lesões da cavidade bucal. | 30 |
| Figura 10 – Exemplos de tecido histológico da cavidade bucal: (a) saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa. | 31 |
| Figura 11 – Exemplo das imagens originais de tecido histopatológico da cavidade oral: (a) saudável, (b) OSCC. | 32 |
| Figura 12 – Exemplo das imagens de tecido histopatológico da cavidade oral: (a) saudável, (b) OSCC, redimensionadas a partir do centro das imagens da Figura 11. | 32 |
| Figura 13 – 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} | 34 |
| Figura 14 – As imagens (a) e (b) são amostras de tecidos com displasia leve, (c) imagem construída por meio da combinação de (a) e (b). | 34 |
| Figura 15 – Exemplo de transformação dos dados de treinamento para cada época. Fonte (Matlab documentation, 2022) | 35 |
| Figura 16 – Arquitetura da rede neural <i>Resnet50</i> . M1, M2, M3, M4 representam os módulos compostos por blocos residuais. Adaptado de (SILVA, 2019). | 35 |

| | |
|---|----|
| Figura 17 – Blocos residuais presentes na arquitetura <i>ResNet50</i> : (a) bloco residual sem convolução 1x1 e (b) bloco residual com convolução 1×1 | 36 |
| Figura 18 – Arquitetura da rede neural <i>DenseNet201</i> . B1, B2, B3, B4 representam os blocos densos. T1, T2, T3 representam as camadas de transição. | 37 |
| Figura 19 – 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) | 38 |
| Figura 20 – Exemplo de uma matriz de confusão para análise das imagens histológicas | 40 |
| Figura 21 – Imagens de carcinoma oral de células escamosas. As imagens das colunas (a) e (b) foram escolhidas aleatoriamente, a coluna (c) representa a combinação resultante das imagens de (a) e (b). | 42 |
| Figura 22 – Imagens de displasia oral. As imagens das colunas (a) e (b) foram escolhidas aleatoriamente, a coluna (c) representa a combinação resultante de (a) e (b). | 43 |
| Figura 23 – Imagens aumentadas utilizando transformações geométricas: (a) imagens originais e (b) imagens transformadas através da combinação aleatória de operações geométricas. | 44 |
| Figura 24 – 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. | 46 |
| Figura 25 – 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. | 46 |
| Figura 26 – 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. | 47 |
| Figura 27 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de displasia oral e a arquitetura <i>ResNet50</i> , onde (a) utilizando <i>Mixup</i> e aumento de dados geométricos, (b) utilizando somente <i>Mixup</i> , (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados. | 49 |

| | |
|---|----|
| Figura 28 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de displasia oral e a arquitetura <i>DenseNet201</i> , onde (a) utilizando <i>Mixup</i> e aumento de dados geométricos, (b) utilizando somente <i>Mixup</i> , (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados. | 49 |
| Figura 29 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de carcinoma escamoso da cavidade oral e a arquitetura <i>ResNet50</i> , onde (a) utilizando <i>Mixup</i> e aumento de dados geométricos, (b) utilizando somente <i>Mixup</i> , (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados. | 50 |
| Figura 30 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de carcinoma escamoso da cavidade oral e a arquitetura <i>DenseNet201</i> , onde (a) utilizando <i>Mixup</i> e aumento de dados geométricos, (b) utilizando somente <i>Mixup</i> , (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados. | 51 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 – Acurácia média obtida para o conjunto de dados de displasia da cavidade oral. | 45 |
| Tabela 2 – Acurácia média obtida para o conjunto de dados de carcinoma escamoso da cavidade oral. | 45 |
| Tabela 3 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral. Arquitetura <i>ResNet50</i> | 47 |
| Tabela 4 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral. Arquitetura <i>DenseNet201</i> | 48 |
| Tabela 5 – Resultados obtidos ao classificar as imagens de teste de carcinoma escamoso da cavidade oral. | 50 |
| Tabela 6 – Avaliação entre o modelo investigados para classificação das lesões e os estudos presentes na Literatura. | 51 |

Lista de abreviaturas e siglas

| | |
|------|---------------------------------------|
| CAD | <i>Computer Aided Diagnosis</i> |
| CNN | <i>Convolutional Neural Network</i> |
| DA | <i>Data Augmentation</i> |
| FNC | <i>Fully Convolutional Network</i> |
| GAN | <i>Generative Adversarial Network</i> |
| H&E | Hematoxilina e Eosina |
| HOP | <i>Hermite Orthogonal Polynomial</i> |
| IA | Inteligência Artificial |
| OSCC | <i>Oral Squamous Cell Carcinoma</i> |
| ReLU | <i>Rectified Linear Unit</i> |
| ROI | <i>Region of Interest</i> |
| RGB | <i>Red, Green and Blue</i> |
| RNA | Redes Neurais Artificiais |
| WSI | <i>Whole Slide Image</i> |

Sumário

| | | |
|------------|---|-----------|
| 1 | INTRODUÇÃO | 14 |
| 1.1 | Objetivos | 15 |
| 1.1.1 | Objetivo Geral | 15 |
| 1.1.2 | Objetivos Específicos | 15 |
| 1.2 | Organização do Trabalho | 16 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 17 |
| 2.1 | Lesões da cavidade oral | 17 |
| 2.2 | Histologia | 18 |
| 2.3 | Sistemas de apoio ao diagnóstico | 19 |
| 2.4 | Redes Neurais Artificiais | 21 |
| 2.5 | Redes Neurais Convolucionais | 22 |
| 2.6 | Aumento de Dados | 25 |
| 2.7 | Estado da Arte | 27 |
| 2.7.1 | Estudos de classificação de displasia da cavidade oral | 27 |
| 2.7.2 | Estudos de classificação de carcinoma escamoso da cavidade oral | 28 |
| 2.8 | Considerações Finais | 28 |
| 3 | METODOLOGIA | 30 |
| 3.1 | Visão Geral | 30 |
| 3.2 | Banco de Imagens | 30 |
| 3.2.1 | Lesões displásicas da cavidade oral | 31 |
| 3.2.2 | Carcinoma escamoso da cavidade oral - OSCC | 31 |
| 3.3 | Aumento dados utilizando a técnica <i>Mixup</i> | 33 |
| 3.4 | Aumento de dados por transformações geométricas | 33 |
| 3.5 | Redes Neurais Convolucionais | 34 |
| 3.5.1 | ResNet50 | 35 |
| 3.5.2 | DenseNet201 | 37 |
| 3.6 | Treinamento das Redes | 37 |
| 3.7 | Métricas de Avaliação | 39 |
| 4 | RESULTADOS | 41 |
| 4.1 | Composição de Imagens com o uso da Técnica <i>Mixup</i> | 41 |
| 4.2 | Composição de Imagens com o uso de Transformações Geométricas | 41 |
| 4.3 | Avaliação da Classificação das Lesões com os Modelos de CNNs | 42 |
| 4.4 | Estudo das Melhores Estratégias Empregadas nos Bancos de Imagens | 44 |

| | | |
|-----|---|----|
| 4.5 | Análise do método investigado com estudos da literatura | 51 |
| 5 | CONCLUSÃO | 52 |
| 5.1 | Desafios Encontrados | 53 |
| | REFERÊNCIAS | 54 |

1 Introdução

O câncer da cavidade oral engloba tumores malignos que afetam os tecidos da boca, incluindo os lábios, gengivas, língua, bochechas e o assoalho da boca. Os principais fatores de risco para o desenvolvimento desse tipo de câncer são o tabagismo, o consumo excessivo de álcool, a exposição prolongada ao sol, o vírus do papiloma humano (HPV) e a má higiene oral (INCA, 2022; TILAKARATNE et al., 2019).

Segundo INCA (2023), estima-se que haverá aproximadamente 15.100 novos casos de câncer de cavidade oral no Brasil anualmente, durante o período de 2023 a 2025. Isso equivale a uma taxa estimada de 6,99 casos por 100 mil habitantes, sendo 10.900 casos em homens e 4.200 casos em mulheres. Esses números indicam um risco estimado de 10,30 novos casos para cada 100 mil homens e 3,83 novos casos para cada 100 mil mulheres.

O diagnóstico precoce do câncer da cavidade oral desempenha um papel fundamental na determinação do prognóstico do paciente. Quando o câncer é detectado em estágios iniciais, as opções de tratamento são mais promissoras e as taxas de sobrevivência são significativamente melhores (ABATI et al., 2020). Além disso, um diagnóstico precoce permite que os profissionais de saúde adotem medidas preventivas, como a remoção de lesões pré-malignas, reduzindo assim o risco de progressão para um estágio mais avançado da doença.

As displasias orais são lesões pré-malignas que podem se desenvolver em tecidos orais. Elas são caracterizadas por alterações celulares anormais, que ainda não atingiram o estágio de câncer invasivo. O risco de transformação da displasia epitelial em carcinoma foi relatado pela primeira vez por Richart e Barron (1969). Isso significa que, se as displasias não forem tratadas adequadamente, elas têm maior probabilidade de progredir para um câncer oral. Portanto, o monitoramento regular e o tratamento das displasias são essenciais para a detecção precoce e a prevenção do carcinoma escamoso (TILAKARATNE et al., 2019). No entanto, a avaliação das lesões orais muitas vezes depende da experiência e habilidade do examinador. A interpretação das características clínicas pode variar entre os profissionais de saúde, o que pode levar a diferentes opiniões e dificuldades no diagnóstico (TILAKARATNE et al., 2019).

Os sistemas de auxílio ao diagnóstico baseados em computador (do inglês, *Computer-aided diagnosis* - CAD) tem sido amplamente utilizados para análise de imagens histopatológicas (DAS; NAIR; PETER, 2020). Esses sistemas utilizam algoritmos de processamento digital de imagens e técnicas de aprendizado profundo de máquina para analisar imagens médicas e fornecer informações adicionais para auxiliar no processo de tomada de decisão (CHAN; HADJIISKI; SAMALA, 2020). No caso do câncer, os sistemas CAD

podem ajudar a identificar características específicas das lesões malignas, assim como, fornecer análises de probabilidades e contribuir com informações suplementares para um diagnóstico mais preciso (LI et al., 2022).

Para desenvolver um sistema de aprendizado de máquina robusto para uma determinada tarefa, é necessário um conjunto de dados de amostra que seja tanto extenso quanto representativo de cada classe da população relevante. Esse conjunto de dados permite que o algoritmo de aprendizado de máquina modele com precisão as propriedades estatísticas da população em questão e avalie corretamente quaisquer novos casos desconhecidos pertencentes à mesma população (CHAN; HADJIISKI; SAMALA, 2020). No entanto, a obtenção desse conjunto de dados é dispendiosa, principalmente devido à necessidade de rotulagem e anotação, que frequentemente requerem o envolvimento de mais de um especialista clínico devido às variações inter e intra-leitor na interpretação das imagens e na avaliação da doença (CHAN; HADJIISKI; SAMALA, 2020).

Um método comumente usado para aumentar o tamanho aparente da amostra de treinamento é o uso de técnicas de aumento de dados (do inglês, *data augmentation* - DA). Essas técnicas envolvem a criação de novas amostras de treinamento, introduzindo variações nas imagens existentes (KHOSLA; SAINI, 2020). A aplicação dessas técnicas ajuda a aumentar a diversidade dos dados disponíveis para o treinamento dos sistemas CAD, melhorando assim a capacidade de classificação. Com base nessas considerações, esse trabalho se propõe a investigar técnicas de aumento de dados em conjuntos limitados de displasia da cavidade oral e carcinoma escamoso da cavidade oral. Os algoritmos de DA discutidos nesse estudo incluem transformações geométricas e a técnica *Mixup* (ZHANG et al., 2017), que consiste na combinação de duas ou mais imagens com o propósito de obter uma nova imagem.

1.1 Objetivos

1.1.1 Objetivo Geral

O objetivo deste trabalho é investigar técnicas de aumento de dados em conjuntos de imagens histológicas, envolvendo a geração de exemplos de treinamento sintéticos, para classificação de lesões baseados em modelos diferentes de redes DL.

1.1.2 Objetivos Específicos

- Implementação da técnica de aumento de dados *Mixup* para geração de imagens sintéticas de lesões da cavidade oral;
- Comparação entre as técnicas de aumento de dados por transformações geométricas

e *Mixup*;

- Investigação das arquiteturas de redes convolucionais *ResNet50* e *DenseNet201* para classificação de imagens histológicas de displasia da cavidade oral e carcinoma escamoso da cavidade oral;
- Comparação dos resultados em diferentes bases de imagens histológicas da cavidade oral.

1.2 Organização do Trabalho

Neste capítulo, são fornecidos, resumidamente, os objetivos da metodologia proposta e a razão por trás do desenvolvimento deste trabalho. Os próximos capítulos deste documento estão estruturados da seguinte maneira:

- Capítulo 2: apresenta conceitos necessários para a compreensão deste trabalho e o estado da arte, onde são listados trabalhos que abordam o processamento de lesões da cavidade oral;
- Capítulo 3: descreve as técnicas e metodologias desenvolvidas, bem como as métricas utilizadas na avaliação do método;
- Capítulo 4: apresenta os resultados obtidos em cada uma das estratégias de aumento de dados adotadas e um comparativo com resultados obtidos por outros métodos presentes da literatura;
- Capítulo 5: apresenta as conclusões do trabalho proposto e uma sugestão de trabalhos futuros.

2 Fundamentação Teórica

Este capítulo apresenta conceitos fundamentais para a compreensão deste trabalho, abrangendo temas de diversas áreas, como Computação e Biologia. Esses temas serão introduzidos de forma clara e concisa ao longo deste capítulo.

2.1 Lesões da cavidade oral

A displasia epitelial é uma condição cientificamente descrita como um distúrbio de crescimento em que ocorre uma proliferação anormal ou atípica das células epiteliais. Após observar a ocorrência de distúrbios potencialmente malignos na mucosa adjacente aos carcinomas escamosos orais, sustentou-se a hipótese de transformação maligna da displasia epitelial oral, o que incentivou pesquisadores a investigarem a relação entre a displasia epitelial e distúrbios potencialmente malignos, assim como a taxa de transformação maligna (CHIERICI, 1968; WRIGHT; SHEAR, 1985).

Em 2005, a Organização Mundial de Saúde (OMS) publicou um artigo (BARNES et al., 2005) em que definiu-se as categorias das displasias da cavidade oral em cinco níveis: hiperplasia de células escamosas, displasia leve, displasia moderada, displasia severa e carcinoma *in situ*. Em 2017, uma atualização foi publicada em (MÜLLER, 2017) onde os níveis de displasia foram reduzidos de cinco para três (leve, moderada e severa), eliminando a hiperplasia (um processo não displásico por definição) e com displasia severa e carcinoma *in situ* considerados sinônimos. Carcinoma *in situ* é caracterizado por alterações celulares acentuadas e perda significativa da arquitetura normal do tecido, com alto risco de evolução para carcinoma escamoso invasivo, estágio onde as células cancerosas invadem o tecido subjacente. Nesse ponto, o câncer já é considerado maligno e capaz de se espalhar para outras áreas do corpo. Recentemente a OMS publicou uma nova edição em (MULLER; TILAKARATNE, 2022) onde foi mantido o sistema de classificação de displasias da cavidade oral em três níveis.

Segundo os autores em (SHULMAN; GONZALES, 2008), a displasia epitelial oral não segue uma progressão sequencial previsível de leve para moderada e ou severa. Não é incomum que uma displasia leve progrida rapidamente para um carcinoma invasivo, no entanto, nem todas as displasias epiteliais evoluem para carcinoma. Carcinomas orais podem surgir de lesões nas quais a displasia epitelial não foi diagnosticada. Por conta dessa imprevisibilidade na progressão das displasias orais, exames histopatológicos são fundamentais para identificação precoce das lesões e tratamento adequado as características. Na Figura 1 são apresentados modelos das alterações histopatológicas do epitélio oral.

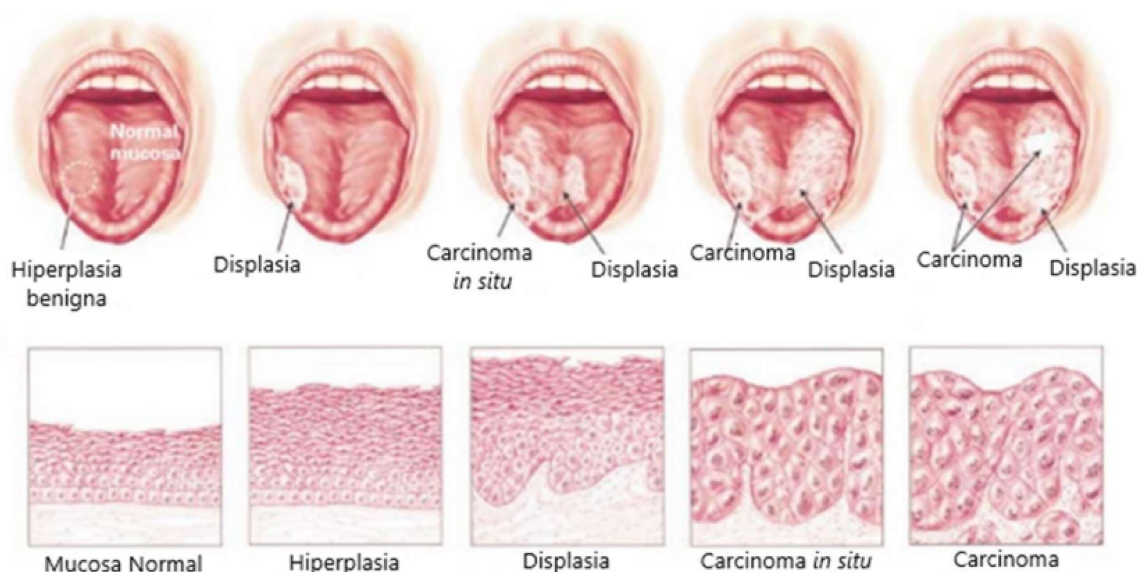


Figura 1 – Alterações histopatológicas do epitélio oral. Fonte (Có, 2019)

2.2 Histologia

A histologia é uma disciplina que se dedica ao estudo da estrutura microscópica dos tecidos biológicos e à organização celular. Para realizar a análise histológica, é necessário realizar uma série de processos que têm como objetivo preservar a integridade estrutural e as características celulares dos tecidos, possibilitando sua observação e análise ao microscópio. Conforme definido pelos autores em (GARTNER, 2007; JUNQUEIRA; CARNEIRO, 2013), os procedimentos para obtenção das imagens são descritos a seguir.

O primeiro processo é a fixação, que consiste em preservar os tecidos, evitando a decomposição pós-morte. O fixador mais comumente utilizado é o formaldeído, que estabiliza as proteínas e previne a degradação enzimática. Existem diferentes métodos de fixação, como a imersão e a perfusão.

Após a fixação, ocorre a desidratação dos tecidos em que a água presente nos tecidos é removida e substituída por solventes hidrofóbicos, como o álcool. A desidratação é realizada por meio de uma série de banhos alcoólicos sucessivos, começando geralmente com álcool etílico diluído e avançando para concentrações mais altas, até que toda a água seja removida. A desidratação é importante porque a água pode interferir na penetração adequada dos reagentes de inclusão, como a parafina. Além disso, a água pode causar encolhimento ou expansão dos tecidos, comprometendo sua estrutura e a análise subsequente.

Em seguida, os tecidos são incluídos em uma substância de inclusão, como parafina ou a resina plástica. Essa etapa enrijece os tecidos, permitindo a obtenção de cortes finos

para a análise microscópica. A parafina é um material de inclusão amplamente utilizado na histologia.

Após a inclusão, os tecidos imersos em parafina são cortados em lâminas finas utilizando um equipamento chamado micrótomo. Os cortes obtidos são montados em lâminas de vidro e podem passar por diferentes técnicas de coloração, que visam realçar estruturas específicas dos tecidos. A coloração é um passo crucial na análise histológica, pois permite a visualização de diferentes componentes e estruturas. A coloração mais comum é a Hematoxilina-Eosina (H&E), amplamente utilizada para visualizar núcleos e citoplasma. Essa técnica utiliza a Hematoxilina para corar os núcleos das células em tons de azul ou roxo e a Eosina para corar o citoplasma em tons de rosa.

Os tecidos são digitalizados com o uso de microscópios equipados com câmeras de alta resolução para avaliação de especialistas. Além disso, essas informações podem ser exploradas por técnicas de Inteligência Artificial (IA) e aprendizado de máquina. Algoritmos de IA podem ser treinados para reconhecer e classificar características histológicas específicas, auxiliando na detecção de doenças ou na diferenciação de lesões. Esses sistemas podem ser integrados a plataformas de diagnóstico por imagem ou a sistemas de apoio ao diagnóstico em patologia, melhorando a precisão e a eficiência dos diagnósticos.

2.3 Sistemas de apoio ao diagnóstico

Os sistemas CAD são ferramentas e métodos que auxiliam profissionais de saúde no processo de diagnóstico médico. Eles são projetados para melhorar a precisão, eficiência e confiabilidade do diagnóstico, fornecendo informações adicionais e análises automatizadas para complementar a expertise do médico (RAMADAN, 2020). O processo de um sistema CAD geralmente envolve várias etapas, conforme apresentado na Figura 2.

A primeira etapa corresponde à aquisição dos dados. Nesta etapa, são coletados os dados médicos necessários para o diagnóstico, como imagens radiológicas, ressonâncias magnéticas, tomografias ou as imagens histopatológicas. Esses dados são adquiridos por meio de equipamentos médicos adequados e armazenados em formato digital.

Após a aquisição dos dados, é realizada uma etapa de pré-processamento para melhorar a qualidade das imagens. Isso pode incluir a remoção de ruídos, a correção de distorções ou o realce de características relevantes (GONZALEZ; WOODS, 2010). O objetivo é garantir que as imagens estejam em condições adequadas para análise.

Em seguida, os dados pré-processados são submetidos à etapa de segmentação, na qual as regiões de interesse (do inglês, *Regions of interest* - ROI) são identificadas e separadas do restante da imagem. Isso pode ser feito por meio de algoritmos que detectam bordas, limiares de intensidade ou padrões específicos (KAUSHAL et al., 2019). Isso ajuda

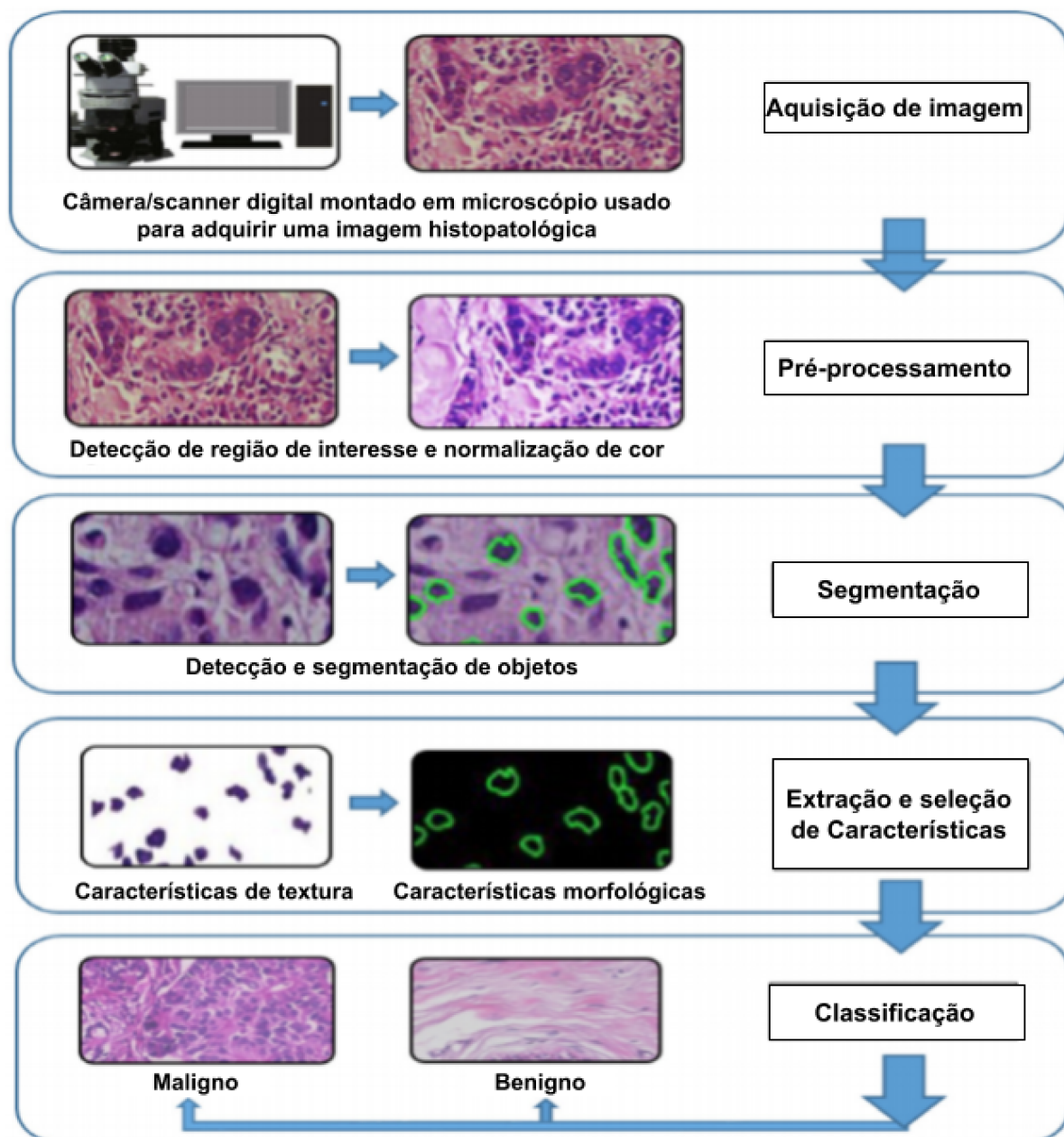


Figura 2 – Visão abrangente das etapas de um sistema CAD para diagnóstico de imagens histopatológicas. Adaptado de (KAUSHAL et al., 2019)

a reduzir a quantidade de dados a serem analisados e concentra a atenção nos pontos de interesse.

Na etapa de extração de características, são identificadas e obtidas as características relevantes das regiões de interesse que foram segmentadas anteriormente que sejam significativas para diferenciação das classes (GONZALEZ; WOODS, 2010). Essas características podem incluir medidas de tamanho, forma, densidade, textura ou outras propriedades específicas das regiões analisadas. A seleção das características a serem extraídas depende da área médica específica e dos objetivos do diagnóstico em questão, como por

exemplo, as características relativas aos núcleos celulares, que podem ser utilizadas para identificar o surgimento de lesões e progressão de câncer (HANNEN et al., 1998).

Por fim, na etapa de classificação, algoritmos e técnicas avançadas, como aprendizado de máquina ou IA, são aplicados aos dados e características extraídas. Esses algoritmos são treinados utilizando grandes conjuntos de dados e podem ajudar a identificar padrões. Nesse processo, um especialista desempenha um papel importante, realizando uma classificação manual inicial de algumas imagens e determinando quais características podem ser utilizadas para identificar a presença de doenças (SILVA, 2019). A classificação é usada para identificar a qual classe uma nova instância pertence com base no conjunto de dados de treinamento disponível (KAUSHAL et al., 2019).

2.4 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no funcionamento do cérebro humano, criados para realizar tarefas de aprendizado e reconhecimento de padrões (ROSEBROCK, 2017; HAYKIN, 2009). Os neurônios são os elementos fundamentais do sistema nervoso e são responsáveis pela recepção, transmissão e processamento de estímulos. Esses elementos são formados basicamente por três componentes: dendritos, corpo celular e axônios. O corpo celular recebe informações provenientes dos dendritos, que são as estruturas receptoras do neurônio. Essas informações podem ser sinais elétricos ou químicos transmitidos por outros neurônios. O corpo celular integra esses sinais, somando e processando as informações recebidas de várias fontes. Essa integração é crucial para a tomada de decisões do neurônio e para a geração de uma resposta apropriada. O axônio fica então responsável pela transmissão dos impulsos gerados a outras células nervosas, musculares ou glandulares. Na porção terminal do axônio, há os neurotransmissores que são substâncias químicas responsáveis pela transmissão do impulso nervoso através das sinapses (FURTADO, 2019; JUNQUEIRA; CARNEIRO, 2013). Na Figura 3 é apresentada uma representação de um neurônio biológico.

Um neurônio artificial é uma estrutura lógico-matemática projetada para imitar a forma, o comportamento e as funções de um neurônio biológico, sendo considerado uma unidade fundamental para processamento de informações. De maneira simplificada, é possível fazer uma associação entre o dendrito e a entrada, o corpo celular (ou soma) e o processamento, e o axônio e a saída (FURTADO, 2019).

Na Figura 4, é apresentada uma representação de um neurônio artificial. No neurônio, os sinais de entrada são apresentados às sinapses associadas a um determinado peso. Uma função somatória realiza uma combinação linear desses sinais ponderados por pesos sinápticos e aplica uma função de ativação para produzir um sinal de saída.

Durante o processo de treinamento, os pesos sinápticos da rede neural são ajus-

tados aos poucos com o objetivo de minimizar uma função de custo. A função de custo mede a diferença entre as saídas previstas pela rede e as saídas desejadas. Esse processo é geralmente realizado por meio do algoritmo de retropropagação do erro (do inglês, *back-propagation*), que calcula o gradiente da função de custo em relação aos pesos e propaga esse gradiente de volta pela rede, atualizando os pesos conforme descreveu (FURTADO, 2019).

Uma das principais vantagens das redes neurais artificiais é sua capacidade de aprender representações de alto nível dos dados, extraindo características relevantes de forma automática. Isso é especialmente valioso em problemas complexos, nos quais definir manualmente as características pode ser difícil ou impraticável. Além disso, as RNAs são altamente adaptáveis e podem lidar com uma ampla variedade de tipos de dados, como imagens, texto, áudio, séries temporais, entre outros. Elas são capazes de lidar com dados não estruturados, aprendendo padrões diretamente dos dados brutos (HAYKIN, 2009).

No entanto, as RNAs têm algumas limitações a serem consideradas. Elas podem demandar um grande volume de dados de treinamento para alcançar uma bom desempenho, além de exigir um processamento computacional considerável durante o treinamento. Além disso, as RNAs podem enfrentar o problema do sobreajuste (do inglês, *overfitting*), no qual o modelo se adapta em excesso aos dados de treinamento e não consegue generalizar de forma adequada para novos dados (HAYKIN, 2009).

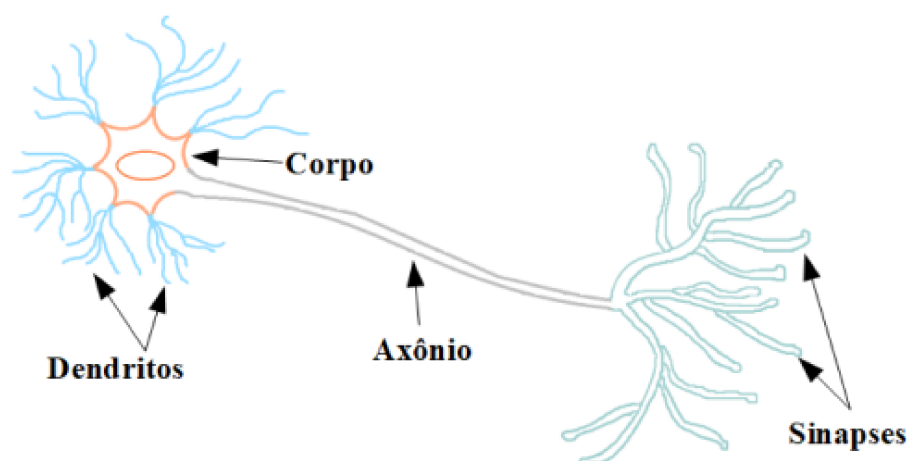


Figura 3 – Representação de um neurônio biológico. Fonte (SILVA, 2019)

2.5 Redes Neurais Convolucionais

As Redes Neurais Convolucionais (do inglês, *convolutional neural network* - CNN) são um tipo especializado de RNAs profundas projetadas para o processamento eficiente de dados estruturados, como imagens e vídeos. Embora tenham sido inicialmente projetadas

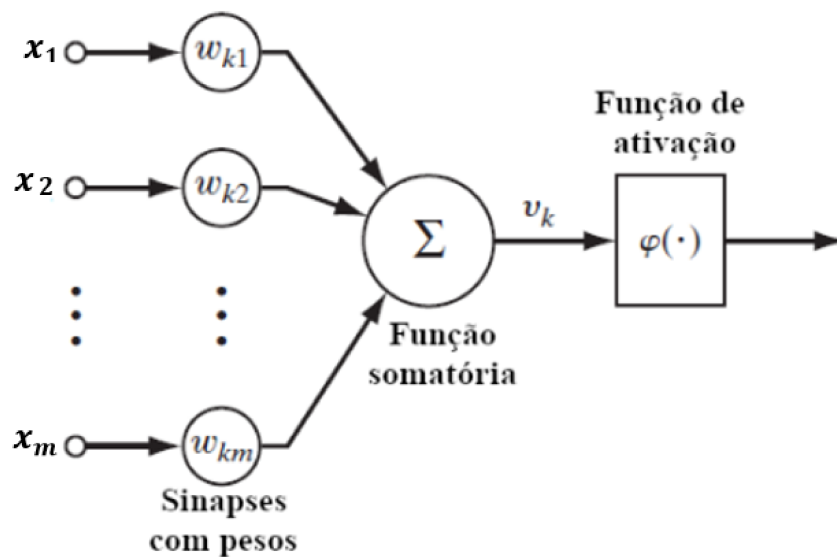


Figura 4 – Representação de um neurônio artificial. Fonte (SILVA, 2019). Adaptado de (HAYKIN, 2009)

para tarefas de visão computacional (LECUN et al., 1998), as CNNs têm mostrado sucesso em diversas outras áreas, como processamento de áudio, processamento de linguagem natural e até mesmo processamento de texto (LI et al., 2021; SHAMSALDIN et al., 2019). Essa abordagem de rede apresenta diversas camadas na composição de um modelo.

A convolução é a operação básica nas CNNs, em que envolve a aplicação de um conjunto de filtros (também conhecidos como *kernels*) sobre a imagem de entrada. Cada filtro é uma pequena matriz de pesos que é deslizada pela imagem de entrada, multiplicando seus valores pelos valores correspondentes da imagem em cada posição. Essa multiplicação é seguida por uma soma, produzindo um valor único para cada posição do filtro (ROSEBROCK, 2017). O resultado é uma nova matriz chamada mapa de características (do inglês, *feature map*), que destaca os padrões relevantes na imagem (ver Figura 5). Conseqüentemente, ao utilizar um maior número de filtros em uma camada, aumenta-se a quantidade de características obtidas pelo processo de convolução executado por esse conjunto de filtros.

Após a convolução, aplica-se a função de ativação ReLU (do inglês, *Rectified Linear Unit*) (HE et al., 2015). Basicamente, essa a função mapeia qualquer valor negativo para zero e mantém qualquer valor positivo inalterado (Figura 6). A função ReLU ajuda a mitigar o problema de desvanecimento do gradiente, que ocorre quando os gradientes se tornam muito pequenos à medida que são propagados para as camadas anteriores da rede durante o treinamento (GLOROT; BORDES; BENGIO, 2011).

Além das operações de convolução, existe um segundo método responsável por reduzir o tamanho dos dados de entrada. Frequentemente aplicada após a convolução e

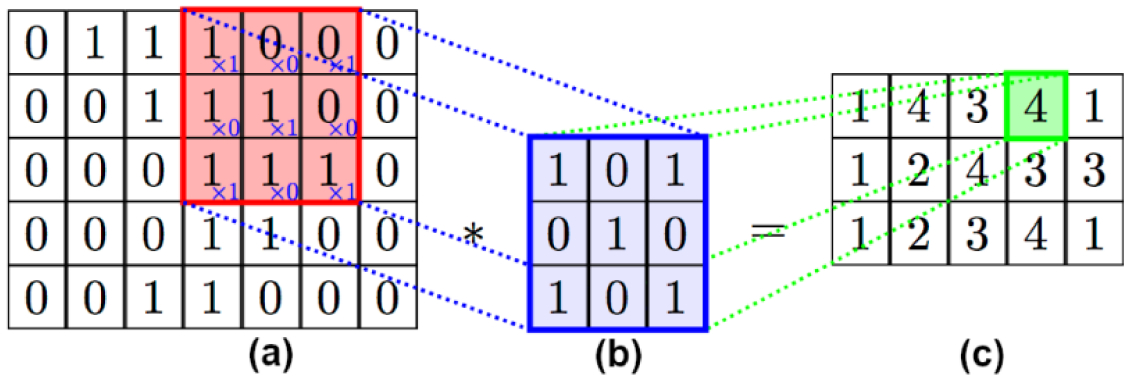


Figura 5 – Exemplo de uma operação de convolução: (a) Representação de uma imagem de tamanho 7×5 pixels, (b) representação de um filtro de convolução de tamanho 3×3 pixels e (c) matriz de característica resultante. Fonte (SILVA, 2019).

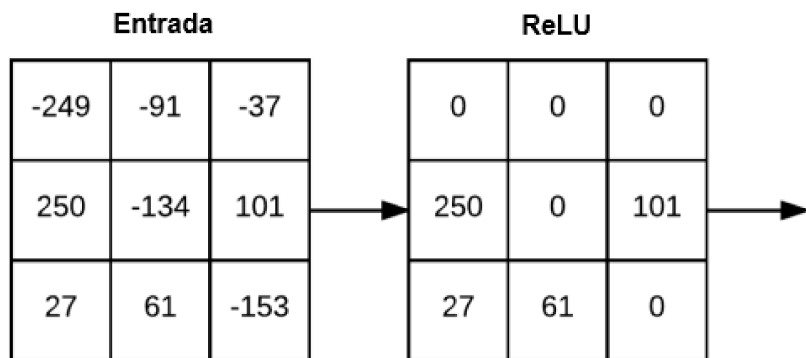


Figura 6 – Exemplo de uma entrada passando por uma ativação ReLU. Adaptado de (ROSEBROCK, 2017).

função de ativação, a operação de *pooling* é responsável por reduzir a dimensionalidade dos *features maps*, extraindo as características mais importantes. Durante a etapa de *pooling*, uma janela deslizante de tamanho fixo (por exemplo, 2×2 pixels) é aplicada na imagem de entrada, movendo-se em incrementos definidos. Para cada posição da janela, é calculado um valor agregado, geralmente o valor máximo (*max pooling*) ou a média (*average pooling*), dependendo do tipo de *pooling* usado. Na Figura 7 são demonstrados exemplos de *max pooling* e *average pooling*.

Para a classificação das imagens, as CNNs utilizam camadas totalmente conectadas. Em uma camada totalmente conectada, cada neurônio recebe entradas de todos os neurônios da camada anterior e produz uma saída que é transmitida para todos os neurônios da camada seguinte. Essa camada é geralmente seguida pela função de ativação *softmax*. Essa função é usada especialmente para problemas de classificação multiclasse, onde o resultado da função *softmax* é interpretado como a probabilidade de um deter-

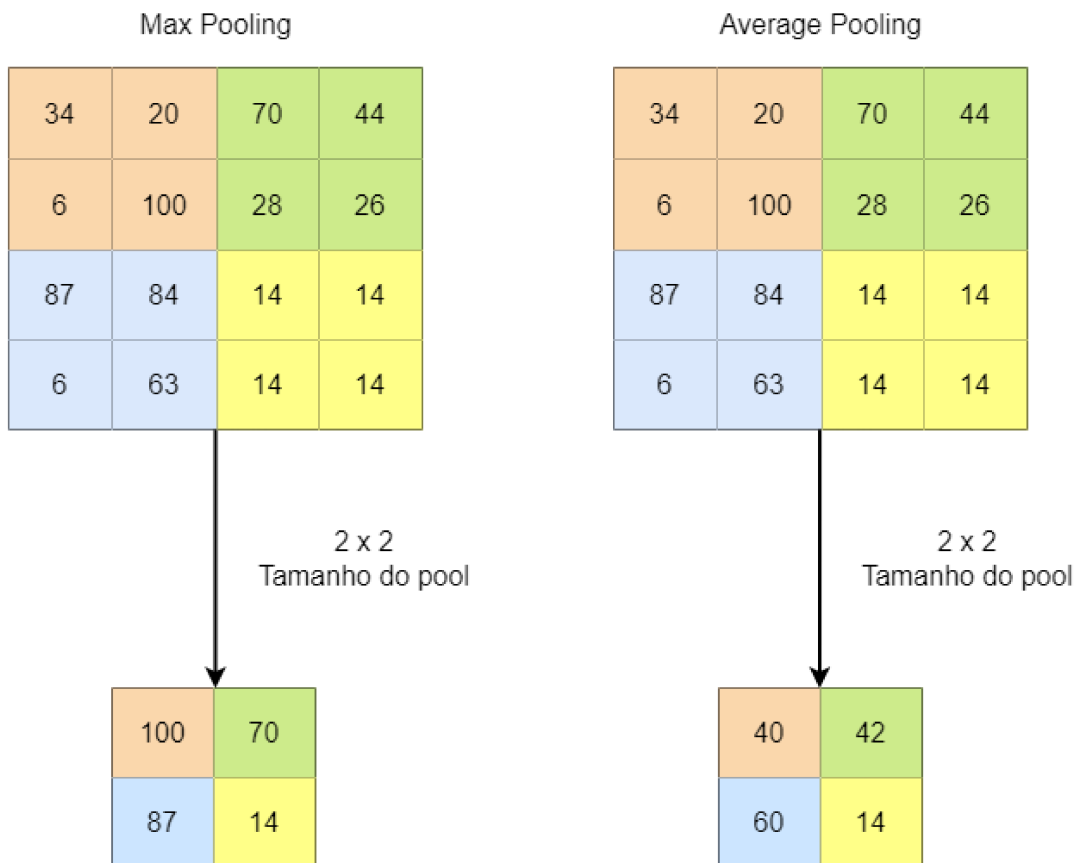


Figura 7 – Exemplo das operações de *max pooling* e *average pooling*. Adaptado de (FREITAS et al., 2021).

minado elemento pertencer a cada uma das classes predefinidas no treinamento (SILVA, 2019; GOODFELLOW; BENGIO; COURVILLE, 2016).

2.6 Aumento de Dados

O aumento de dados é uma técnica amplamente utilizada no aprendizado de máquina para aumentar a quantidade e diversidade dos dados disponíveis para treinamento dos modelos. Essa abordagem visa melhorar a capacidade de generalização dos modelos, tornando-os mais robustos e eficientes na realização de tarefas específicas.

O aumento de dados é especialmente útil quando há restrições de dados de treinamento, o que é comum em muitos cenários reais, como por exemplo, a geração de conjuntos de imagens médicas. Ao aumentar o conjunto de dados existente, é possível evitar o problema da escassez de dados e reduzir o risco de *overfitting* (SHORTEN; KHOSHGOFTAAR, 2019; DAS; DASH, 2022; YING, 2019).

Basicamente, o aumento de dados envolve a aplicação de transformações ou manipulações nos dados existentes para criar novos exemplos de treinamento que sejam semelhantes aos dados originais (KHOSLA; SAINI, 2020; MUMUNI; MUMUNI, 2022).

O aumento de dados por transformações geométricas é uma das técnicas mais populares. Diversas operações geométricas são empregadas nessa etapa, mas neste trabalho foram utilizadas as seguintes transformações:

- a) Espelhamento: consiste em espelhar as imagens horizontal ou verticalmente. Essa técnica não altera a classe ou o conteúdo da imagem, entretanto adiciona variação ao conjunto de dados.
- b) Rotação e corte: envolve a rotação da imagem em diferentes ângulos e o corte subsequente para manter o tamanho original, simulando diferentes pontos de vista.
- c) Ampliação ou redução: consiste em ampliar ou reduzir a escala da imagem, representando variações de distância ou tamanho.
- d) Cisalhamento: essa técnica distorce a imagem ao deslocar as linhas paralelas em ângulos diferentes, adicionando variações de inclinação e perspectiva nas imagens.

Na Figura 8 é apresentado exemplos das transformações citadas anteriormente.

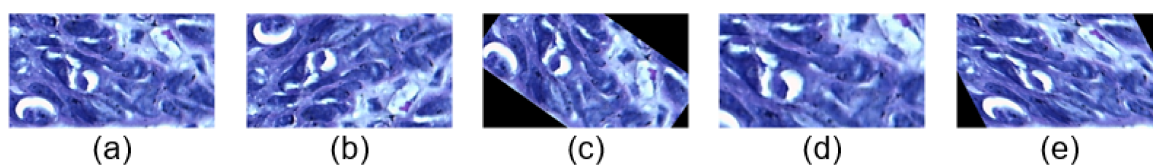


Figura 8 – Exemplo de aumento de dados por transformações geométricas: (a) imagem original, (b) imagem espelhada verticalmente, (c) rotação seguida de corte, (d) ampliação e (e) cisalhamento.

Outra técnica de aumento de dados utilizada como objeto de aumento de dados é o *Mixup*. Esse também é um conceito que foi investigado no estudo. O conceito dessa técnica combina duas ou mais amostras do conjunto de dados de treinamento para criar uma nova amostra aumentada (ZHANG et al., 2017). Isso é alcançado pela combinação linear ponderada das características de entrada e seus rótulos correspondentes.

Conforme demonstrado pelos autores em Zhang et al. (2017) e Zhang et al. (2020), a abordagem *Mixup* oferece várias vantagens no treinamento de modelos de aprendizado profundo (do inglês, *deep learning*), pois incentiva o modelo a aprender características mais robustas e invariantes, mesclando as características de diferentes amostras. Isso ajuda o modelo a generalizar melhor para dados não vistos anteriormente. Uma outra vantagem é a função de atuar como regularizador, ajudando a combater o *overfitting*, ao forçar o modelo a realizar uma interpolação entre as amostras e evitando uma dependência

excessiva em exemplos de treinamento individuais. Isso promove uma maior generalização e melhora o desempenho do modelo em dados não vistos anteriormente. Por fim, o *Mixup* aumenta implicitamente o conjunto de dados de treinamento, aumentando efetivamente seu tamanho e diversidade sem a necessidade de dados rotulados adicionais.

É importante ressaltar que o aumento de dados não substitui a necessidade de um conjunto de dados de treinamento de alta qualidade. É fundamental obter dados de treinamento relevantes e representativos para que o aumento de dados seja eficaz.

2.7 Estado da Arte

A prevenção e o tratamento precoce de lesões da cavidade oral são essenciais para preservar a saúde bucal e evitar complicações graves. O estudo de imagens histológicas desempenha um papel crucial nesse processo, fornecendo informações precisas sobre o tipo de lesão, sua natureza e extensão. Nesta seção são apresentadas as principais contribuições relacionados ao estudo de lesões da cavidade oral.

2.7.1 Estudos de classificação de displasia da cavidade oral

Em [Silva et al. \(2022a\)](#) é apresentada uma nova abordagem para a classificação automatizada de lesões de displasia oral. Essa abordagem combina a segmentação utilizando uma CNN com a classificação baseada em um algoritmo polinomial. Na etapa de segmentação, foi utilizado um modelo de CNN chamado *Mask R-CNN*, treinado com máscaras de núcleos para detectar os objetos de interesse. O estudo utilizou 66 imagens histológicas, catalogadas por três especialistas, como tecidos saudáveis, displasia leve, moderada e severa. Ao todo foram extraídas 296 regiões de interesse, distribuídas igualmente entre as quatro classes. Após a segmentação e o pós-processamento, foram extraídas características morfológicas e não morfológicas. Essas características foram utilizadas como entrada para um classificador polinomial, que foi empregado para distinguir diferentes graus de lesão nas imagens. Os resultados mostraram que a etapa de segmentação obteve acurácias entre 88,92% e 90,35%, enquanto a etapa de classificação apresentou área sob a curva ROC variando de 0,88 a 0,97.

O estudo de [Silva et al. \(2022b\)](#) propôs um método para a classificação da displasia epitelial oral em imagens histopatológicas combinando características profundas e um classificador HOP (do inglês, *Hermite orthogonal polynomial*). A metodologia foi aplicada em um conjunto de imagens construído a partir de seções de tecido de língua de 30 camundongos coradas com H&E. Foram extraídos 296 regiões de interesse classificadas em saudáveis, displasia leve, moderada e severa. As características das imagens foram adquiridas utilizando a técnica de transferência de aprendizado a partir das arquiteturas *ResNet50* e *AlexNet*, que foram previamente treinadas no conjunto de dados *ImageNet*.

Em seguida, as características mais relevantes foram selecionadas usando o algoritmo *ReliefF*, que as ranqueou e as utilizou como entrada para um classificador polinomial. A etapa de classificação alcançou valores de área sob a curva ROC variando de 0,9663 a 0,9800 e os resultados obtidos foram comparados com outros algoritmos de aprendizado de máquina presentes na literatura.

2.7.2 Estudos de classificação de carcinoma escamoso da cavidade oral

O trabalho proposto por [Das et al. \(2018\)](#) utilizou uma abordagem em duas etapas para a segmentação e classificação de imagens de carcinoma escamoso da cavidade oral (do inglês, *oral squamous cell carcinoma* - OSCC). Na primeira etapa, foi utilizado uma rede neural convolucional de 12 camadas para segmentação das camadas constituintes das imagens. Na etapa seguinte, o algoritmo *Random Forests* em conjunto com um filtro *Garbor* foi utilizado para identificar as regiões queratinizadas. No estudo, foi utilizado um conjunto de dados composto por 120 imagens de tecidos com câncer e 6 imagens de tecidos saudáveis. Das imagens totais, 80 foram segmentadas por um especialista e usadas no processo de treinamento da rede neural. A partir de cada imagem segmentada, sub-regiões RGB de tamanho 7×7 *pixels* foram extraídas para o treinamento da rede. Essa extração resultou em um total de um milhão de sub-regiões que foram utilizadas como dados de entrada para o treinamento do modelo. O método proposto alcançou valores de acurácia de 98,42% para a segmentação da camada epitelial e 98,05% para a segmentação das regiões queratinizadas.

Em [Santos et al. \(2022\)](#), é apresentado uma abordagem baseada em redes neurais completamente convolucionais (do inglês, *fully convolutional network* - FCN) para detecção e classificação de regiões de OSCC em imagens histológicas de toda a lâmina (do inglês, *whole slide image* - WSI). Os autores propuseram a implementação de um detector de tecido baseado em cores como etapa de pré-processamento. Um total de 1050 fragmentos de imagem de tamanho 640×640 *pixels* foram extraídos aleatoriamente de 15 WSIs. Para treinar e validar o método, um conjunto de dados WSI proprietário de tecidos corados com H&E de regiões OSCC foi empregado. O método foi testado nesta e em outras bases de imagens disponíveis na literatura, incluindo uma base de WSIs de outro tipo de câncer, alcançando 90% de F1-score na base proprietária e 83% de F1-score para as demais bases de imagens de domínio público, mostrando-se robusto e com potencial para criação de ferramentas de apoio ao diagnóstico em casos de OSCC.

2.8 Considerações Finais

Os trabalhos apresentados destacam a importância e o potencial das abordagens computacionais na análise de imagens histológicas para diagnóstico de lesões orais, bem

como a detecção de núcleos celulares. Embora cada trabalho tenha suas próprias limitações e áreas para melhoria, eles contribuem para o avanço do conhecimento na área de histopatologia oral. Os estudos mostram que essa área ainda existem muitos desafios em aberto e novas pesquisas podem contribuir para o avanço de métodos computacionais empregados no diagnóstico de lesões da cavidade oral.

No trabalho desenvolvido por [Silva et al. \(2022a\)](#) embora tenha alcançado resultados relevantes, foram observadas algumas limitações, como regiões de falsos positivos e valores de classificação mais baixos em certos grupos. No entanto, vale ressaltar que essa metodologia representa um avanço, uma vez que não havia estudos anteriores sobre a quantificação de lesões displásicas.

Em [Silva et al. \(2022b\)](#), os resultados mostraram que as características extraídas utilizando o modelo *ResNet50* foram mais eficazes, e o classificador HOP apresentou os melhores resultados. Essa metodologia contribuiu para o estado-da-arte na classificação de lesões displásicas, oferecendo *insights* valiosos para patologistas na análise histológica.

A abordagem proposta por [Das et al. \(2018\)](#) mostrou-se suficientemente boa para extrair regiões epiteliais, subepiteliais e de queratina, com desempenho comparável ou melhor em relação à literatura existente na época, se mostrando útil para triagem de lesões pré-cancerosas e classificação de carcinoma de células escamosas de cavidade oral, auxiliando os médicos em diagnósticos rápidos e livres de viés.

Por fim, [Santos et al. \(2022\)](#) abordou a segmentação de imagens histológicas coradas com H&E para apoiar a tomada de decisões dos patologistas em casos de câncer da cavidade oral. A proposta envolveu o uso de um método baseado em redes neurais totalmente convolucionais para segmentar regiões tumorais e núcleos em imagens histopatológicas. Embora tenha enfrentado desafios devido à falta de dados de treinamento rotulados e à necessidade de estratégias de aumento de dados, o método proposto mostrou resultados promissores. Além disso, a criação de um novo conjunto de dados rotulados de carcinoma escamoso da cavidade oral contribuiu para o avanço da pesquisa nessa área.

Em conjunto, esses trabalhos mostram o potencial das técnicas computacionais, como redes neurais convolucionais e classificadores, para auxiliar os patologistas na análise histológica de lesões orais, permitindo diagnósticos mais rápidos e precisos. No entanto, há espaço para melhorias contínuas, como a exploração de novas técnicas de pré-processamento, combinação de características manuais e profundas, ampliação de dados e utilização de conjuntos de dados mais abrangentes. O avanço nessas áreas pode levar a ferramentas ainda mais eficazes e confiáveis para o diagnóstico de câncer oral e outras condições relacionadas.

3 Metodologia

3.1 Visão Geral

Conjuntos 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 CNNs 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 investigadas abordagens tradicionais de aumento de dados baseadas em representações geométricas. Na Figura 9 é apresentado um diagrama das principais etapas do método proposto e investigadas neste trabalho.

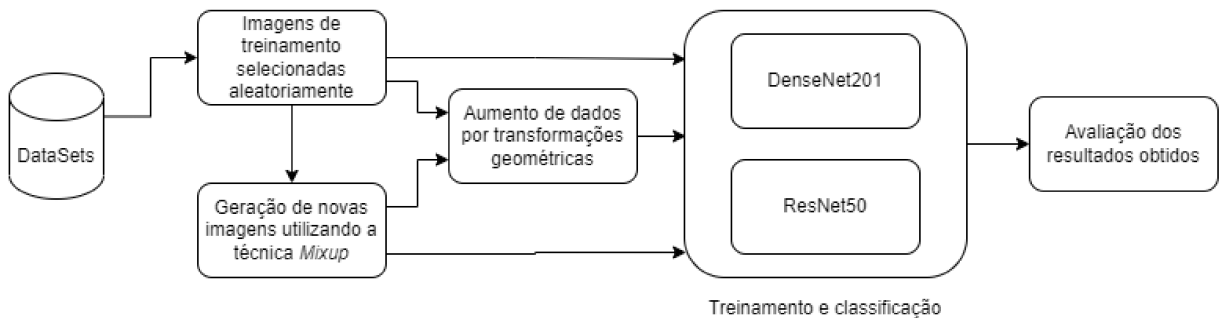


Figura 9 – Visão geral das principais etapas de classificação de lesões da cavidade bucal.

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 conjuntos de dados para avaliar e gerar imagens aumentadas, sendo um composto por lesões displásicas e outro composto por imagens de OSCC.

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 400×, 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 296 ROIs com tamanho de 450×250 pixels, divididas igualmente para cada classe de tecido.

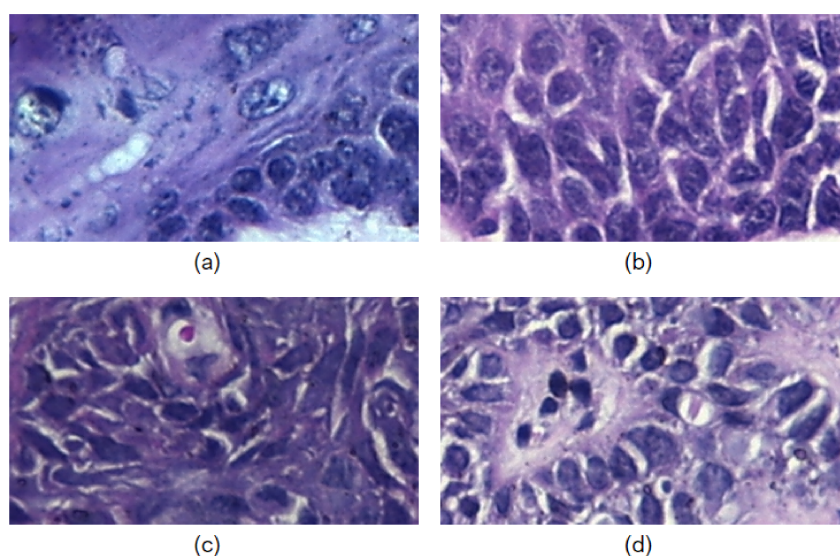


Figura 10 – Exemplos de tecido histológico da cavidade bucal: (a) saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa.

3.2.2 Carcinoma escamoso da cavidade oral - OSCC

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 $100\times$, contém 89 imagens histopatológicas com o epitélio normal da cavidade oral e 439 imagens de OSCC. O segundo, com magnificação de $400\times$, contém 201 imagens histopatológicas com o epitélio normal da cavidade oral e 495 imagens de OSCC.

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 $400\times$ (ver Figura 11), 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 12).

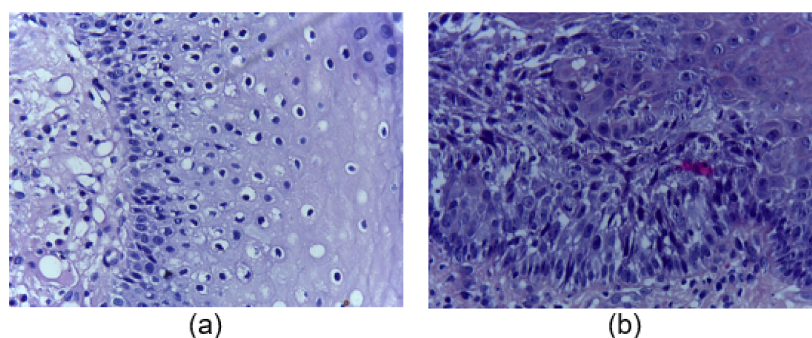


Figura 11 – Exemplo das imagens originais de tecido histopatológico da cavidade oral: (a) saudável, (b) OSCC.

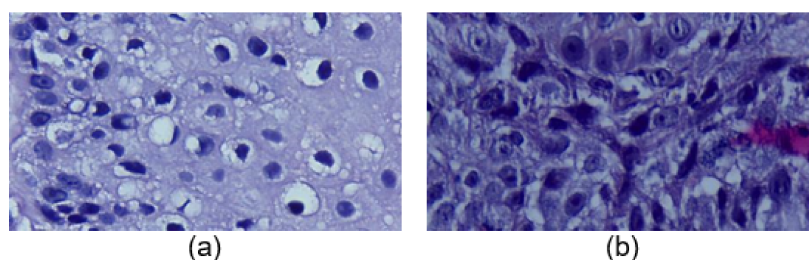


Figura 12 – Exemplo das imagens de tecido histopatológico da cavidade oral: (a) saudável, (b) OSCC, redimensionadas a partir do centro das imagens da Figura 11.

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:

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

onde (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 13 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 14).

3.4 Aumento de dados por transformações geométricas

O aumento de dados 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 (KHOSLA; SAINI, 2020; MUMUNI; MUMUNI, 2022). Neste trabalho, as seguintes transformações foram empregadas: rotação aleatória, reflexão aleatória, cisalhamento horizontal e redimensionamento aleatório.

No caso do programa *Matlab*, a função denominada *imageDataAugmenter* foi empregada para realizar essas operações. Na Figura 15 é possível observar as etapas do

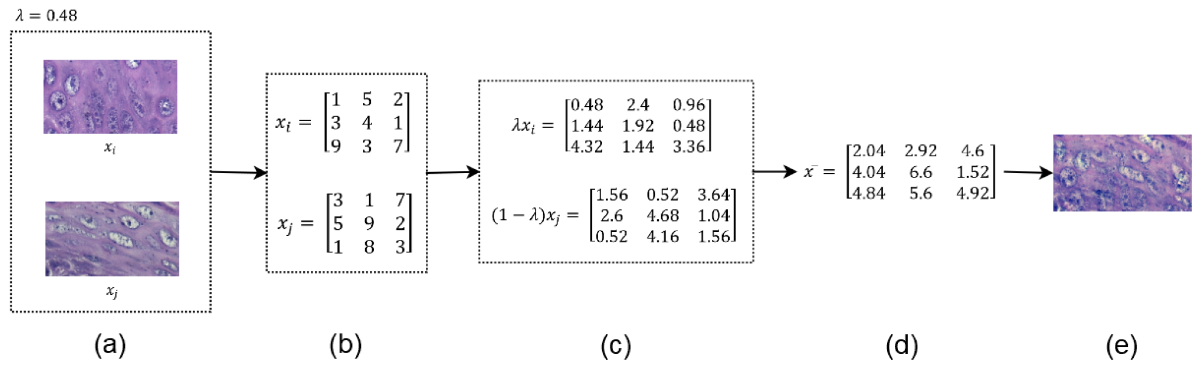


Figura 13 – Exemplo das etapas da criação de imagens sintéticas utilizando *Mixup*: (a) Imagens obtidas aleatoriamente, (b) Vetorização das imagens, (c) Matrizes proporcionais de x_i e x_j , (d) Matriz \tilde{x} resultante da soma das matrizes proporcionais e (e) Imagem equivalente a \tilde{x} .

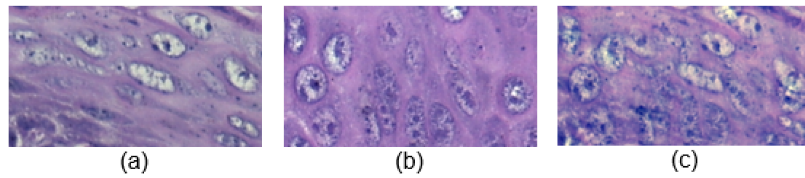


Figura 14 – As imagens (a) e (b) são amostras de tecidos com displasia leve, (c) imagem construída por meio da combinação de (a) e (b).

processo de construção das imagens transformadas pelo emprego da função *imageDataAugmenter*. 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

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 neste 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.

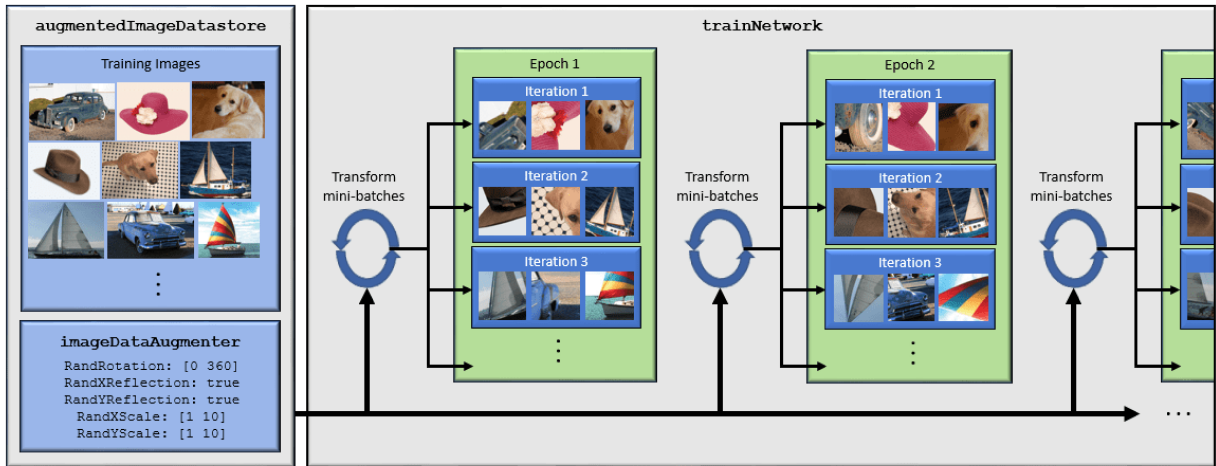


Figura 15 – Exemplo de transformação dos dados de treinamento para cada época. Fonte (Matlab documentation, 2022)

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 16).

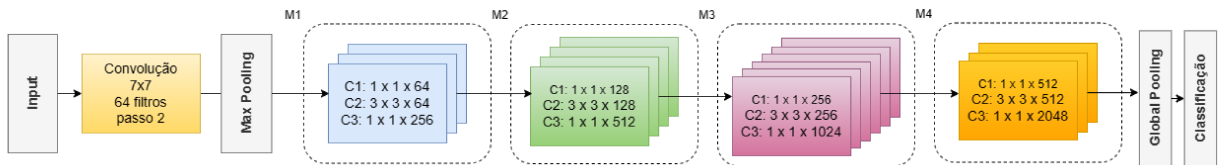


Figura 16 – Arquitetura da rede neural *Resnet50*. M1, M2, M3, M4 representam os módulos compostos por blocos residuais. Adaptado de (SILVA, 2019).

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 17).

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.

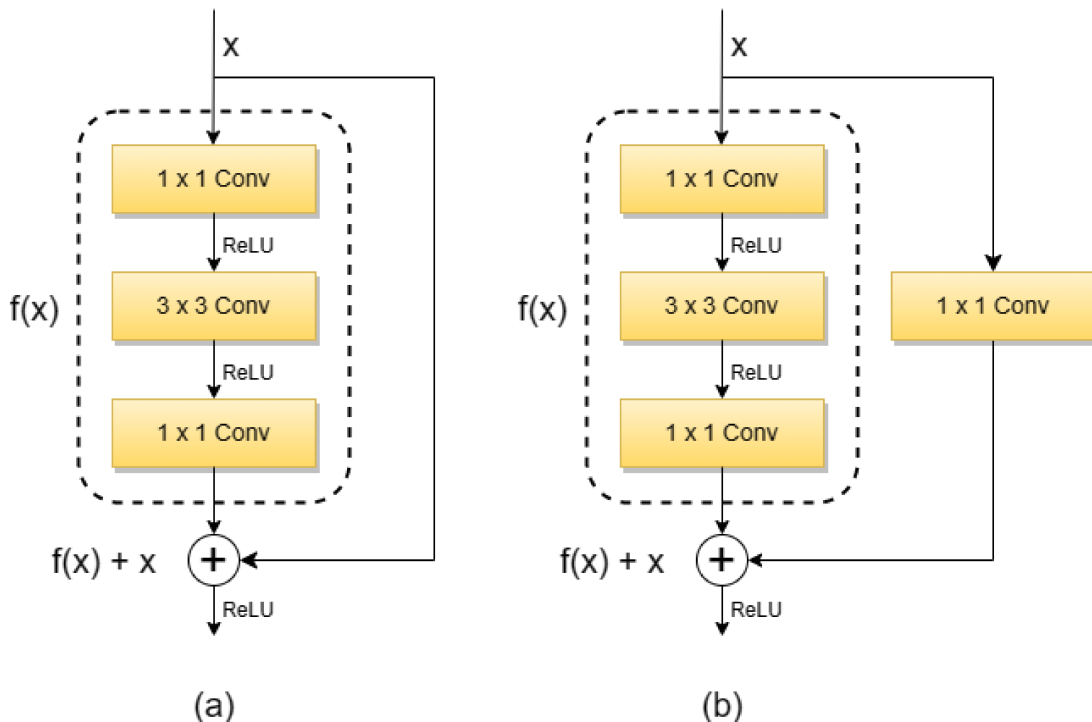


Figura 17 – Blocos residuais presentes na arquitetura *ResNet50*: (a) bloco residual sem convolução 1×1 e (b) bloco residual com convolução 1×1 .

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 18). 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 19.

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*.

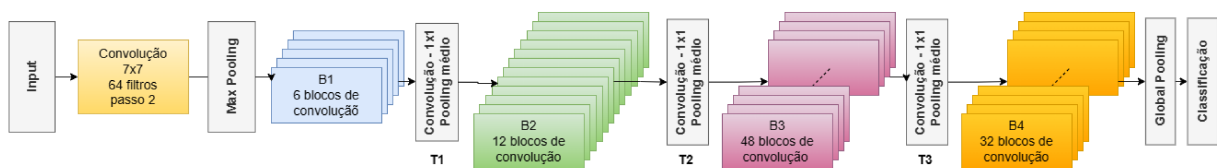


Figura 18 – Arquitetura da rede neural *DenseNet201*. B1, B2, B3, B4 representam os blocos densos. T1, T2, T3 representam as camadas de transição.

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 au-

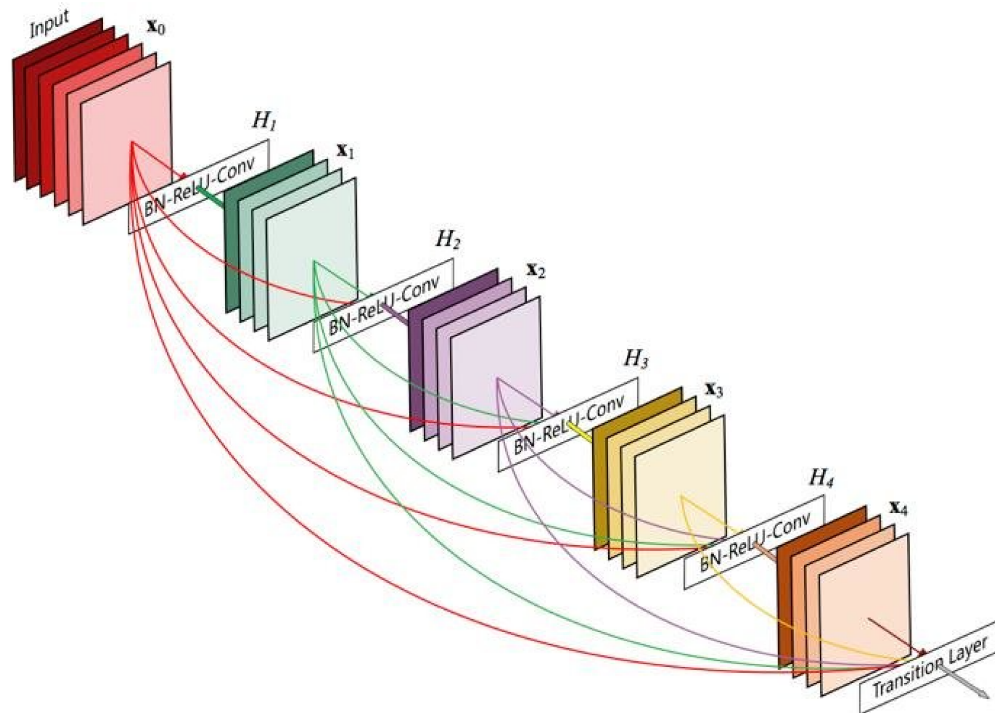


Figura 19 – Representação de um bloco denso, composto por 5 blocos de convolução, onde cada bloco de convolução tem como entrada a saída dos blocos anteriores. Fonte (HUANG et al., 2017)

mento 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 empregada 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 125×225 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 (saudável, leve, moderada e severa no conjunto de displasia da cavidade oral e saudável e não saudável para o conjunto de carcinoma escamoso da cavidade oral).

Os modelos foram treinados utilizando 50 épocas, com taxa inicial de aprendizagem de 0,0003, mini-lote 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 20 é apresentado uma matriz com as informações de cada tipo.

Para esse estudo foram empregadas as métricas de acurácia, precisão, sensibilidade, especificidade e *F1-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.

$$Acurácia = \frac{tp + tn}{tp + tn + fp + fn}. \quad (3.1)$$

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.

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

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.

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

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

Matriz de confusão

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

Figura 20 – Exemplo de uma matriz de confusão para análise das imagens histológicas

negativos.

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

A *F1-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 *F1-Score* será baixa. Desse modo, para que *F1-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.

$$F1 - Score = 2 * \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}}. \quad (3.5)$$

4 Resultados

Neste capítulo são apresentados os resultados obtidos ao utilizar as diferentes técnicas de aumento de dados para as diversas arquiteturas investigadas. Diferentes abordagens foram investigadas com objetivo de avaliar a influência sobre a classificação de lesões da cavidade oral.

4.1 Composição de Imagens com o uso da Técnica *Mixup*

Nas Figuras 21 e 22 são mostrados os resultados da aplicação da técnica de aumento de dados *Mixup*. Nessa Figura, as colunas (a) e (b) representam diferentes imagens de para uma determinada lesão e a coluna (c) traz o resultado da operação de *Mixup*. A escolha das imagens ocorreu de forma aleatória, porém dentro de uma mesma classe, em que a imagem resultante sempre é gerada pela composição de duas outras imagens da mesma classe. Ao aplicar a técnica, a quantidade de amostras de treinamento de displasia saltou de 236 imagens para 472 imagens e as amostras de treinamento de OSCC saltou de 557 imagens para 1114 imagens. É possível observar nas imagens resultantes (Figuras 21(c) e 22(c)), características das imagens originais utilizadas na composição (Figuras 21(a), 21(b) e 22(a), 22(b)). Também nota-se que há perda de nitidez de características, problema que se agrava à medida que a proporção que cada imagem representa na composição se torna mais desbalanceada. Isso mostra que o uso dessa abordagem pode auxiliar na composição de uma nova imagem, mas o uso sem controle das proporcionalidades pode gerar uma imagem com baixa nitidez.

4.2 Composição de Imagens com o uso de Transformações Geométricas

Na Figura 23, são apresentados exemplos das imagens geradas aplicando transformações geométricas em amostras de treinamento. As operações empregadas foram escolhidas de modo que a imagem não perdesse características relevantes à classificação, porém adicionasse variação ao conjunto de treinamento. As imagens foram submetidas a rotações aleatórias em ângulos de até 45 graus, reflexão aleatória vertical e horizontal, cisalhamento em ângulos de até 30 graus e redimensionamento entre 90% e 110% da imagem original. As imagens de treinamento foram divididas em mini-lotes escolhidos aleatoriamente e então submetidas à uma combinação randômica dessas operações, ou seja, uma imagem sofria mais de uma transformação geométrica ao mesmo tempo.

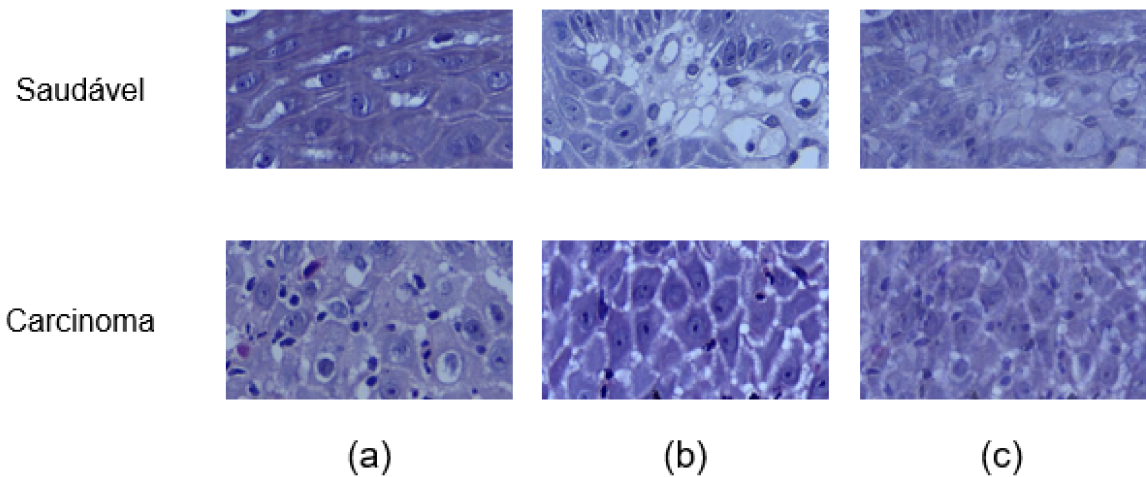


Figura 21 – Imagens de carcinoma oral de células escamosas. As imagens das colunas (a) e (b) foram escolhidas aleatoriamente, a coluna (c) representa a combinação resultante das imagens de (a) e (b).

4.3 Avaliação da Classificação das Lesões com os Modelos de CNNs

Inicialmente, as etapas de treinamentos devem ser realizadas sob as mesmas condições em relação aos dados. No entanto, a escolha do mini-lote utilizado em cada época ocorre de forma aleatória, o que pode influenciar o resultado de um modelo investigado. Além disso, as estratégias que utilizam o aumento de dados por transformações geométricas também exploram a aleatoriedade na elaboração das imagens. Dessa forma, o estudo comparativo das arquiteturas foi realizado com cinco treinamentos para cada um dos modelos. 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. As Tabelas 1 e 2 apresentam os valores de acurácia obtidos em cada uma das execuções, por arquitetura e conjunto de dados.

Os resultados mostram que ocorre uma variação nos dados devido aos processos randômicos empregados nos experimentos. Na investigação multiclasse com a base de displasia da cavidade oral, nota-se que os melhores resultados ocorrem com o modelo DenseNet201 e a composição entre as duas técnicas de aumento de dados com acurácia média de $85,6 \pm 0,05$. Esse mesmo comportamento também foi observado com a base carcinoma escamoso da cavidade oral, em que a mesma arquitetura proporcionou resultados médio de acurácia de $88,6 \pm 0,02$.

As Figuras 24, 25 e 26 apresentam os resultados das arquiteturas *ResNet50* e

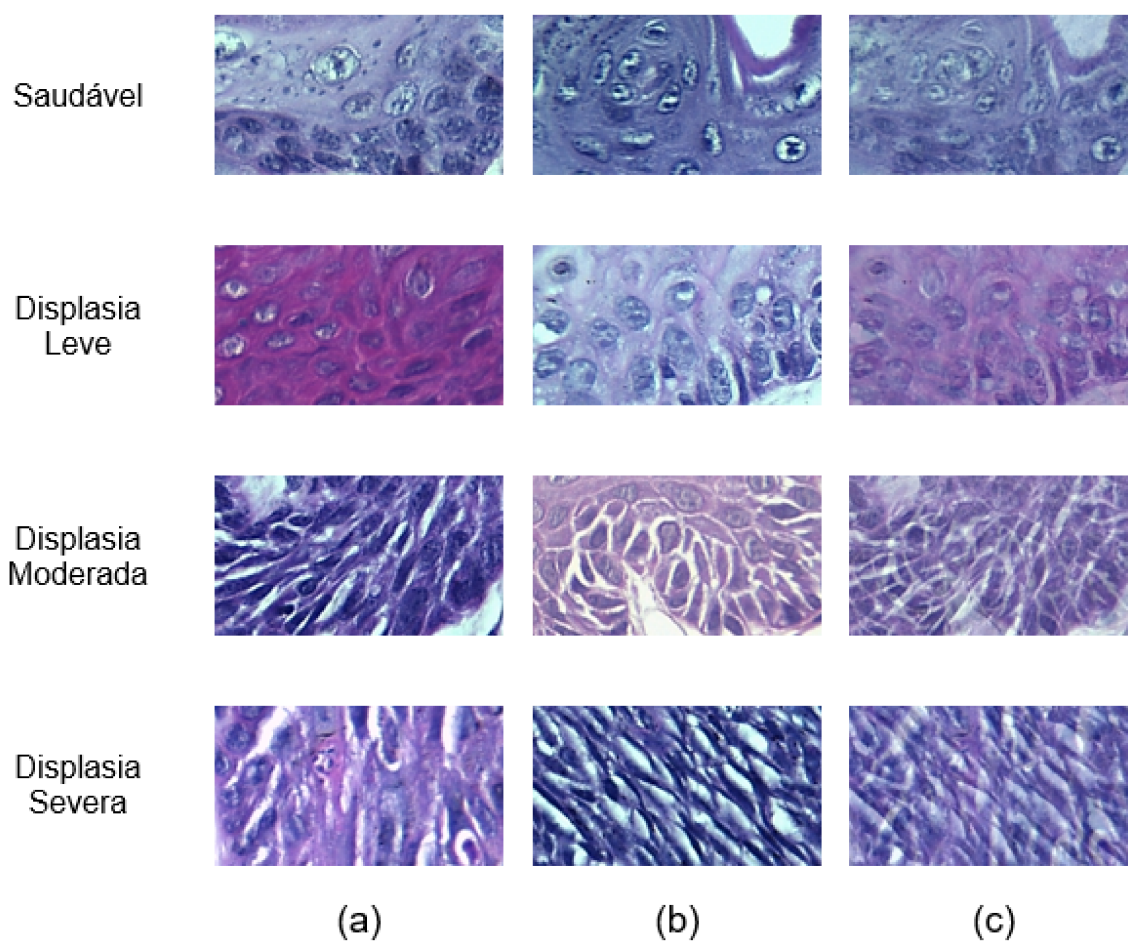


Figura 22 – Imagens de displasia oral. As imagens das colunas (a) e (b) foram escolhidas aleatoriamente, a coluna (c) representa a combinação resultante de (a) e (b).

DenseNet201 nas etapas de testes para os 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. O conjunto da base de carcinoma escamoso da cavidade oral possui dados desbalanceados e por esse motivo a acurácia pode não trazer valores efetivos dos modelos. Na Figura 26 são apresentados os resultados da classificação dos dados de carcinoma utilizando a métrica F1-Score. É possível observar que as curvas obtidas são semelhantes às apresentadas na Figura 25 o que indica que os valores de acurácia obtidos não sofreram influência de uma determinada classe em detrimento da outra.

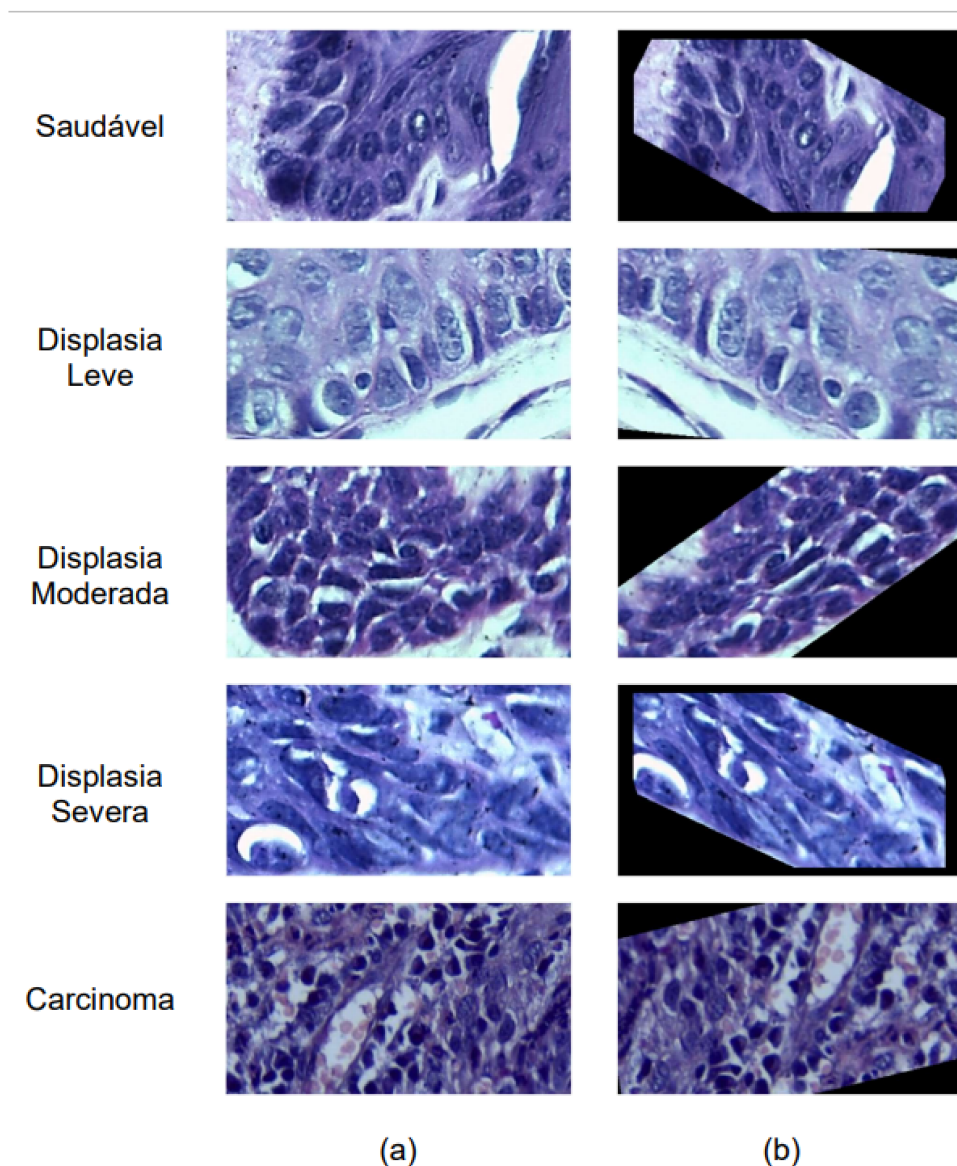


Figura 23 – Imagens aumentadas utilizando transformações geométricas: (a) imagens originais e (b) imagens transformadas através da combinação aleatória de operações geométricas.

4.4 Estudo das Melhores Estratégias Empregadas nos Bancos de Imagens

Com os resultados obtidos na etapa anterior, nessa fase foram escolhidos os melhores valores de acurácia entre as execuções (1-5) para os dois bancos de imagens. Os valores mais altos em relação a métrica de acurácia foram selecionados para estudo comparativo em relação a função de perda para cada tipo de aumento de dados, assim como, o desempenho de cada uma das classes das lesões. No caso do banco de displasia, foram utilizadas 32 imagens para avaliação dos modelos treinados, divididas igualmente entre quatro classes: tecido saudável, displasia leve, displasia moderada e displasia severa. Estas

Tabela 1 – 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 | Aumento de dados geométricos e <i>Mixup</i> | 84,4 | 90,6 | 87,5 | 87,5 | 75,0 | 85,0 ± 0,05 |
| | <i>Mixup</i> | 81,2 | 84,4 | 75,0 | 81,2 | 84,4 | 81,2 ± 0,03 |
| | Aumento de dados geométricos | 78,1 | 75,0 | 84,4 | 81,2 | 81,2 | 80,0 ± 0,03 |
| | Sem Aum. de dados geométricos ou <i>Mixup</i> | 81,2 | 78,1 | 78,1 | 84,4 | 84,4 | 81,2 ± 0,03 |
| DenseNet201 | Aumento de dados geométricos e <i>Mixup</i> | 84,4 | 84,4 | 90,6 | 90,6 | 78,1 | 85,6 ± 0,05 |
| | <i>Mixup</i> | 81,2 | 90,6 | 81,2 | 81,2 | 84,4 | 83,7 ± 0,04 |
| | Aumento de dados geométricos | 81,2 | 81,2 | 81,2 | 81,2 | 84,4 | 81,8 ± 0,01 |
| | Sem Aum. de dados geométricos ou <i>Mixup</i> | 81,2 | 81,2 | 87,5 | 81,2 | 87,5 | 83,7 ± 0,03 |

Tabela 2 – 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 | Aumento de dados geométricos e <i>Mixup</i> | 87,1 | 87,1 | 80,0 | 88,6 | 82,9 | 85,1 ± 0,03 |
| | <i>Mixup</i> | 77,1 | 81,4 | 74,3 | 75,7 | 80,0 | 77,7 ± 0,03 |
| | Aumento de dados geométricos | 82,9 | 81,4 | 81,4 | 80,0 | 84,3 | 82,0 ± 0,01 |
| | Sem Aum. de dados geométricos ou <i>Mixup</i> | 81,4 | 81,4 | 78,6 | 78,6 | 78,6 | 79,7 ± 0,01 |
| DenseNet201 | Aumento de dados geométricos e <i>Mixup</i> | 85,7 | 87,1 | 88,6 | 88,6 | 92,9 | 88,6 ± 0,02 |
| | <i>Mixup</i> | 82,9 | 80,0 | 84,3 | 75,7 | 75,7 | 79,7 ± 0,04 |
| | Aumento de dados geométricos | 90,0 | 85,7 | 82,9 | 85,7 | 87,1 | 86,3 ± 0,02 |
| | Sem Aum. de dados geométricos ou <i>Mixup</i> | 84,3 | 82,9 | 81,4 | 84,3 | 81,4 | 82,9 ± 0,01 |

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 3 e 4 são mostradas 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 associada a outra

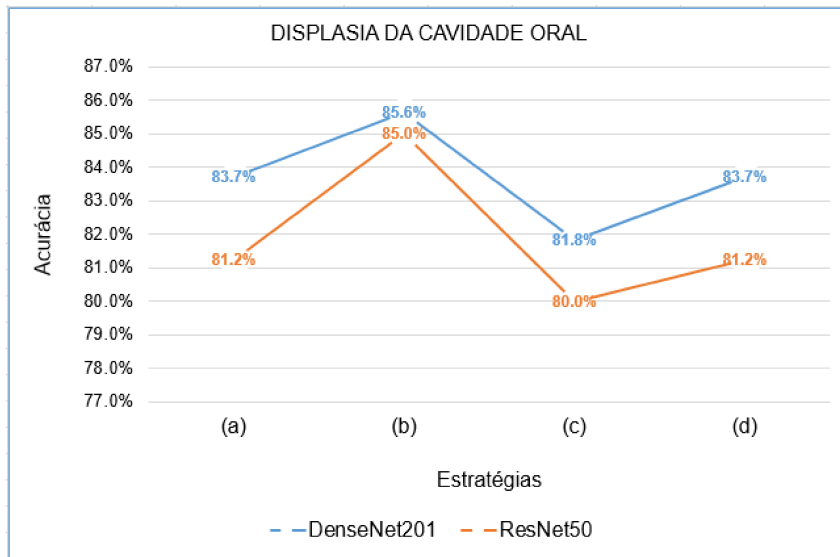


Figura 24 – Acurácia média obtida para o conjunto de dados de displasia da cavidade oral, onde (a) utilizando somente a técnica *Mixup*, (b) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

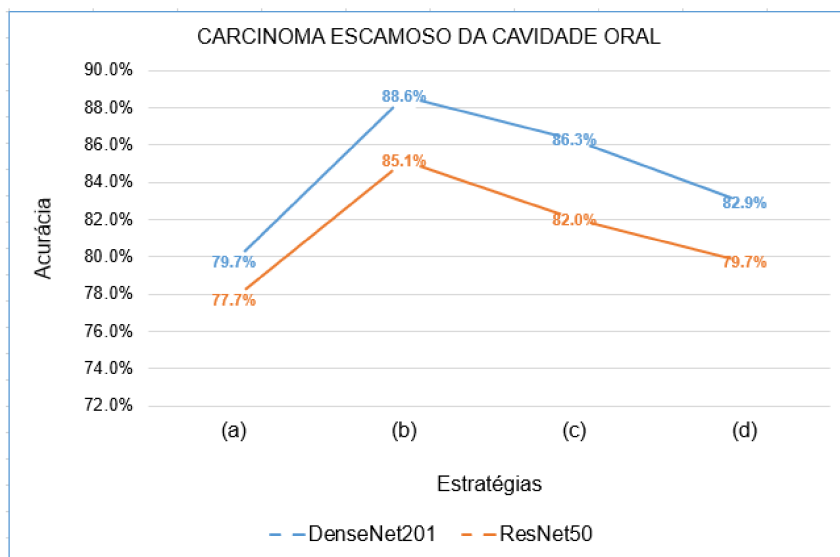


Figura 25 – Acurácia média obtida para o conjunto de dados de carcinoma escamoso da cavidade oral, onde (a) utilizando somente a técnica *Mixup*, (b) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

técnica de aumento de dados ou não. Para o modelo *ResNet50*, a estratégia adotada que empregou a técnica *Mixup* combinada com aumento de dados geométricos apresentou o melhor resultado. No entanto, o uso da técnica de forma isolada não apresentou resultados relevantes às demais estratégias (utilizando somente aumento de dados geométricos e utilizando somente as imagens originais).

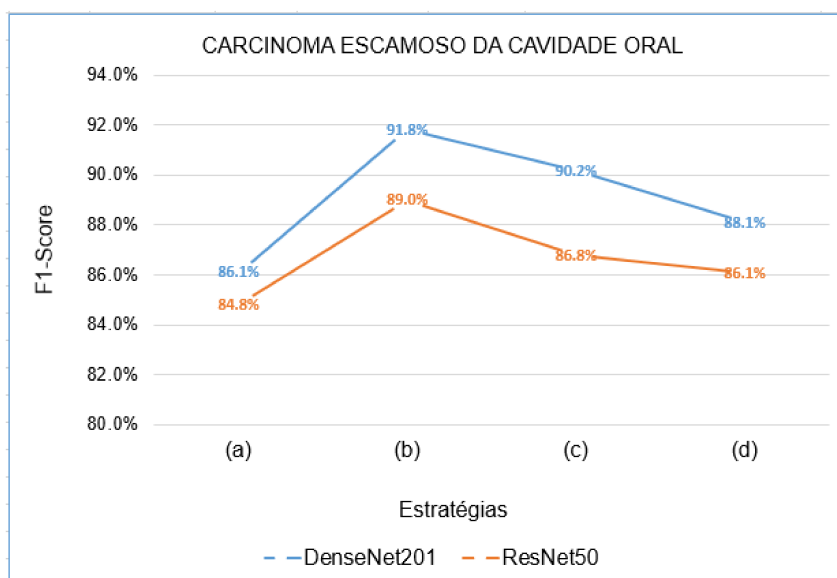


Figura 26 – F1-Score médio obtido para o conjunto de dados de carcinoma escamoso da cavidade oral, onde (a) utilizando somente a técnica *Mixup*, (b) utilizando a combinação das técnicas *Mixup* e aumento de dados geométricos, (c) utilizando aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

Tabela 3 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral com arquitetura *ResNet50*.

| Displasia da cavidade oral - <i>ResNet50</i> (%) | | | | | |
|--|----------|----------------|----------|---------------|----------|
| Estratégias | Classes | Especificidade | Precisão | Sensibilidade | Acurácia |
| Aumento de dados geométricos e <i>Mixup</i> | 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 Aum. de dados geométricos ou <i>Mixup</i> | 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 | |

Nas Figuras 27 e 28 são apresentados as taxas de perda para os modelos *ResNet50* e *DenseNet201*, respectivamente. As taxas correspondem ao treinamento realizado com os dados de displasia oral nas quatro estratégias empregadas neste trabalho. O eixo X representa a quantidade de iterações e o eixo Y o valor da taxa. Os gráficos mostram que

Tabela 4 – Resultados (%) das classificações das imagens de teste de displasia da cavidade oral com arquitetura *DenseNet201*.

| Displasia da cavidade oral - <i>DenseNet201</i> (%) | | | | | |
|---|----------|----------------|----------|---------------|----------|
| Estratégias | Classes | Especificidade | Precisão | Sensibilidade | Acurácia |
| Aumento de dados geométricos e <i>Mixup</i> | 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 Aum. de dados geométricos ou <i>Mixup</i> | 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 | |

as estratégias que empregaram somente a técnica *Mixup* obtiveram resultados semelhantes às estratégias que não utilizaram nenhuma técnica de aumento de dados para as duas arquiteturas, *ResNet50* e *DenseNet201*. As curvas de validação se afastaram das curvas de treinamento, indicando que não houve generalização da classificação, mas sim um sobreajuste das amostras de teste. Ao contrário, as estratégias que empregaram a técnica *Mixup* combinadas com aumento de dados geométricos apresentaram boa generalização, o que justifica as acurácias obtidas nas Tabelas 3 e 4.

Os resultados da etapa de testes para o conjunto de dados de OSCC são apresentados na Tabela 5. Os modelos foram submetidos aos testes aplicando as estratégias avaliadas neste estudo utilizando uma amostra de 70 imagens, sendo 20 imagens de tecidos saudáveis e 50 imagens de tecidos com carcinoma. Trata-se de um conjunto de dados binário e para fins de esclarecimento considerou-se como resultado positivo a presença de carcinoma. Neste experimento é possível observar que a estratégia que fez uso da técnica *Mixup* combinada ao aumento de dados geométricos apresentou resultados melhores do que as estratégias que fazem usos isolados dessas técnicas. Nota-se que tanto no modelo *ResNet50* quanto no modelo *DenseNet201* houve uma melhora não só na capacidade de classificar corretamente tecidos não saudáveis, como também de classificar corretamente amostras saudáveis, o que pode ser comprovado ao analisar, respectivamente, a métrica F1-Score e especificidade. No entanto, observa-se que o uso isolado da técnica *Mixup* não apresentou resultados melhores do que não usar nenhuma técnica de aumento de dados. Ao analisar as taxas de perda do modelo *ResNet50* (Figura 29) e do modelo *DenseNet201*

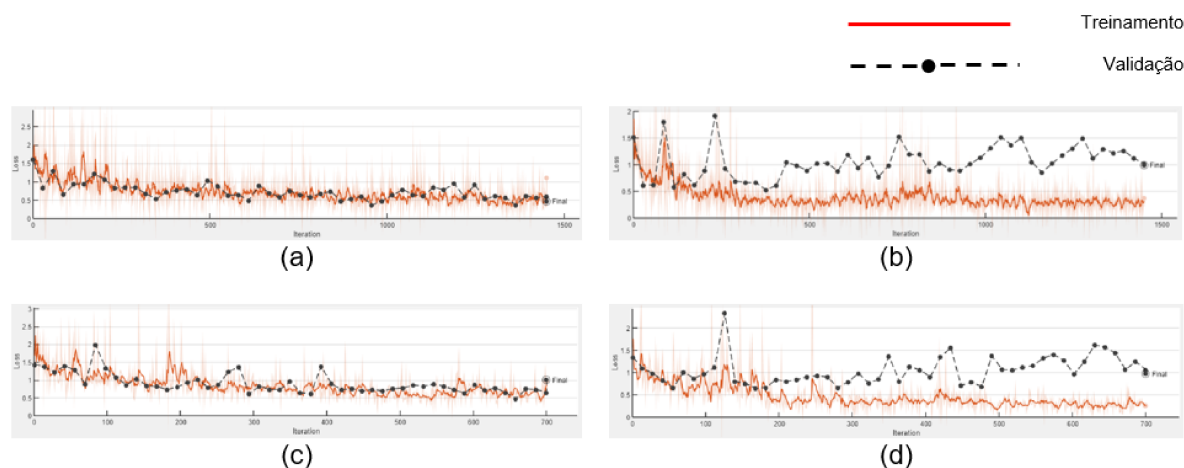


Figura 27 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de displasia oral e a arquitetura *ResNet50*, onde (a) utilizando *Mixup* e aumento de dados geométricos, (b) utilizando somente *Mixup*, (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

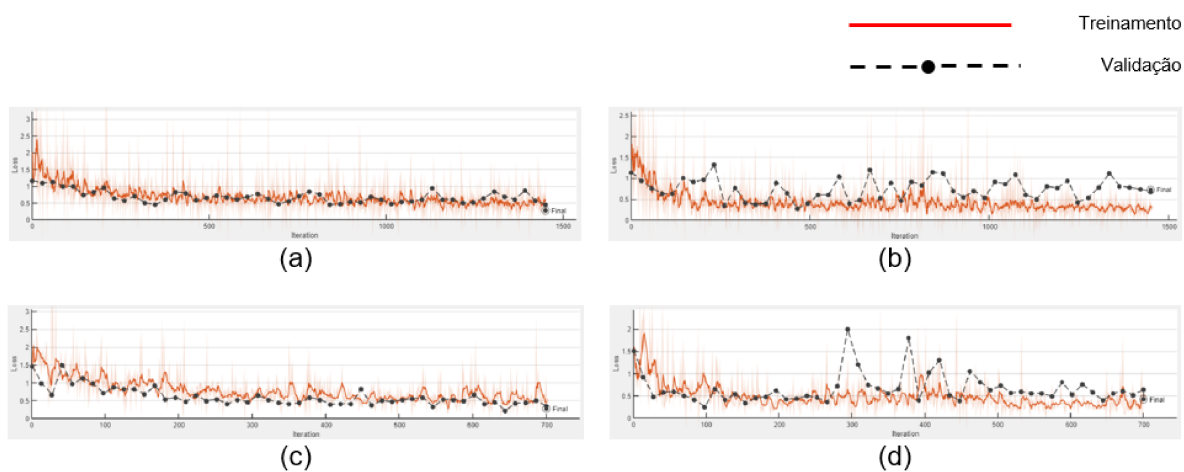


Figura 28 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de displasia oral e a arquitetura *DenseNet201*, onde (a) utilizando *Mixup* e aumento de dados geométricos, (b) utilizando somente *Mixup*, (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

(Figura 30) percebe-se que o uso apenas da técnica *Mixup* não garantiu resultados relevantes. Observe que as taxas de perda das amostras de validação foram maiores que as taxas de perda das amostras de treinamento, indicando que não houve generalização, mas sim um possível *overfitting*.

Tabela 5 – Resultados obtidos ao classificar as imagens de teste de carcinoma escamoso da cavidade oral.

| Carcinoma escamosos da cavidade oral (%) | | | | | | |
|--|---|----------|----------|---------|----------|----------|
| Modelos | Estratégias | Especif. | Precisão | Sensib. | Acurácia | F1-Score |
| <i>ResNet50</i> | Aum. de dados geométricos e <i>Mixup</i> | 90,0 | 95,7 | 88,0 | 88,6 | 91,7 |
| | <i>Mixup</i> | 70,0 | 87,8 | 86,0 | 81,4 | 86,9 |
| | Aum. de dados geométricos | 85,0 | 93,3 | 84,0 | 84,3 | 88,4 |
| | Sem <i>Mixup</i> ou Aum. de dados geométricos | 75,0 | 89,4 | 84,0 | 81,4 | 86,6 |
| <i>DenseNet201</i> | Aum. de dados geométricos e <i>Mixup</i> | 90,0 | 95,9 | 94,0 | 92,9 | 94,9 |
| | <i>Mixup</i> | 80,0 | 91,5 | 86,0 | 84,3 | 88,7 |
| | Aum. de dados geométricos | 85,0 | 93,9 | 92,0 | 90,0 | 92,9 |
| | Sem <i>Mixup</i> ou Aum. de dados geométricos | 75,0 | 89,8 | 88,0 | 84,3 | 88,9 |

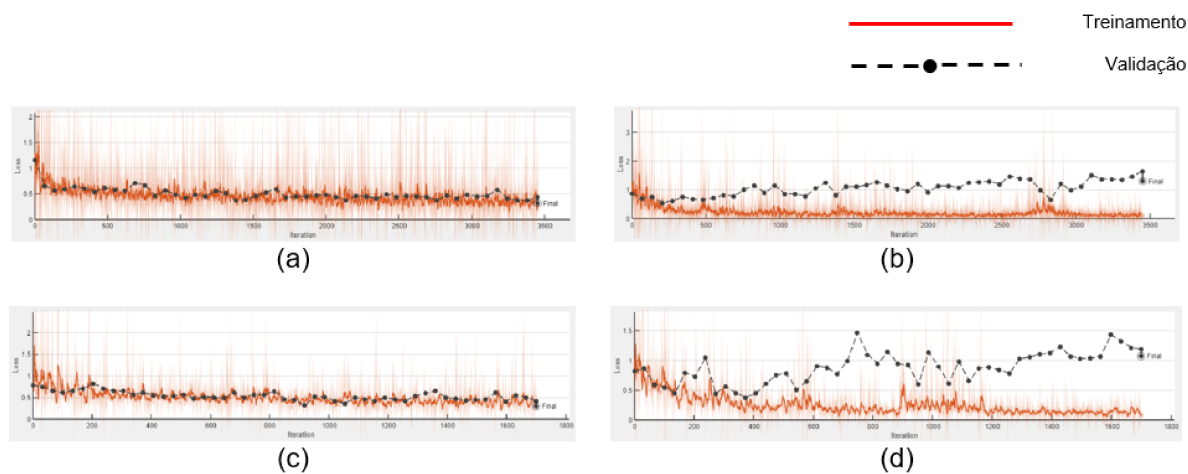


Figura 29 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de carcinoma escamoso da cavidade oral e a arquitetura *ResNet50*, onde (a) utilizando *Mixup* e aumento de dados geométricos, (b) utilizando somente *Mixup*, (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

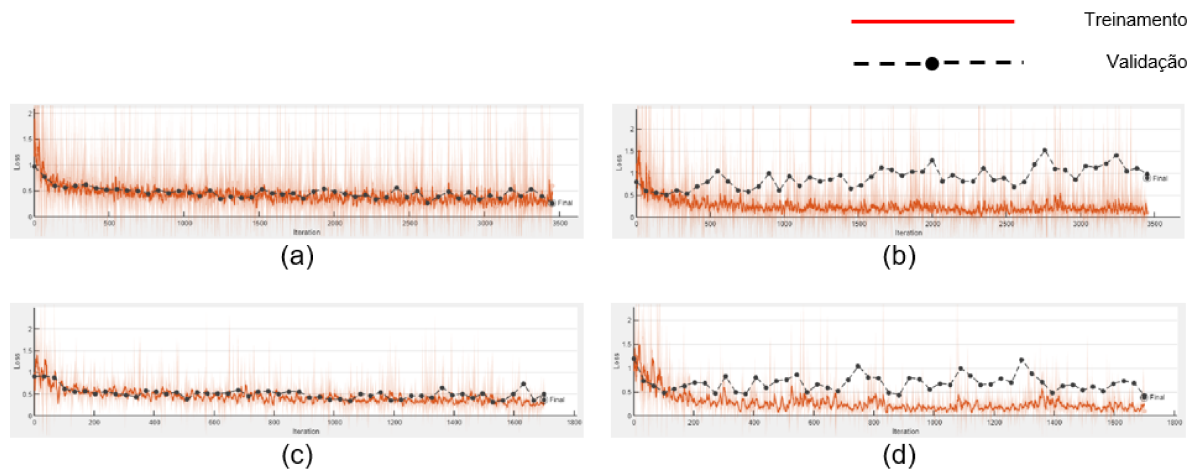


Figura 30 – Taxas de perda obtidas ao realizar treinamentos utilizando o conjunto de dados de carcinoma escamoso da cavidade oral e a arquitetura *DenseNet201*, onde (a) utilizando *Mixup* e aumento de dados geométricos, (b) utilizando somente *Mixup*, (c) utilizando somente aumento de dados geométricos e (d) não utilizando nenhuma técnica de aumento de dados.

4.5 Análise do método investigado com estudos da literatura

A Tabela 6 apresenta os valores dos métodos disponíveis na literatura para as bases de imagens investigadas nesse estudo. Nota-se que os conjuntos de dados empregados nos trabalhos citados foram semelhantes aos dados usados nesta pesquisa. É possível notar que as abordagens propostas também podem desempenhar um papel significativo nessa área, oferecendo estratégias relevantes para a classificação de lesões na cavidade oral.

Tabela 6 – Avaliação entre o modelo investigados para classificação das lesões e os estudos presentes na Literatura.

| Estudos | Displasia | OSCC |
|----------------------------------|-----------|--------|
| (RAHMAN et al., 2022) | - | 90,1% |
| (PANIGRAHI; DAS; SWARNKAR, 2022) | - | 97,3% |
| (DAS; DASH; MISHRA, 2023) | - | 97,8% |
| (SILVA et al., 2022a) | 90,3% | - |
| (SILVA et al., 2022b) | 98,0% | - |
| Método Investigado | 90,6 % | 94,9 % |

5 Conclusão

A obtenção de imagens histológicas de lesões da cavidade oral apresenta várias dificuldades e desafios, desde questões técnicas relacionadas à aquisição e processamento das amostras até preocupações éticas e políticas envolvendo a privacidade do paciente.

O principal objetivo desse trabalho foi investigar as técnicas de aumento de dados por transformações geométricas e a técnica *Mixup* como solução alternativa à escassez de conjuntos de treinamento. As técnicas foram utilizadas no treinamento dos modelos *ResNet50* e *DenseNet201* utilizando quatro estratégias diferentes: utilizando somente a técnica *Mixup*, utilizando somente transformações geométricas, utilizando ambas as técnicas combinadas e utilizando somente as imagens originais. As estratégias foram empregadas tanto nos treinamentos utilizando o conjunto de dados de displasia oral, quanto no conjunto de dados de OSCC.

A utilização da técnica *Mixup* combinada ao aumento de dados por transformações geométricas se mostrou promissor. Para o conjunto de displasia oral, os modelos *ResNet50* e *DenseNet201* alcançaram acurácias de 90,6%. O mesmo ocorreu para o conjunto de dados de OSCC, onde os maiores valores de F1-Score também foram obtidos ao utilizar as duas técnicas de aumento de dados combinadas, alcançando 91,7% com o modelo *ResNet50* e 94,9% com o modelo *DenseNet201*. Ao analisar os gráficos de perda, foi possível observar que as curvas correspondentes aos dados de treinamento e classificação estão sobrepostas, indicando boa generalização em todos os treinamentos realizados.

Por outro lado, o uso isolado da técnica *Mixup* não apresentou resultados significativos quando comparado às outras estratégias. Embora tenha apresentado bons resultados de acurácia em alguns casos, ao observar os gráficos de perda é possível notar um afastamento das curvas correspondentes ao treinamento e classificação, indicando um possível sobreajuste aos dados de treinamento.

O estudo mostrou que a técnica *Mixup* não substitui métodos convencionais de aumento de dados, porém pode complementá-los e sua aplicação deve ser considerada ao realizar treinamento com conjunto de dados limitados, uma vez que ela aumenta efetivamente o tamanho e a diversidade do conjunto sem a necessidade de dados rotulados adicionais.

Futuramente, pretende-se investigar a utilização de Redes Adversárias Generativas (do inglês, *Generative Adversarial Networks* - GAN) para geração de imagens sintéticas de lesões da cavidade oral. Pretende-se ainda investigar a utilização de Redes Neurais *Autoencoders* como etapa preliminar ao treinamento de modelos CNNs removendo ruídos dos dados gerados sinteticamente.

A implementação dos códigos utilizados para o desenvolvimento deste trabalho podem ser encontrados no *GitHub*¹.

5.1 Desafios Encontrados

Durante o desenvolvimento do projeto, foram encontradas algumas dificuldades significativas. Dentre elas, destacam-se o entendimento de tópicos relacionados à histologia, devido à falta de experiência e conhecimento prévio na área. Também foi necessário adaptar-se à linguagem *Matlab*, devido à falta de familiaridade com a mesma. Além disso, o tempo de execução dos experimentos foi afetado pelas limitações dos recursos computacionais disponíveis.

Esses desafios representaram oportunidades para o incremento dos conhecimentos já obtidos e reforçaram a importância de várias práticas e conceitos abordados durante o curso. Além disso, a superação das dificuldades encontradas contribuiu para a consolidação de uma aplicação cada vez mais robusta, que pode desempenhar um papel significativo na área interdisciplinar de biologia e computação.

¹ <https://github.com/diegorjesus/classificacaoLesoesOrais.git>

Referências

- ABATI, S. et al. Oral cancer and precancer: a narrative review on the relevance of early diagnosis. *International journal of environmental research and public health*, MDPI, v. 17, n. 24, p. 9160, 2020. Citado na página 14.
- BARNES, L. et al. World health organization classification of tumours: pathology and genetics of head and neck tumours. IARC publications, 2005. Citado na página 17.
- CHAN, H.-P.; HADJIISKI, L. M.; SAMALA, R. K. Computer-aided diagnosis in the era of deep learning. *Medical physics*, Wiley Online Library, v. 47, n. 5, p. e218–e227, 2020. Citado 2 vezes nas páginas 14 e 15.
- CHIERICI, G. A tumor registry study of oral squamous cell carcinoma. *J Oral Med*, v. 23, p. 91–98, 1968. Citado na página 17.
- Có, A. C. G. *Expressão da proteína Bcl-2 como potencial biomarcador de progressão tumoral em carcinoma epidermóide oral*. Dissertação (Mestrado) — Universidade Federal do Espírito Santo, 2019. Disponível em: <<http://repositorio.ufes.br/handle/10/10951>>. Citado 2 vezes nas páginas 7 e 18.
- DAS, A.; NAIR, M. S.; PETER, S. D. Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review. *Journal of digital imaging*, Springer, v. 33, p. 1091–1121, 2020. Citado na página 14.
- DAS, D. K. et al. Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. *Tissue and Cell*, v. 53, p. 111–119, 2018. ISSN 0040-8166. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0040816618301137>>. Citado 2 vezes nas páginas 28 e 29.
- DAS, M.; DASH, R. Data augmentation and its application in histopathological oral cell image classification. In: *2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS)*. [S.l.: s.n.], 2022. p. 194–198. Citado na página 25.
- DAS, M.; DASH, R.; MISHRA, S. K. Automatic detection of oral squamous cell carcinoma from histopathological images of oral mucosa using deep convolutional neural network. *International Journal of Environmental Research and Public Health*, v. 20, n. 3, 2023. ISSN 1660-4601. Disponível em: <<https://www.mdpi.com/1660-4601/20/3/2131>>. Citado na página 51.
- FREITAS, A. D. et al. Investigação de normalização de corantes para classificação de imagens da cavidade oral. Universidade Federal de Uberlândia, 2021. Citado 2 vezes nas páginas 7 e 25.
- FURTADO, M. I. V. Redes neurais artificiais: uma abordagem para sala de aula. *Ponta Grossa, PR. Atena Editora*, p. 19, 2019. Citado 2 vezes nas páginas 21 e 22.
- GARTNER, L. *Tratado de Histologia Em Cores*. Elsevier, 2007. ISBN 9788535223477. Disponível em: <<https://books.google.com.br/books?id=-mUoQXHQxO4C>>. Citado na página 18.

- GLOROT, X.; BORDES, A.; BENGIO, Y. Deep sparse rectifier neural networks. In: JMLR WORKSHOP AND CONFERENCE PROCEEDINGS. *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. [S.l.], 2011. p. 315–323. Citado na página 23.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento Digital de Imagens*. [S.l.]: Pearson Prentice Hall, 2010. Citado 2 vezes nas páginas 19 e 20.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado na página 25.
- HANNEN, E. J. et al. An image analysis study on nuclear morphology in metastasized and non-metastasized squamous cell carcinomas of the tongue. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, Wiley Online Library, v. 185, n. 2, p. 175–183, 1998. Citado na página 21.
- HAYKIN, S. *Neural networks and learning machines, 3/E*. [S.l.]: Pearson Education India, 2009. Citado 4 vezes nas páginas 7, 21, 22 e 23.
- HE, K. et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: IEEE. *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.], 2015. p. 1026–1034. Citado na página 23.
- 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 34.
- HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012. Citado na página 38.
- HUANG, G. et al. *Densely Connected Convolutional Networks*. arXiv, 2017. Disponível em: <<https://arxiv.org/abs/1608.06993>>. Citado 4 vezes nas páginas 8, 34, 37 e 38.
- INCA. *Diagnóstico Precoce do Câncer de Boca*. INCA, 2022. Disponível em: <<https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//livro-diagnostico-precoce-cancer-boca-2022.pdf>>. Citado na página 14.
- INCA. *Estimativa 2023: Incidência de Câncer no Brasil*. 2023. <<https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2023.pdf>>. Acesso em: 23 de junho de 2023. Citado na página 14.
- JUNQUEIRA, L.; CARNEIRO, J. *Histologia Básica*. 12th. ed. Rio de Janeiro: Guanabara Koogan, 2013. Citado 2 vezes nas páginas 18 e 21.
- KAUSHAL, C. et al. Recent trends in computer assisted diagnosis (cad) system for breast cancer diagnosis using histopathological images. *Irbm*, Elsevier, v. 40, n. 4, p. 211–227, 2019. Citado 4 vezes nas páginas 7, 19, 20 e 21.
- KHOSLA, C.; SAINI, B. S. Enhancing performance of deep learning models with different data augmentation techniques: A survey. In: IEEE. *2020 International Conference on Intelligent Engineering and Management (ICIEM)*. [S.l.], 2020. p. 79–85. Citado 3 vezes nas páginas 15, 25 e 33.

LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 23.

LI, X. et al. A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. *Artificial Intelligence Review*, Springer, v. 55, n. 6, p. 4809–4878, 2022. Citado na página 15.

LI, Z. et al. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, IEEE, 2021. Citado na página 23.

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 35.

MÜLLER, S. Update from the 4th edition of the world health organization of head and neck tumours: tumours of the oral cavity and mobile tongue. *Head and neck pathology*, Springer, v. 11, p. 33–40, 2017. Citado na página 17.

MULLER, S.; TILAKARATNE, W. M. Update from the 5th edition of the world health organization classification of head and neck tumors: tumours of the oral cavity and mobile tongue. *Head and Neck Pathology*, Springer, v. 16, n. 1, p. 54–62, 2022. Citado na página 17.

MUMUNI, A.; MUMUNI, F. Data augmentation: A comprehensive survey of modern approaches. *Array*, v. 16, p. 100258, 2022. ISSN 2590-0056. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2590005622000911>>. Citado 2 vezes nas páginas 25 e 33.

PANIGRAHI, S.; DAS, J.; SWARNKAR, T. Capsule network based analysis of histopathological images of oral squamous cell carcinoma. *Journal of King Saud University-Computer and Information Sciences*, Elsevier, v. 34, n. 7, p. 4546–4553, 2022. Citado na página 51.

RAHMAN, A.-u. et al. Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning. *Sensors*, v. 22, n. 10, 2022. ISSN 1424-8220. Disponível em: <<https://www.mdpi.com/1424-8220/22/10/3833>>. Citado na página 51.

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 31.

RAMADAN, S. Z. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review. *Journal of healthcare engineering*, Hindawi, v. 2020, 2020. Citado na página 19.

RICHART, R. M.; BARRON, B. A. A follow-up study of patients with cervical dysplasia. *American Journal of Obstetrics and Gynecology*, v. 105, n. 3, p. 386–393, 1969. ISSN 0002-9378. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0002937869902683>>. Citado na página 14.

- ROSEBROCK, A. *Deep Learning for Computer Vision*. San Francisco, CA: PyImageSearch, 2017. Citado 4 vezes nas páginas 7, 21, 23 e 24.
- SANTOS, D. F. D. d. et al. Automated segmentation of tumor regions from oral histological whole slide images using fully convolutional neural networks. Universidade Federal de Uberlândia, 2022. Citado 2 vezes nas páginas 28 e 29.
- SHAMSALDIN, A. S. et al. A study of the convolutional neural networks applications. *UKH Journal of Science and Engineering*, v. 3, n. 2, p. 31–40, 2019. Citado na página 23.
- SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of big data*, SpringerOpen, v. 6, n. 1, p. 1–48, 2019. Citado na página 25.
- SHULMAN, J. D.; GONZALES, C. B. Chapter 3 - epidemiology/biology of oral cancer. In: CAPPELLI, D. P.; MOBLEY, C. C. (Ed.). *Prevention in Clinical Oral Health Care*. Saint Louis: Mosby, 2008. p. 27–43. ISBN 978-0-323-03695-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780323036955500070>>. 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 8 vezes nas páginas 7, 21, 22, 23, 24, 25, 31 e 35.
- SILVA, A. B. et al. Computational analysis of histological images from hematoxylin and eosin-stained oral epithelial dysplasia tissue sections. *Expert Systems with Applications*, Elsevier, v. 193, p. 116456, 2022. Citado 3 vezes nas páginas 27, 29 e 51.
- SILVA, A. B. et al. Assessment of the association of deep features with a polynomial algorithm for automated oral epithelial dysplasia grading. In: IEEE. *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.], 2022. v. 1, p. 264–269. Citado 3 vezes nas páginas 27, 29 e 51.
- TILAKARATNE, W. M. et al. Oral epithelial dysplasia: Causes, quantification, prognosis, and management challenges. *Periodontology 2000*, Wiley Online Library, v. 80, n. 1, p. 126–147, 2019. Citado na página 14.
- WRIGHT, A.; SHEAR, M. Epithelial dysplasia immediately adjacent to oral squamous cell carcinomas. *Journal of Oral Pathology & Medicine*, Wiley Online Library, v. 14, n. 7, p. 559–564, 1985. Citado na página 17.
- YING, X. An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, IOP Publishing, v. 1168, n. 2, p. 022022, feb 2019. Disponível em: <<https://dx.doi.org/10.1088/1742-6596/1168/2/022022>>. Citado na página 25.
- ZHANG, H. et al. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. Citado 3 vezes nas páginas 15, 26 e 33.
- ZHANG, L. et al. How does mixup help with robustness and generalization? *arXiv preprint arXiv:2010.04819*, 2020. Citado na página 26.