

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Rafael Henrique de Oliveira Carvalho

**Classificação de Lesões da Cavidade Bucal
Baseada em Aprendizagem Profunda em
Ambiente Remoto**

Uberlândia, Brasil

2021

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Rafael Henrique de Oliveira Carvalho

**Classificação de Lesões da Cavidade Bucal Baseada em
Aprendizagem Profunda em Ambiente Remoto**

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 Engenharia Mecatrônica.

Orientador: Marcelo Zanchetta do Nascimento

Universidade Federal de Uberlândia
Bacharelado em Engenharia Mecatrônica

Uberlândia, Brasil

2021

Ao professor Marcelo pela paciência na orientação, pelo convívio, pelo apoio, pela compreensão e pela amizade que tornaram possível a conclusão deste projeto. À minha futura esposa por todo o apoio, incentivos e discussões que tornaram esse trabalho possível. À minha família por todo suporte que eles propiciaram.

Resumo

A displasia epitelial oral é um precursor relativamente comum do câncer de boca em que sua progressão para o câncer varia de 6% a 36%. Existem várias técnicas de imagem empregadas em seu diagnóstico. A análise histológica de imagens suportada pelos sistemas computacionais mostrou-se bastante eficaz no diagnóstico da doença. Neste trabalho, é apresentada uma abordagem para quantificar e classificar amostras de tecido bucal com uso de redes neurais convolucionais (AlexNet, Vgg16 e ResNet) e transferência de aprendizado. A transferência de aprendizado utilizada foi a baseada em rede, em que se emprega a reutilização de uma parte da rede pré-treinada no domínio da fonte, para aplicação no domínio destino. Devido ao tamanho do banco de imagens, aplicou-se uma etapa de aumento de dados para avaliar a acurácia das arquiteturas das redes. Definiu-se 30% das imagens, aleatoriamente selecionadas, para o grupo de teste e dentre 70% restantes, 90% ficariam para o grupo de treinamento e outros 10% para a validação. Ao final do treinamento obteve-se resultados relevantes, atingindo 96,56% de acurácia com a ResNet18 e 94,33% com a VGG16. Esse estudo ainda apresenta uma aplicação para dispositivos móveis para disponibilizar ao usuário um meio de classificar imagens histológicas. Essa aplicação é composta por um bot do *Telegram* e algoritmos em linguagem MatLab para as execução das CNNs. Levando em consideração o contexto pandêmico da COVID-19, esses resultados são considerados relevantes e essa abordagem pode ser útil como um protocolo que contribuirá na análise de diagnóstico de lesões da cavidade oral.

Palavras-chave: Telegram, Imagens Histológicas, Cavidade Oral, CNNs, Transferência de Aprendizado, Dispositivos Móveis.

Lista de ilustrações

Figura 1 – Exemplos de tecidos histológicos da região da cavidade oral: (a) tecido saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa. As setas em cor vermelha indicam núcleos celulares e as setas em cor azul indicam o epitélio. Fonte: (SILVA et al., 2019).	15
Figura 2 – Diagrama de Venn descrevendo a localização do DL Adaptado de: (ROSEBROCK, 2017).	16
Figura 3 – Novo paradigma de Programação empregado nos algoritmos de AM. Adaptado de: (CHOLLET et al., 2018).	17
Figura 4 – Modelo de Uma Rede <i>Perceptron</i> Adaptado de: (ROSEBROCK, 2017)	18
Figura 5 – Representação de uma convolução em uma CNN. Adaptado de: (BIANCHI, 2020)	20
Figura 6 – Exemplos de função de ativação (Step, Sigmoid, tanh, ReLU e suas variações). Adaptado de: (ROSEBROCK, 2017)	21
Figura 7 – Exemplos da operação de um <i>Max-Pooling</i> . Adaptado de: (SAVYAKHOSLA, 2019)	21
Figura 8 – A rede apresentada em (a) mostra uma arquitetura padrão e (b) mostra as conexões existentes após uso do <i>dropout</i> . Adaptado de: (SHANMUGAMANI, 2018)	22
Figura 9 – Exemplos de um CNN com as diversas camadas. Adaptado de: (GURUCHARAN, 2020)	22
Figura 10 – Exemplo de tecido histológico da cavidade bucal: (a) saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa. Adaptado de: (SILVA et al., 2019)	27
Figura 11 – Arquitetura do método proposto nesse trabalho.	27
Figura 12 – A estrutura da do modelo de CNN Vgg16 Adaptado de: (HASSAN, 2018).	30
Figura 13 – Exemplo do processo do processo residual da arquitetura ResNet. Adaptado de: (HE et al., 2015)	31
Figura 14 – Transferência de aprendizado por uma rede neural convolucional <i>AlexNet</i> . Adaptado de: (LU; LU; ZHANG, 2019)	31
Figura 15 – Arquitetura da solução para dispositivo móvel para classificação das imagens histológicas.	32
Figura 16 – Tela de criação do bot para o aplicativo Telegram com o software Bot-Father.	33
Figura 17 – Representação de imagens sintéticas geradas pela técnica de aumento de dados.	36

Figura 18 – Representação da etapa de treinamento utilizando o modelo AlexNet: avaliação da acurácia (imagem superior) e função de perda (imagem inferior).	37
Figura 19 – Etapa de treinamento utilizando o modelo ResNet: avaliação da acurácia (imagem superior) e função de perda (imagem inferior).	38
Figura 20 – O modelo da VGG na etapa de treinamento: a acurácia (imagem superior) e função de perda (imagem inferior).	39
Figura 21 – A representação de CAM (direita) para uma imagem da classe normal (esquerda) empregando a Resnet18 para a classificação.	40
Figura 22 – Imagem do displasia leve (esquerda) e mapeamento de ativação de classe de uma imagem com displasia leve usando a Resnet18 (direita).	41
Figura 23 – Representação da CAM (esquerda) de uma imagem da categoria moderado (direita) usando a Resnet18.	41
Figura 24 – Imagem de classe grave (esquerda) e mapeamento de ativação de classe (direita) usando o modelo Resnet18 para classificação.	41
Figura 25 – Exemplo de funcionamento do Broker com o tópico <i>NovaRede</i>	42
Figura 26 – Operação do Broker com o tópico <i>NovaFoto</i>	42
Figura 27 – Resultado do tratamento do fluxo das informações recebidas no Node-RED.	43
Figura 28 – Exemplo de funcionamento do bot na interação do usuário para escolha das opções.	44
Figura 29 – Tela com informações da submissão de uma imagem e o processo de retorno do servidor com o valor de acurácia para uma imagem histológica.	44

Lista de tabelas

Tabela 1 – Alguns critérios usados para diagnosticar a displasia. **Fonte:** (WAR-
NAKULASURIYA et al., 2008). 14

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
CAD	Computer-Aided Diagnosis
CNN	Convolution Neural Network
CPU	Central Processing Unit
DL	Deep Learning
FC	Fully-Connected
FSO	Fibrose Submucosa Oral
GPU	Graphics Processing Unit
GUI	Graphical User Interface
H&E	Hematoxilina-Eosina
IA	Inteligência Artificial
OMS	Organização mundial da Saúde
RNA	Redes Neurais Artificiais
ROI	Region of Interest
TA	Transferência de Aprendizagem

Sumário

1	INTRODUÇÃO	10
1.1	Objetivos	12
1.2	Organização desse Trabalho de Conclusão de Curso	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Displasias	14
2.2	Sistema de Apoio ao Diagnóstico	15
2.3	Aprendizagem Profunda	16
2.3.1	Redes Neurais Convolucionais	19
2.4	Transferência de Aprendizado	22
2.5	Trabalhos Relacionados	23
3	METODOLOGIA	26
3.1	Banco de Imagens	26
3.2	Algoritmos para Classificação de Lesões de Displasias	27
3.2.1	Pré-Processamento	28
3.2.2	Redes Neurais Convolucionais	28
3.2.2.1	AlexNet	28
3.2.2.2	Vgg16	29
3.2.2.3	ResNet	30
3.2.3	Transferência de Aprendizado	31
3.3	Avaliação de Desempenho	32
3.4	Aplicação para Dispositivos Móveis	32
3.4.1	Telegram	32
3.4.2	Node-RED	33
3.4.3	Comunicação MQTT	33
3.4.3.1	Eclipse Mosquitto MQTT Broker	34
4	RESULTADOS	36
4.1	Classificação das Imagens Histológicas com CNN	36
4.2	Aplicação em Dispositivo Móvel	42
5	CONCLUSÃO	45
5.1	Desafios Encontrados	45
5.2	Trabalhos Futuros	46

REFERÊNCIAS	47
--------------------------	-----------

1 Introdução

A displasia oral é um precursor relativamente comum do câncer de boca ou carcinoma de células escamosas em que sua progressão para o câncer varia de 6% a 36% (SMITH et al., 2009). Estima-se que 95% dos cânceres da cavidade oral são carcinomas de células escamosas, sendo que as outras lesões são compostas em grande parte de adenocarcinomas de glândulas salivares. O tumor maligno epitelial agressivo é o sexto tipo de neoplasia mais comum no mundo e, apesar dos inúmeros avanços no tratamento, a taxa de sobrevivência de longo prazo é inferior a 50% (ABBAS et al., 2008).

O carcinoma de células escamosas está diretamente associado ao aparecimento de lesões potencialmente malignas em tecidos das regiões da boca (AXÉLL, 1987). Dentre esses tipos, as leucoplasias configuram como as mais importantes, especialmente, quando há alterações morfológicas conhecidas como as displasias. Assim, essa neoplasia se desenvolve a partir da progressão de uma hiperplasia epitelial, passando por diferentes graus que vão da displasia leve, moderada e severa, até carcinoma *in situ*. Essa neoplasia ainda pode progredir para um estágio mais avançado em que há uma invasão do tecido conjuntivo subjacente (SCULLY, 2014). Embora existam fatores que possam ser preditivos para a transformação maligna dessas lesões bucais, a identificação da displasia e o grau da lesão são tarefas complexas. Mas a melhor estratégia é a predição da evolução para que não ocorra uma evolução para um estado maligno.

Em histopatologia, um dos principais métodos para análise e investigação de tecidos com esse tipo de lesão se dá com o de emprego corantes como a Hematoxilina e Eosina (H&E). Assim, estudos moleculares usando essas imagens que avaliam perda de heterozigoticidade, têm auxiliado na predição da evolução do estado dessas lesões. Isso tem permitido estudos investigativos para o diagnóstico e evolução das lesões displásicas, mas no geral, essas estratégias ainda apresentam um grande grau de subjetividade e carga de trabalho que reduz a acurácia das análises pelo especialista (KUJAN et al., 2006).

Pelo fato das neoplasias não apresentarem padrão histológico e ter comportamento clínico diferenciado, a determinação de um diagnóstico preciso não é uma tarefa trivial (WRIGHT; ADDIS; LEONG, 2011). Normalmente, a discriminação entre os diferentes graus da lesão é complexa e repleta de subjetividade. Nessa tarefa, um especialista analisa com uso de um microscópio as características presentes na estrutura externa e interna dos núcleos e estruturas celulares presentes na amostra.

Nas últimas décadas, o uso dos microscópios de luz, equipados com câmeras digitais, auxiliam consideravelmente os patologistas na avaliação de lâminas dos tecidos histológicos. As imagens digitais obtidas por esses equipamentos permitem, como um ponto

inicial, o desenvolvimento de algoritmos computacionais para processamento digital das imagens das amostras, podendo fornecer uma ferramenta para o sistema computacional de apoio ao diagnóstico (do inglês, *computer-aided diagnosis* - CAD) (SERTEL et al., 2010).

Nas análises assistidas por um sistema CAD, os modelos de redes neurais convolucionais (do inglês, *convolutional neural networks* - CNN) têm sido empregados para classificação de imagens (HE et al., 2016). Em histologia, alguns estudos trazem ótimos resultados ao desenvolver metodologias que envolvam CNN na classificação de imagens de lesões de tumores (CRUZ-ROA et al., 2017; ARAÚJO et al., 2017). No entanto, o uso de algoritmos de aprendizado de máquina para uma etapa de classificação exige dados necessários para treinamento dos modelos e dados para teste destes modelos para que seja empregado nas aplicações. Então, desenvolver novos modelos CNN é um processo custoso, pois demanda bastante recursos computacionais e uma grande quantidade de imagem para o treinamento da rede. As CNN ainda requerem GPU, (do inglês *graphics processing unit*) e CPU (do inglês, *central processing unit*) de alto desempenho para a fase de treinamento dos dados.

Uma estratégia para acelerar esse processo é o uso de Transferência de Aprendizado (TA). Essa técnica consiste em reciclar o conhecimento de modelos previamente treinados com grande volume de dados. Isso possibilita empregar pesos pré-treinados em diversas bases de imagens como a ImageNet. Esses pesos diminuem a necessidade de retrainar todos os parâmetros, assim como utilizar informações existentes na rede pré-treinada (YOSINSKI et al., 2014). Então, essa abordagem usa o conhecimento existente em dados e modelos de um domínio, para melhorar o desempenho de um modelo treinado em um outro domínio que ainda há pouco conhecimento prévio. Isso tem permitido usar esses modelos que demandam grande poder computacional em equipamentos como os computadores pessoais. Com isso, não é necessário muito tempo para a fase de treinamento, pois geralmente esses processos requerem apenas alguns minutos (LU; LU; ZHANG, 2019). Existem várias arquiteturas de CNN disponíveis para o uso. Exemplos de modelos CNN são AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), Resnet18 (HE et al., 2016) e Vgg16 (SIMONYAN; ZISSERMAN, 2015).

Devido a COVID-19, seguindo as orientações do Comité de Monitoramento os campi da UFU foram parcialmente fechados. Esse problema tem limitado o número de usuários em laboratórios de histologia para análise e investigação de tecidos histológicos. Uma forma de contribuir é permitir o uso dessas tecnologias empregadas em CAD por meio de acesso remoto. Essa condição possibilita aos estudantes e docentes novas experiências, além de ser um método promissor no processo de análise de tecidos histológicos.

Nesse trabalho, um método composto pela utilização de CNN pré-treinada e técnica de transferência de aprendizado e aumento de dados é proposto para diagnosticar

lesões de displasia epitelial oral, separadas em quatro grandes grupos: *Leve*, *Moderado*, *Grave* e *Normal*. Esse estudo dedicou-se em investigar o comportamento dos modelos AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), Resnet18 (HE et al., 2016) e Vgg16 (SIMONYAN; ZISSERMAN, 2015) em um banco de imagens de displasia oral. Além disso, é apresentada uma ferramenta para dispositivos móveis baseada no programa de mensagens instantâneas *Telegram Messenger*, que permite ao usuário uma interação de forma simples para investigação de imagens.

1.1 Objetivos

Este trabalho pretende atender a necessidade de especialistas na diferenciação das regiões suspeitas em imagens histológicas de displasia bucal por meio de novas metodologias computacionais. Assim, este trabalho propõe desenvolver algoritmos capazes de classificar tecidos histológicos baseados em arquiteturas CNN com abordagem de TA para aplicação em dispositivos móveis. Portanto, pontualmente, pretende-se:

- Propor metodologias com uso de CNN e TA para classificação de imagens histológicas de displasia bucal em diferentes níveis do estágio da lesão;
- Investigar os diferentes modelos das CNN para classificação das diferentes classes das lesões do tipo displasia;
- Desenvolver uma ferramenta para dispositivos móveis para uso dos algoritmos de classificação de lesões displásicas.

1.2 Organização desse Trabalho de Conclusão de Curso

Neste capítulo, foram apresentadas as considerações iniciais, motivação e objetivos para o desenvolvimento do método proposto. O restante deste trabalho possui a seguinte organização:

- **Capítulo 2 - Fundamentação Teórica:** apresenta a teoria e estado da arte de tópicos relacionados a histologia, processamento digital de imagens médicas, CNN e transferência de aprendizado;
- **Capítulo 3 - Metodologia:** apresenta os materiais e métodos utilizados para o desenvolvimento deste trabalho, sendo eles: descrição do banco de imagens histológicas utilizado, descrição das técnicas usadas no treinamento e teste das redes, método para o desenvolvimento da aplicação para os dispositivos móveis, bem como

as métricas empregadas para avaliação quantitativa do método proposto;

- **Capítulo 4 - Resultados:** apresenta os resultados obtidos em cada etapa empregada, bem como os resultados quantitativos dos experimentos e uma breve discussão desses resultados;
- **Capítulo 5 - Conclusão:** apresenta as considerações e pesquisas futuras deste trabalho.

2 Fundamentação Teórica

Neste capítulo, são apresentados os conceitos básicos necessários para compreensão deste trabalho bem como os trabalhos relacionados.

2.1 Displasias

As displasias são anormalidades benignas pré-cancerígenas que apresentam potencial de se tornar malignas podendo assim evoluir para um câncer ([WARNAKULASURIYA et al., 2008](#)). A Tabela 1 contém alguns critérios utilizados para classificar uma lesão como uma displasia. As displasias são divididas em três classes ou grau de intensidade da anormalidade: leve, moderado e grave. O grau de displasia é determinado pela intensidade da anormalidade acima da membrana basal epitelial como originalmente definido pela Organização Mundial da Saúde (OMS) ([SMITH et al., 2009](#)). Em grande parte dos casos, essas lesões não evoluem para um câncer naturalmente, mas por estimularem a proliferação regenerativa e exporem as células as condições que podem ocasionar inflamações podem aumentar as chances da evolução ao câncer. Na Figura 1 são mostradas lesões e também uma comparação entre os diversos níveis dessa lesão.

Tabela 1 – Alguns critérios usados para diagnosticar a displasia. **Fonte:** ([WARNAKULASURIYA et al., 2008](#)).

Arquitetura	Citologia
Estratificação epitelial irregular	Varição anormal no tamanho nuclear (anisonucleose)
Perda de polaridade das células basais	Varição anormal da forma nuclear
Cumes rete em forma de gota	Varição anormal no tamanho das células (anisocitose)
Maior número de figuras mitóticas	Varição anormal da forma celular (pleomorfismo celular)
Mitoses anormalmente superficiais	Aumento do tamanho nuclear

Existem diversos tipos de displasias que ocorrem em várias partes do corpo. A displasia oral, a qual é objeto de estudo neste trabalho, é um precursor relativamente comum do câncer de boca ou carcinoma de células escamosas e sua progressão para o câncer varia em intervalos de 6% a 36% ([SMITH et al., 2009](#)). Aproximadamente, 95% dos cânceres da cavidade oral são carcinomas de células escamosas, sendo que os restantes são compostos em grande parte de adenocarcinomas de glândulas salivares. Esse tumor maligno epitelial agressivo é a sexta neoplasia mais comum no mundo apesar dos inúmeros avanços no tratamento. A taxa de sobrevivência de longo prazo tem resultados inferiores a 50% ([ABBAS et al., 2008](#)).

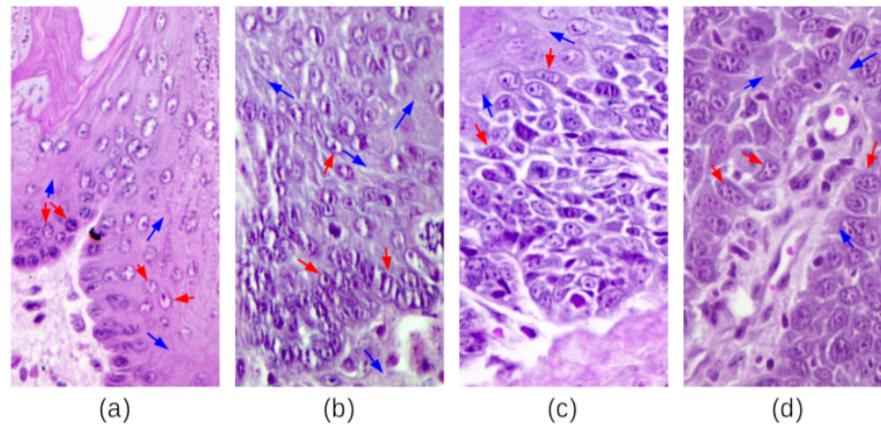


Figura 1 – Exemplos de tecidos histológicos da região da cavidade oral: (a) tecido saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa. As setas em cor vermelha indicam núcleos celulares e as setas em cor azul indicam o epitélio. **Fonte:** (SILVA et al., 2019).

Nas lesões pré-cancerosas, como a displasia, geralmente há grande similaridade entre células doentes e saudáveis, e sua taxa de crescimento é reduzida, podendo levar anos para ocorrer uma transformação mais severa (SILVA et al., 2019). Além disso, no processo de diagnóstico, as lesões de mesmo tamanho podem ser classificadas em graus diferentes pelo especialista, o que torna o diagnóstico subjetivo e dificulta a distinção entre as lesões (WARNAKULASURIYA et al., 2008). Por isso, torna-se necessária a investigação de métodos quantitativos que possam caracterizar essas lesões. Em processamento digital de imagens, técnicas empregadas na etapa de extração de características podem contribuir por meio de sistemas de diagnósticos assistidos por computador para padronização e quantificação dessas lesões.

2.2 Sistema de Apoio ao Diagnóstico

O diagnóstico assistido por computador (do inglês, *computer-aided diagnosis - CAD*), também conhecido como sistema de apoio ao diagnóstico, nos últimos anos se tornou uma importante ferramenta empregada na rotina clínica no trabalho da detecção e classificação de cânceres (DOI, 2007). Essa ferramenta emprega técnicas com objetivo de facilitar o diagnóstico para um tratamento adequado, funcionando como um solução suplementar para a tomada de decisão do especialista. A implementação dessas técnicas consiste em desenvolver algoritmos capazes de separar os objetos, extrair informações das imagens e assim classificá-las para a detecção mais precisa das anormalidades. As técnicas devem ser capazes de fazer uma descrição das características da imagem e, posteriormente, uma classificação. Um sistema CAD tem como base as fases: pré-processamento, segmentação, descrição e reconhecimento. A etapa de pré-processamento tem o intuito de melhorar a qualidade das características da imagem, acentuando-a e tornando-a mais adequada para

seu manuseio em determinadas aplicações. Na etapa de segmentação, a imagem passa por um processo de identificação e separação das regiões de interesse ou objetos existentes. O reconhecimento e descrição do processamento de imagens digitais refere-se a extração de características ou propriedades que discriminam as classes de objetos contidas na imagem. No sistema CAD, a etapa de segmentação pode ocorrer de forma manual ou automática, sendo já explorado na literatura há um certo tempo com resultados promissores. O uso de Inteligência Artificial (IA), principalmente Aprendizado Profundo (Do inglês, *Deep learning* - DL), em sistemas CAD tem se popularizado e cada vez mais diferentes tipos de algoritmos são empregados para contribuir com soluções para os desafios ainda presentes.

2.3 Aprendizagem Profunda

O Aprendizado Profundo (*Deep Learning* - DL) é subcampo do Aprendizado de Máquina (do inglês, *Machine Learning* - ML), em especial de redes neurais, que por sua vez é um subcampo de IA (ver Figura 2). O estudo de [Chollet et al. \(2018\)](#) traz que a IA nasceu na década de 1950, com a ideia de que computadores ou programas de computadores poderiam ser feitos para "pensar" e conseqüentemente permitiriam automatizar tarefas intelectuais normalmente realizadas por humanos.

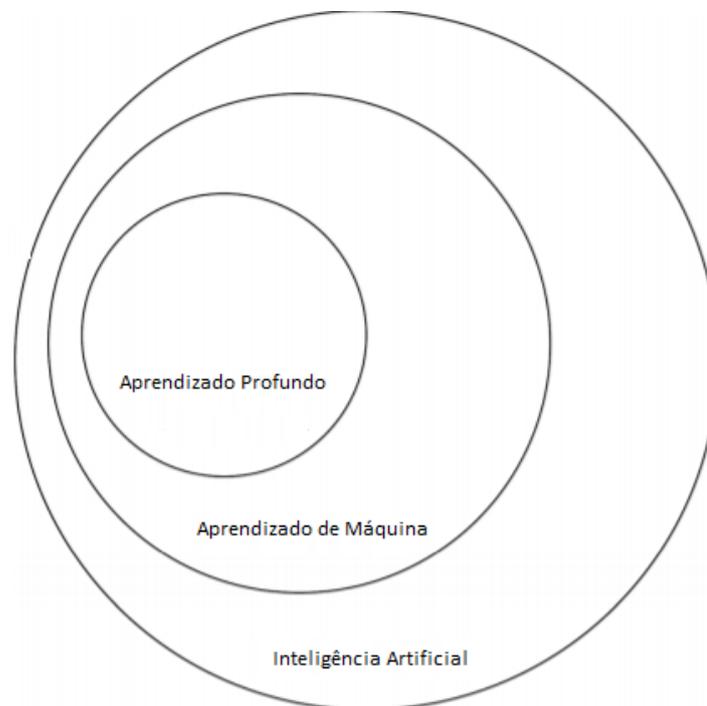


Figura 2 – Diagrama de Venn descrevendo a localização do DL **Adaptado de:** ([ROSE-BROCK, 2017](#)).

Em ([CHOLLET et al., 2018](#)), os autores citam que os primeiros programas de xadrez, tratadas como IA simbólicas, envolviam apenas regras codificadas elaboradas por

programadores, não se qualificando como AM. Essa abordagem com uso de IA simbólica se mostrou adequada para resolver problemas lógicos bem definidos, como jogo de xadrez, porém ao se tratar problemas complexos e imprecisos, como classificação de imagem, reconhecimento de fala e tradução de idioma, a abordagem baseada em AM tornou-se promissora.

Nessa abordagem, diferente do paradigma da IA clássica, em que regras e dados eram as entradas do programa, tendo as respostas como saída, houve uma inversão, alterando esse fluxo. Agora passou-se a ter como entradas os dados e respostas, e como saída as regras (ver Figura 3). Com essa abordagem, as regras podem ser aplicadas a novos dados para produzir respostas originais. Nesse sentido, segundo (ROSEBROCK, 2017), o subcampo de AM passou a ter especificamente interesse em reconhecimento de padrões e aprendizagem de dados.



Figura 3 – Novo paradigma de Programação empregado nos algoritmos de AM. **Adaptado de:** (CHOLLET et al., 2018).

Chollet et al. (2018) trazem ainda que para explorar o AM, são necessários:

- Dados de entrada: áudio para reconhecimento de fala ou foto para análise de imagens.
- Exemplos da saída esperada: transcrições de arquivos de som e categoria da imagem.
- Um método de avaliação: Necessário para determinar a distância entre a saída atual do algoritmo e sua produção esperada. É usado como um sinal de retorno (*feedback*) para ajustar a forma como o algoritmo funciona, permitindo que essa técnica possa estar aprendendo.

Assim o AM busca representações úteis de alguns dados de entrada, dentro de um espaço predefinido de possibilidades, usando orientação de um sinal de retorno.

As estratégias baseadas em DL são técnicas obtidas de Redes Neurais Artificiais (RNA) (SHANMUGAMANI, 2018). As RNA's são uma classe de algoritmos de aprendizado de máquina que aprendem com dados e são especializados no reconhecimento de

padrões, inspirado na estrutura, como os neurônios se relacionam, e função do cérebro, como a informação é processada. Uma RNA então é composta por várias camadas (neurônios), das quais cada uma tem uma função específica no processo de análise e classificação das informações.

A primeira estrutura de uma RNA surgiu ainda em 1950, com o algoritmo denominado *Perceptron* publicado por Rosenblatt (ROSENBLATT, 1958). Esse modelo se utiliza de diversas entradas e de somatórios ponderados para produzir uma saída, uma classificação. Na Figura 4 é apresentado um exemplo de uma rede *Perceptron*. Nesse modelo há um processo automático, em que os pesos para os neurônios são obtidos durante o processo de treinamento do algoritmo. Para um problema de classificação binária, por exemplo, as entradas são ponderadas e somadas passando por uma função de degrau unitário ou função de ativação, a qual permite definir a categoria (classe) dos dados.

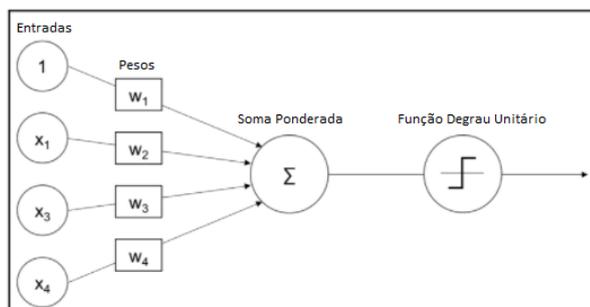


Figura 4 – Modelo de Uma Rede *Perceptron* Adaptado de: (ROSEBROCK, 2017)

Rosebrock (2017) explica que esse método evoluiu ao longo dos anos, porém no final da década de 1960, por meio da publicação de Minsky e Papert Block (1970), essa área estagnou por cerca de quase uma década. Isso ocorreu porque o trabalho de Minsky e Papert (BLOCK, 1970) demonstrou que a rede *Perceptron* com uma função de ativação (independentemente da profundidade) era um simples classificador linear, incapaz de resolver problemas não lineares como a função lógica XOR. De acordo com Rosebrock (2017) esse artigo sozinho, fez com que essa área ficasse sem novas contribuições e quase totalmente esquecida. Somente após 1980 com os trabalhos de (RUMELHART; HINTON; WILLIAMS, 1986) e (LECUN et al., 2012), e suas investigações com algoritmos de retropropagação, as pesquisas em redes neurais retomaram força, pois tinha-se RNA feedforward de multicamadas, as quais permitiram melhorar as soluções ao aprenderem com seus erros.

Novos estudos foram surgindo até chegar-se nas DLs com múltiplas camadas ocultas. Atualmente, com a possibilidade de bons recursos computacionais baseados em GPU, pode-se processar dados em alta velocidade e isso tem permitido novas pesquisas e contribuições para novas arquiteturas baseadas em DL. Nesse contexto, pode-se treinar redes com muito mais camadas ocultas que são capazes de explorar aprendizagem hierárquica,

em que conceitos simples são aprendidos nas camadas inferiores e mais padrões abstratos nas camadas superiores da rede. Um exemplo de DL aplicado ao aprendizado de recursos são as Redes Neurais Convolucionais (do Inglês, *Convolutional Neural Network - CNN*).

2.3.1 Redes Neurais Convolucionais

As CNNs (LECUN et al., 1998) estão entre os classificadores de imagem mais potentes. Vários trabalhos abordam o seu uso e maneiras de aprimorar as redes como os trabalhos de (HUSSAIN; BIRD; FARIA, 2018) e (LEI; PAN; HUANG, 2019). (ROSE-BROCK, 2017) descreve que a grande diferença de uma CNN para as demais abordagens é que, ao invés de definir manualmente um conjunto de regras e algoritmos para extrair características de uma imagem, esses atributos são aprendidos automaticamente com o processo de treinamento, os quais são usados na etapa de classificação. Esse modelo permite unir um sistema capaz de extrair características e ao mesmo tempo fazer a classificação de dados baseados em um método supervisionado.

A CNN, é um tipo especializado de rede neural para processamento de dados que tem uma topologia semelhante a uma grade, as qual permite a exploração de série temporal, (1-D), e com dados de imagem (2-D). O nome convolucional, vem de operações matemáticas dadas pela operação de convolução.

A convolução é uma operação linear, que a partir de duas funções dadas, resulta numa terceira que mede a soma do produto dessas funções ao longo da região subentendida pela superposição delas em função do deslocamento existente entre elas, conforme descrito na Equação 2.1. Outra forma comumente usada é a equação 2.2. Vale salientar que as funções devem ser integráveis no intervalo analisado.

$$s(t) = \int (x(a)w(t-a)da), \quad (2.1)$$

$$s(t) = (x * t)(t), \quad (2.2)$$

Em que, o primeiro argumento $x(a)$ da operação, é muitas vezes referido como a entrada a matriz que corresponde a imagem, enquanto $w(a)$ corresponde ao *kernel* aplicado no processo convolucional, que produzirá o mapa de características denominado $s(t)$ (GOODFELLOW; BENGIO; COURVILLE, 2016). Em seu trabalho Bianchi (2020), explica que durante a aplicação da convolução em uma imagem, o *kernel* vai se deslocando ao longo da imagem, como uma janela deslizante, em que operações de multiplicação e soma são aplicadas seguindo um deslocamento determinado pela operação de *stride*. Os detalhes dessa operação podem ser visto na Figura 5.

Nas CNNs outros tipos de camadas são usadas para sua construção. A camada de ativação, o *pooling*, as camadas Totalmente conectadas (do inglês, Fully-connected - FC),

a normalização (do inglês, *Batch Normalization*) e *dropout*. A arquitetura da CNN então é formada pela interação entre uma série dessas camadas de uma maneira específica (ROSEBROCK, 2017). Como os dados de treinamento não mostram a saída desejada para cada uma dessas camadas, essas são também conhecidas como camadas ocultas (SHANMUGAMANI, 2018).

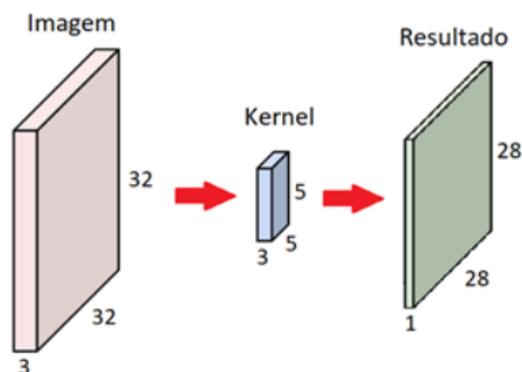


Figura 5 – Representação de uma convolução em uma CNN. Adaptado de: (BIANCHI, 2020)

A camada de ativação é geralmente subsequente a uma camada convolucional, em que normalmente são aplicadas funções de ativação. Essa é uma área ativa que vem sendo explorada para propor novos modelos para uma função de ativação. As principais funções são: binária, linear, sigmóide, tanh, ReLU e Leaky ReLU (ver Figura 6).

Uma função muito utilizada é a ReLU (do inglês, *Rectified Linear Unit*). Sua principal característica é que ela não ativa todos os neurônios ao mesmo tempo, ou seja, a entrada for negativa, ela será convertida em zero e o neurônio não será ativado. Isso faz com que apenas alguns neurônios sejam ativados, tornando a rede esparsa e eficiente para a computação.

O *pooling* é uma camada responsável por reduzir progressivamente o tamanho do espaço de uma entrada. Isso permite reduzir a quantidade de parâmetros e uso computacional da rede, assim como minimizar os efeitos de sobre ajuste (do inglês, *overfitting*). Geralmente, essa camada é inserida entre conjuntos consecutivos formados por camada convolucionais e de ativação. Uma ilustração desse processo pode ser vista na Figura 7, em que mostra uma técnica conhecida como *max-pooling*, em que somente o máximo valor da janela é usado como saída.

As camadas FCs são sempre colocadas no final antes de aplicar a camada de classificação (ROSEBROCK, 2017). Essas camadas são aquelas camadas onde todas as entradas de uma camada são conectadas a cada unidade de ativação da próxima camada. As camadas FC são conectadas por meio de um peso adaptado por meio de treinamento. O ajuste dos pesos geralmente é feito pelo uso do algoritmo retropropagação. Geralmente,

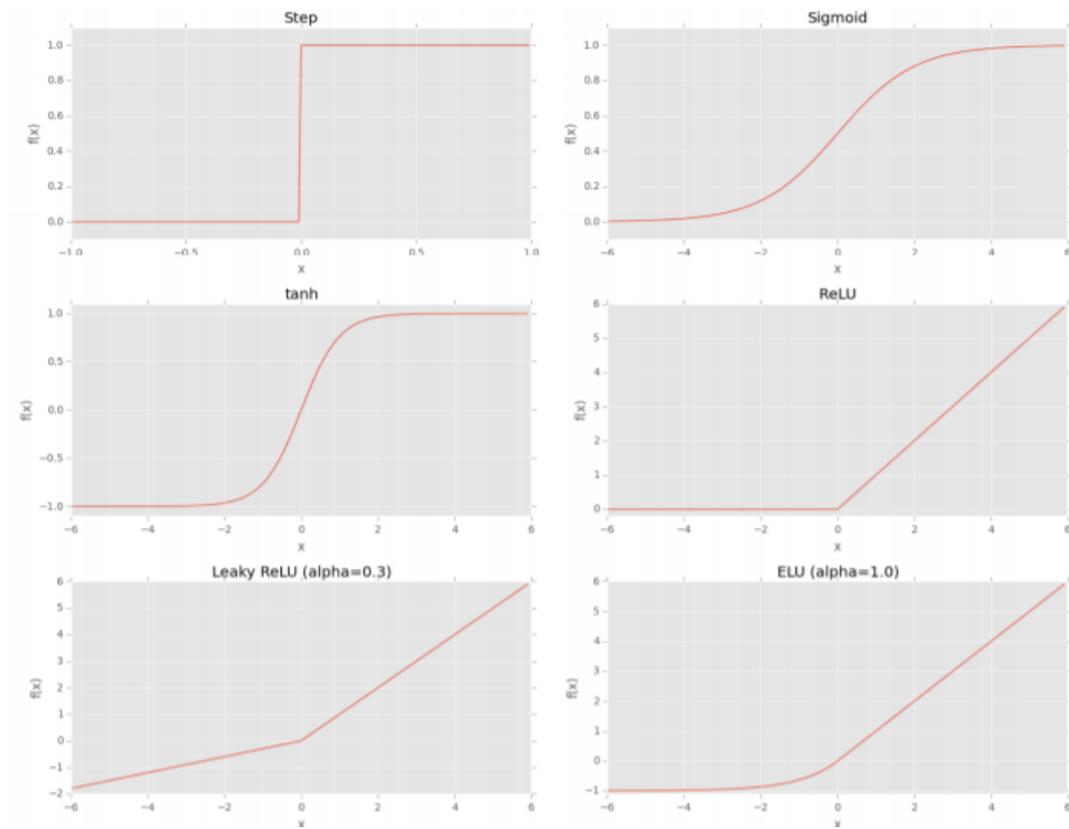


Figura 6 – Exemplos de função de ativação (Step, Sigmoid, tanh, ReLU e suas variações). Adaptado de: (ROSEBROCK, 2017)

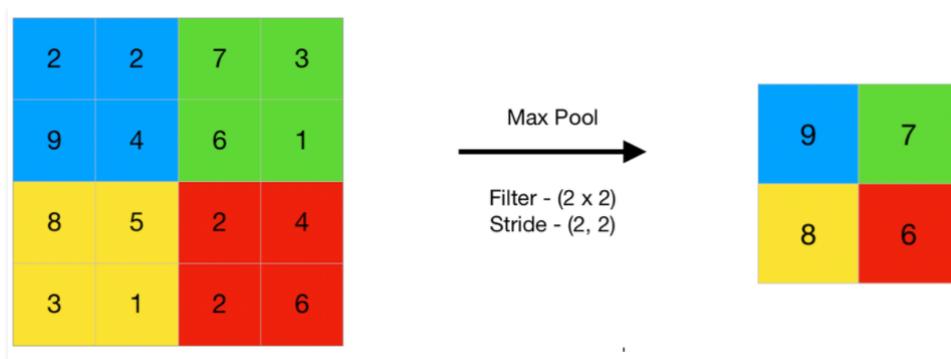


Figura 7 – Exemplos da operação de um *Max-Pooling*. Adaptado de: (SAVYAKHOSLA, 2019)

o número de *perceptrons* da última camada FC é igual ao número de classes da base de dados em treinamento (SINGH, 2018).

A camada denominada *batch normalization* foi apresentada pela primeira vez por (IOFFE; SZEGEDY, 2015). Essas camadas são usadas para normalizar as ativações de um determinado volume de entrada antes passando-o para a próxima camada da rede. Em seu trabalho Rosebrock (2017) explica que essa camada tem se mostrado extremamente eficaz na redução do número de épocas que leva para treinar uma rede neural.

Apresenta ainda o benefício adicional de ajudar a “estabilizar” o treinamento, permitindo uma maior variedade de taxas de aprendizagem e intensidades de regularização.

Por fim, o *dropout*, é uma forma eficaz de regularizar as redes neurais para evitar o *overfitting*. (SHANMUGAMANI, 2018) explica que durante o treinamento, o *Dropout* paralisa a rede neural, removendo camadas ocultas aleatoriamente, conforme mostrado na Figura 8. Com isso, é possível prevenir o *overfitting* e aumentar a precisão dos testes.

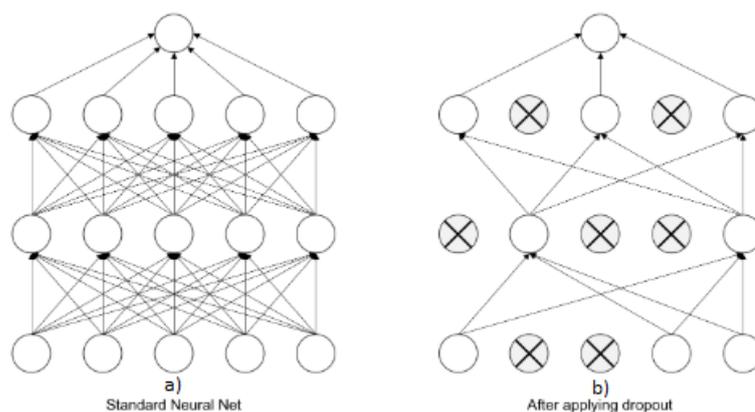


Figura 8 – A rede apresentada em (a) mostra uma arquitetura padrão e (b) mostra as conexões existentes após uso do *dropout*. Adaptado de: (SHANMUGAMANI, 2018)

Na Figura 9 é ilustrado como essas camadas estão conectadas em um CNN. Existem diversas arquiteturas CNNs que empregam essas camadas para tarefa de classificação de imagens. Dentre as principais destacam-se a AlexNet, a VGG16 e a Resnet18, as quais serão investigadas na metodologia.

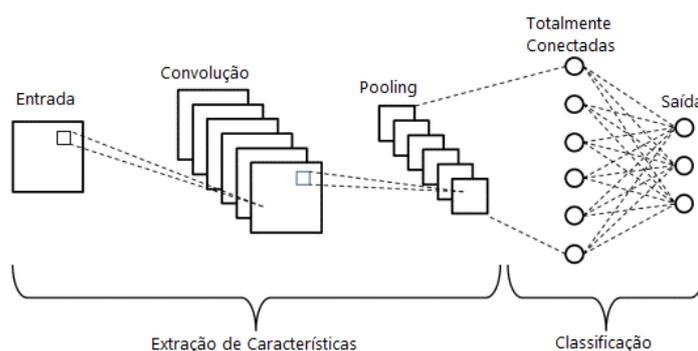


Figura 9 – Exemplos de um CNN com as diversas camadas. Adaptado de: (GURUCHARAN, 2020)

2.4 Transferência de Aprendizado

A transferência de aprendizado é um método muito conveniente e eficaz para treinar redes neurais visando reduzir o tempo gasto no treinamento de grandes modelos

CNN, além de tentar suprir a falta de grandes conjuntos de dados (TAN et al., 2018). Essa técnica consiste em reciclar o conhecimento de modelos previamente treinados com grande volume de dados. As CNN requerem GPU e CPU de alto desempenho para serem treinadas, mas com a transferência de aprendizado, elas podem ser implementadas em computadores pessoais. O treinamento não leva muito tempo, pois geralmente esses processos requerem apenas alguns minutos (LU; LU; ZHANG, 2019). Na transferência de aprendizado, deve-se primeiro treinar uma rede básica, em um conjunto de dados e tarefas básicas e, em seguida, transferir esses recursos aprendidos para uma outra rede para ser treinada em um conjunto de dados e tarefa de destino (YOSINSKI et al., 2014)

Em seu trabalho Tan et al. (2018), classifica o aprendizado de transferência em quatro categorias: baseado em instâncias, baseado em mapeamento, baseada em rede e baseada em contradição. A aprendizagem por transferência baseada em instâncias se sustenta pela suposição de que embora haja diferenças entre dois domínios, instâncias parciais no domínio de origem podem ser utilizadas pelo domínio de destino com pesos apropriados, assim ela se refere ao uso de uma estratégia de ajuste de peso específica, selecionando instâncias parciais do domínio de origem como suplementos ao treinamento definido no domínio de destino.

Uma segunda abordagem é a aprendizagem de transferência baseada em mapeamento. Ela se sustenta na suposição de que, embora existam diferenças entre dois domínios de origem, eles podem ser mais semelhantes em um novo espaço de dados elaborado. Assim mapeia-se o domínio de origem e domínio de destino em um novo espaço de dados. Neste novo espaço, instâncias de dois domínios são semelhantes e adequadas para o treinamento.

Uma terceira forma é a aprendizagem de transferência baseada em rede. Ela se baseia no pressuposto de que a rede neural é semelhante ao mecanismo de processamento do cérebro humano, assim camadas frontais da rede podem ser tratadas como um recurso extrator, e os recursos extraídos são versáteis. Refere-se então a reutilização da rede parcial pré-treinada no domínio de origem, incluindo sua estrutura de rede e conexão de parâmetros, transferindo-a para ser uma parte da rede neural profunda que é usada no domínio alvo. Por fim, a aprendizagem de transferência profunda baseada em contradição, baseia-se no pressuposto de que, para uma transferência eficaz, uma boa representação deve ser discriminativa para a tarefa de aprendizagem principal e indiscriminada entre o domínio de origem e o domínio de destino. Logo, busca encontrar representações transferíveis que são aplicáveis tanto ao domínio de origem quanto ao domínio de destino.

2.5 Trabalhos Relacionados

Dentro da área de saúde oral, as displasias, são investigadas para detectar as características e anomalias que contribuem de forma significativa no desenvolvimento e

detecção de patologias. Dentre soluções apresentadas na literatura, destaca-se o trabalho proposto pelos autores [Krishnan et al. \(2009\)](#), em que um método para classificação automatizada de células de tecidos com Fibrose Submucosa Oral (FSO) foi proposto. Nesse trabalho, 2894 células foram, manualmente, identificadas e classificadas por dois especialistas de forma independente. Foi empregado o algoritmo SVM(do inglês, *Support Vector Machines*) como classificador em que 1054 células foram usadas no conjunto de treinamento e 1840 células para o conjunto de teste utilizando o método de validação cruzada. Destas 730 células eram de amostras normais e 1110 células de FSO. O método proposto pelos autores, apesar de focar em características geométricas como compacidade e excentricidade, obteve uma acurácia de 88,69%.

Outro estudo apresentado por [Krishnan et al. \(2011\)](#) propôs um método para classificação de FSO baseado em informações de textura. Nesse estudo, foram utilizadas 158 imagens, sendo 90 de tecidos normais, 42 de FSO sem displasia e 26 de FSO com displasia. Para retirar as características, os autores usaram dimensão fractal, padrões binários locais e wavelet. Na etapa de classificação, uma variação do classificador SVM, denominada *one-against-one SVM*, foi empregada para caracterizar as imagens entre saudáveis e doentes. O sistema obteve acurácia de 91,67%, 85,71% e 87,76% na classificação de imagens normais, com FSO sem displasia e FSO com displasia, respectivamente.

No ano de 2014, o estudo de [Baik et al. \(2014\)](#) apresentou um método para identificar lesões com alto risco de evoluir para câncer invasivo. Nesse estudo foram calculadas 110 características para medir o tamanho, a forma, a quantidade de DNA nuclear e a distribuição da cromatina, porém somente um subconjunto desses atributos foi utilizado nas análises finais. A descrição desses atributos pode ser encontrada em ([DOUDKINE et al., 1995](#)). Eles utilizaram *10-fold cross validation* para ajustar os grupos de treinamento e teste. Por fim, foram utilizado algoritmos baseados em *Random Forests* para realizar a classificação, obtendo uma acurácia de 80% na classificação dos núcleos celulares.

Os autores em [Adel et al. \(2018\)](#), introduziram uma classificação de FSO usando o algoritmo denominado ORB (do inglês, *Oriented FAST e Rotated BRIEF*), SIFT (do inglês, *Scale Invariant Feature Transform*) e a técnica SURF (do inglês, *Speeded Up Robust Features*) para extração de características e o método SVM (do inglês, *Support Vector Machine*) para classificação. A abordagem proposta alcançou 92,8% de precisão em classificação de displasia epitelial oral. O sistema foi treinado e testado em um total de quarenta e seis casos em imagens com uma magnificação de 100x.

Mais recentemente, os autores em [Silva et al. \(2020\)](#), propuseram uma metodologia para segmentação de núcleos em imagens de tecidos displásicos usando redes neurais convolucionais baseadas em Mask-R CNN. Foram avaliados vários métodos de normalização de corantes e sua influência em relação a etapa de segmentação. O algoritmo foi comparado a outros métodos de segmentação e apresentou resultados relevantes, con-

seguinto segmentar as imagens com acurácia média de 0,887, sensibilidade de 0,762 e especificidade de 0,942 para os tecidos de displasia em diferentes graus da displasia.

Com base nos estudos apresentados de classificação de tecidos da mucosa oral, os resultados demonstram que este tema ainda possuiu grandes desafios pela frente e novas pesquisas. O emprego de CNNs com transferência de aprendizado pode contribuir para a área, e trazer resultados promissores.

3 Metodologia

O desenvolvimento é composto pelas seguintes etapas: o pré-processamento com aumento de dados, transferência de aprendizado baseada em rede, classificação das lesões na redes e construção da interface mobile. Cada uma dessas etapas engloba técnicas de processamento de imagens digitais aplicadas sobre amostras de estruturas histológicas.

3.1 Banco de Imagens

O banco é composto por imagens microscópicas de lesões da língua de camundongos. Essas imagens foram obtidas por meio de experimento in vivo devidamente aprovado pelo Comitê de Ética na utilização de animais sob número de protocolo 038/09 e realizado no período de 2009 a 2010. Essas imagens já foram empregadas nos trabalhos apresentados por (SANT'ANA et al., 2011), (FARIA et al., 2011) e (SILVA et al., 2019).

Na obtenção das imagens foi empregado o protocolo de indução de lesões em língua de camundongos. Nesse caso, foram usados 30 animais pesando entre 20 e 23 gramas, de idade de 6 a 8 semanas e distribuídos de forma aleatória em dois subgrupos denominados de semana 16 e semana 32. Todos os camundongos foram colocados em gaiolas de plástico contendo serragem e mantidos à temperatura de 22° C, ciclo claro/escuro de 12h, no centro de bioterismo da Universidade Federal de Uberlândia. O protocolo obedeceu os critérios estabelecidos por (TANG et al., 2004), como segue: o carcinógeno 4NQO foi previamente diluído em propilenoglicol na concentração de 5 mg/ml e depois em água filtrada até atingir a concentração final de 100 ug/ml. Após o sacrifício de ambos os grupos, as línguas foram removidas e colocadas em solução de formalina tamponada a 4% por 24 horas. Depois desse período, as línguas foram processadas e incluídas em parafina para confecção dos blocos, recortadas e coradas por hematoxilina e eosina (H&E) para estudo histopatológico. Com a ajuda de um patologista, as lâminas foram classificadas em tecido saudável, displasias leve, displasia moderada e displasia grave.

As imagens histológicas digitais foram obtidas com o uso do microscópio óptico Leica DM500, em magnificação de 400x, e armazenadas em arquivos em formato TIFF, utilizando o modelo de cores RGB, com quantificação em 8 bytes e resolução de 2048 × 1536 pixels. Ao todo, 43 lâminas foram digitalizadas e as imagens foram recortadas em regiões de interesse (do inglês, *Region of Interest* - ROI) com tamanho de 450 × 250 pixels, sendo 74 ROI para cada classe de tecido, totalizando 296 ROI. Na Figura 10 é apresentado um exemplo de cada tipo de tecido (displasia leve, moderada e severa e tecido saudável) desse banco.

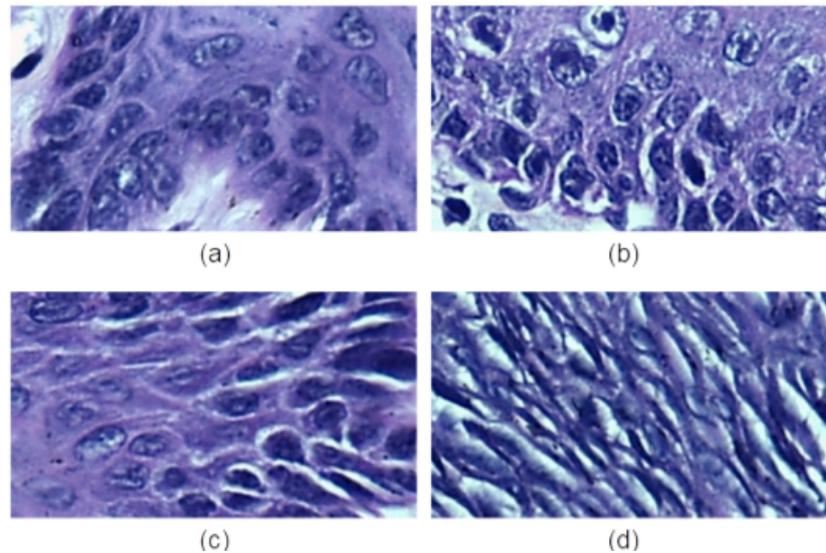


Figura 10 – Exemplo de tecido histológico da cavidade bucal: (a) saudável, (b) displasia leve, (c) displasia moderada e (d) displasia severa. **Adaptado de:** (SILVA et al., 2019)

3.2 Algoritmos para Classificação de Lesões de Displasias

Na Figura 11, é apresentado um fluxograma das etapas empregadas para a classificação de lesões dos diversos níveis de displasia. Os algoritmos das CNNs foram desenvolvidos sem uma interface GUI (do inglês, *Graphical User Interface*) para o usuário. Desta forma, todos os parâmetros utilizados nos códigos foram configurados previamente para execução dos algoritmos. Eles foram desenvolvidos utilizando a linguagem MATLAB® conectado ao *Telegram* (Usuário), por meio do *Node-RED* e do *Mosquitto Broker*. Os experimentos foram realizados utilizando um *ultrabook* (ACER PREDATOR HELIOS 300) com processador de 2,6 GHz (Intel Core I7-9750H) e 32 GB de memória RAM.

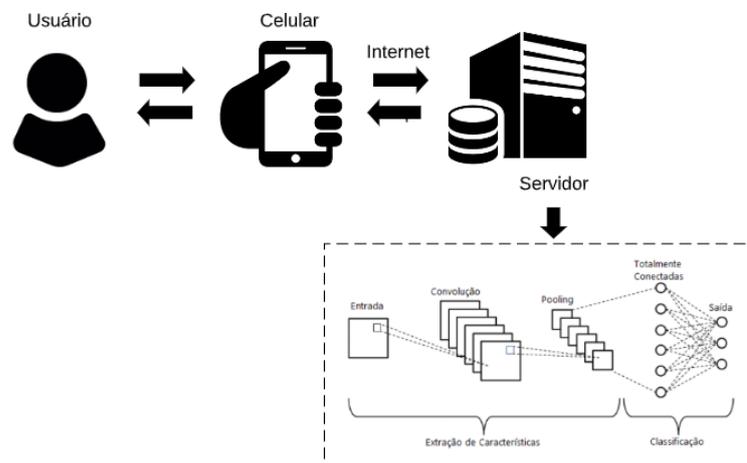


Figura 11 – Arquitetura do método proposto nesse trabalho.

3.2.1 Pré-Processamento

Antes de treinar as redes neurais com o conjunto de dados, é necessário verificar se a quantidade de dados existentes é suficiente para conseguir um margem de erro aceitável. Sabe-se que quanto mais imagens, mais acurada será a rede. Conforme descrito na seção 3.1, o banco apresenta 74 imagens de cada um dos três grupos (Normal, leve, moderado e grave), assim foi empregado um aumento dessa quantidade de dados através da técnica de *data augmentation*.

Para cada imagem foram realizados cortes a partir do centro para redimensionamento da imagem. A AlexNet (seção 3.2.2.1), em sua camada de entrada, requer uma imagem de 256×256 pixels e a Vgg16 (seção 3.2.2.2) ou a Resnet18 (seção 3.2.2.3) requerem uma imagem de 224×224 pixels. Além disso, foram realizadas operações de rotações, inserções de ruído e *color shifting*. Ao final do processo as 74 imagens de cada grupo, resultaram em um grupo de 1406 imagens, totalizando um banco de imagens com 5624 imagens.

3.2.2 Redes Neurais Convolucionais

Feito o tratamento inicial das imagens, a próxima etapa foi investigar a classificação sobre os modelos de CNN. Nessa etapa, com uso da linguagem Matlab, foi empregada a extensão *Deep Network Designer* para elaboração das arquiteturas. Nesse trabalho foram investigadas as arquiteturas AlexNet, Vgg16 e ResNet18.

Com os modelos definidos, o próximo passo é estabelecer os conjuntos de dados para etapa de treinamento, validação e teste. O grupo de treinamento é utilizado pela rede durante o aprendizado e a validação é utilizado para validar esse treinamento. Ao final do processo de aprendizado, se emprega o grupo de teste para verificar a acurácia da rede. Nessa etapa, 30% das imagens foram aleatoriamente selecionadas para compor o grupo de teste e dentre os outros 70% do dados, 90% ficaram para o grupo de treinamento e outros 10% para a validação. Usou-se ainda 6 épocas para o treinamento com 354 iterações por época definidos empiricamente.

3.2.2.1 AlexNet

A AlexNet pertence a uma estrutura CNN profunda, proposta por (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), apresentada no concurso ImageNet ILSVRC-2010 com resultados surpreendentes para classificação de dados. O modelo AlexNet empolgou toda a comunidade com um desempenho superior na classificação em relação aos métodos anteriores. Ela possui, em sua estrutura, 60 milhões de parâmetros e 650.000 neurônios, uma nova função de ativação e melhorias para evitar a sobreposição. Os detalhes da arquitetura podem ser observados na Figura ???. Como mencionado na seção anterior, a função

de ativação é usada em redes neurais para fornecer não-linearidade. Nesse modelo é usada a ReLU como função de ativação, *max-pooling* para a etapa de *pooling* e as camadas FCs e *dropout*.

Essa arquitetura foi desenvolvida sobre um conjunto de dados denominado ImageNet. O ImageNet é um conjunto de dados com mais de 15 milhões de imagens rotuladas de alta resolução pertencentes a cerca de 22.000 categorias. As imagens foram coletadas na web e rotuladas por rotuladores humanos usando a ferramenta de *crowdsourcing Mechanical Turk* da Amazon. Esse subconjunto recebe o nome ILSVRC, e contém cerca de 1000 imagens em cada uma de suas 1000 categorias. Ao todo, existem cerca de 1,2 milhão de imagens de treinamento, 50.000 imagens de validação e 150.000 imagens de teste. A resolução da imagem de entrada da rede foi definida de 256×256 .

Na implementação investigada nesse trabalho, foi empregada uma entrada do tipo imagem de $227 \times 227 \times 3$. Foram usadas as camadas convolucionais padrões da rede, com operadores *max pooling* para etapa de *pooling*. Na camada de ativação, foi empregada a função ReLU. O *dropout* foi definido com 50% e foram mantidas as camadas totalmente contactadas. Já a camada de saída foi adaptada para os 4 tipos de imagens histológicas investigadas.

3.2.2.2 Vgg16

A Vgg16 é um modelo de CNN proposto por (SIMONYAN; ZISSERMAN, 2014). O modelo também foi treinado no banco de imagens ImageNet, porém com imagens de dimensões de $224 \times 224 \times 3$. É considerada uma evolução em relação à AlexNet substituindo grandes filtros do tamanho do núcleo (*kernel*) (11 e 5 na primeira e segunda camada convolucional, respectivamente) por vários filtros 3×3 do tamanho do *kernel* um após o outro. Vgg16 foi treinada por semanas e usando uma GPU NVIDIA Titan Black. Na Figura 12 é apresentada a estrutura dessa rede.

Para essa implementação manteve-se a entrada de dados com tamanho de imagem de $224 \times 224 \times 3$. Todas as camadas convolucionais da estrutura foram preservadas, usando a ReLU como função de ativação e *max pooling* para realizar o *pooling*. Assim como para a AlexNet, o *dropout* foi definido como 50% e as camadas totalmente conectadas foram mantidas com a configuração padrão do MatLab. Já a camada de saída foi adaptada para uma classificação de 4 classes dada de acordo com os tipos de imagens histológicas investigadas.

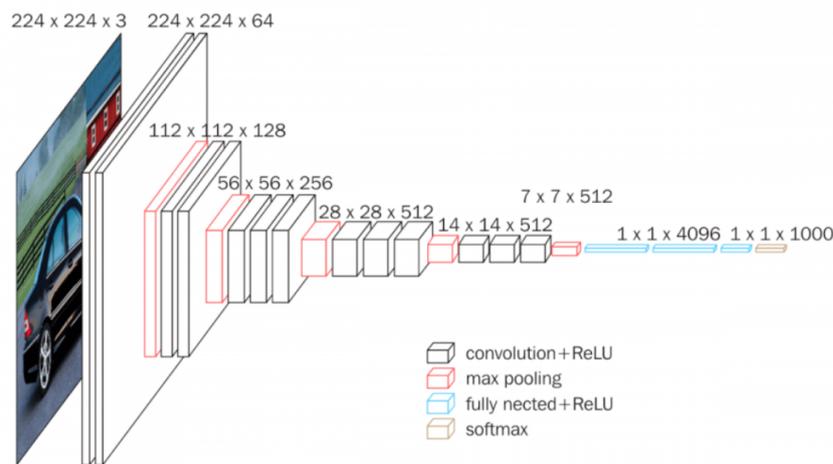


Figura 12 – A estrutura da do modelo de CNN Vgg16 **Adaptado de:** (HASSAN, 2018).

3.2.2.3 ResNet

Um outro modelo muito empregado na literatura que também foi treinada sobre a base ImageNet é a ResNet - *Residual Learning* (HE et al., 2015). Modelos como a AlexNet tinham somente o aumento de camadas, o que provocavam degradação se a rede torna-se demasiadamente profunda e podia saturar a acurácia. Deste modo, esse modelo usa o conceito de blocos residuais, que consistem no uso de “conexões de atalho”. A sua principal particularidade é a capacidade de saltar algumas camadas, tornando o processo de treinamento consideravelmente mais rápido. Na Figura 13 é representado esse processo. (HE et al., 2015) apresentou diferentes arquiteturas baseadas nesse conceito, o que pode variar de 18 até 1202 camadas. Nesse estudo foi empregado o modelo ResNet18 devido aos relevantes resultados e a necessidade de recursos computacionais para um elevado número de camadas.

Para essa implementação manteve-se a entrada de imagens com o tamanho de 224x224x3. De forma semelhante aos outros modelos, nessa arquitetura foram empregadas a função de ativação ReLU, o método *max pooling* para as camadas de *pooling* e finalmente as camadas FCs com modificação na saída para quatro classes de imagens histológicas (tecidos com displasia e saudáveis).

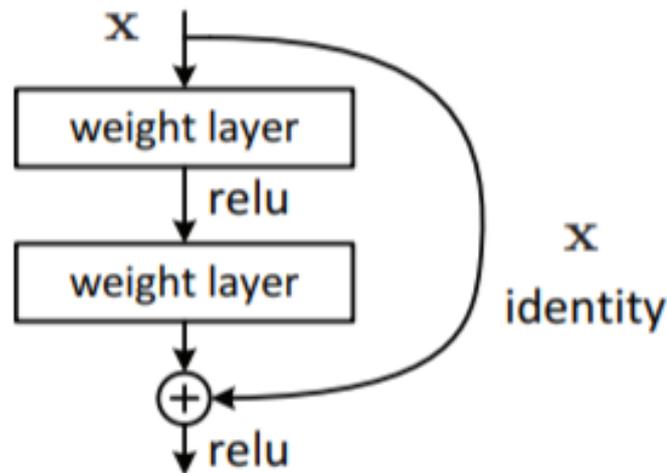


Figura 13 – Exemplo do processo do processo residual da arquitetura ResNet. **Adaptado de:** (HE et al., 2015)

3.2.3 Transferência de Aprendizado

As CNNs têm uma dependência muito forte na fase de treinamento de dados em massa em relação aos métodos tradicionais para entender os padrões latentes dos dados. O uso de transferência de aprendizado é uma solução para esse problema. Uma das formas de uso de transferência de aprendizado, conforme descrito na seção 2.4, é baseada em rede, ou seja, emprega a reutilização da rede parcial pré-treinada no domínio da fonte, incluindo sua estrutura de rede, conexão e parâmetros, transferindo para utilização em um outro domínio. Esse modelo apenas ocorre uma adaptação na última camada para o número de classes empregadas ao domínio alvo. Na Figura 14 é apresentado um exemplo desse processo de adaptação para transferência de aprendizado com a AlexNet, a qual foi empregada em todos os modelos investigados.

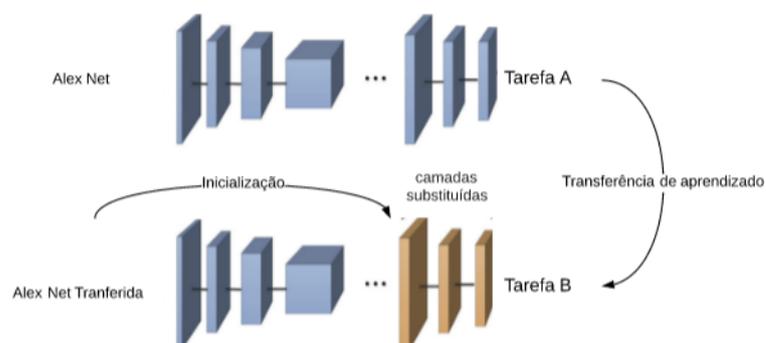


Figura 14 – Transferência de aprendizado por uma rede neural convolucional *AlexNet*. **Adaptado de:** (LU; LU; ZHANG, 2019)

3.3 Avaliação de Desempenho

a métrica escolhida para avaliar o desempenho dos treinamento foi a acurácia. A acurácia é um conceito amplamente conhecido que determina o quão próximo um resultado está em relação ao seu valor de referência real. A equação 3.1 demonstra como essa medida é calculada. Dessa forma, quanto maior a acurácia, melhor será o desempenho da rede.

$$Acurácia = \frac{N \text{ de amostras corretas}}{N \text{ total de amostras}} \quad (3.1)$$

3.4 Aplicação para Dispositivos Móveis

Após o treinamento da rede e a validação, o próximo passo consiste em configurar a interface para celulares pelo o qual o usuário irá se comunicar com essa solução de classificação de imagens. Nesse projeto foi desenvolvido uma interface via *chat-bot* do *Telegram* que se comunica com o software *Matlab*. Nesse processo também foi empregado o protocolo *Message Queuing Telemetry Transport (MQTT)* e a plataforma *Node-RED*, a qual permite conectar os dispositivos. Na Figura 15 é ilustrado um exemplo do processo para comunicação entre o *Telegram* e o software *MatLab*.



Figura 15 – Arquitetura da solução para dispositivo móvel para classificação das imagens histológicas.

3.4.1 Telegram

O Telegram é um aplicação de mensagens para dispositivos móveis (celular, *Tablet*) que possui uma versão para computador (*desktop e web*), com arquitetura é baseada em nuvem com foco em segurança e velocidade. Esse serviço está disponível gratuitamente. Nele os usuários podem fazer chamadas com vídeo, enviar mensagens e trocar fotos, vídeos e arquivos de qualquer tipo (TELEGRAM, 2021).

Sua principal vantagem é a possibilidade de construir *apps e bots* de uma forma fácil e rápida. Assim, essa plataforma apresenta um grande potencial para explorar uma conexão simples e fácil, utilizável diretamente na palma da mão do usuário.

Na Figura 16 é ilustrado como ocorre esse procedimento para criação do bot. O software BotFather retorna um *token* para acessar com o protocolo HTTP API. Esse *token* é necessário para configurar as respostas do bot assim como também tratar as informações recebidas do usuário. O bot não possui restrição de acesso, assim qualquer

peessoa pode enviar uma mensagem ferramenta para análise de imagens (acesse o [link](#) para avaliar essa ferramenta).

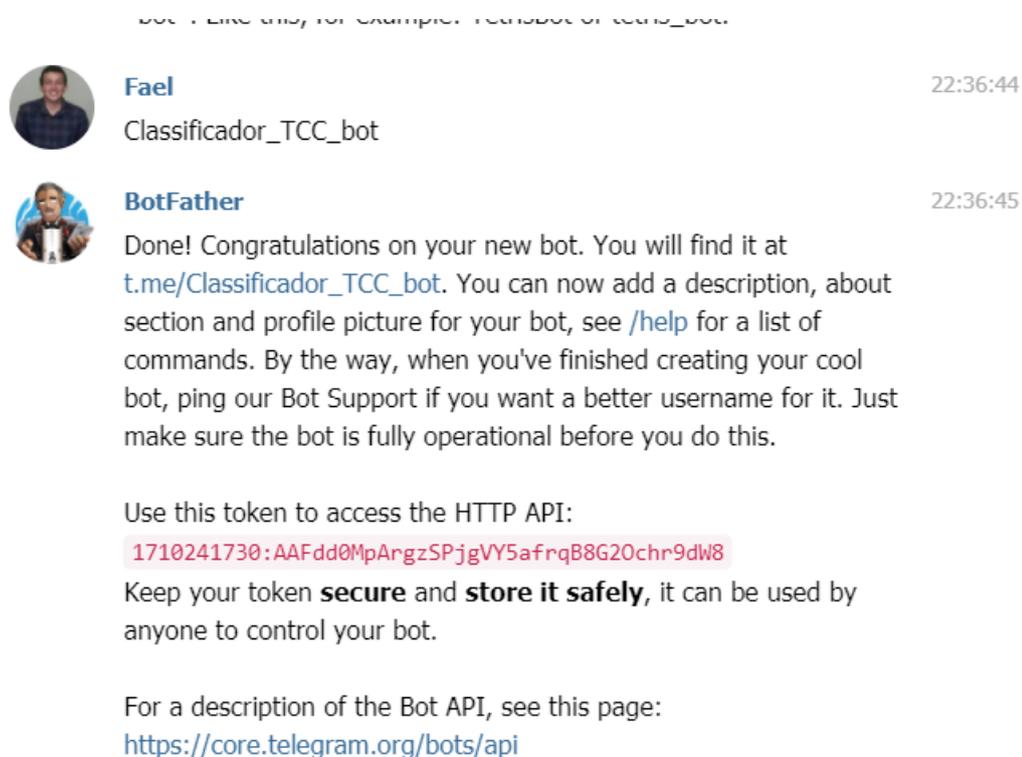


Figura 16 – Tela de criação do bot para o aplicativo Telegram com o software BotFather.

3.4.2 Node-RED

O softwar Node-RED é uma ferramenta de programação para conectar dispositivos de hardware, APIs e serviços online. Ele fornece um editor baseado em navegador que torna mais fácil conectar fluxos usando a ampla gama de nós na paleta que podem ser implementados em seu tempo de execução com um único clique (NODE-RED, 2021). Seu funcionamento se dá através de fluxo e nós (*nodes*), que são as funções do programa. É muito utilizado em automação residencial devido a facilidade de comunicação com uma Raspberry Pi.

Então, uma mensagem enviada por meio do bot do Telegram é processada por alguns *nodes* e quando o usuário envia uma imagem, essa mensagem passa por um *node* específico(Mqtt out) seguindo para um MQTT Broker.

3.4.3 Comunicação MQTT

O protocolo MQTT fornece um método de execução de mensagens usando um modelo de publicação e assinatura. Isso o torna adequado para mensagens da internet

das coisas (*IoT*), como sensores de baixa potência ou dispositivos móveis, como telefones, computadores integrados ou microcontroladores.

A ideia é que esse protocolo permita que um cliente se conecte a um broker, o qual é responsável por tratar as publicações e subscrições do protocolo MQTT. A conexão ocorre por TCP/IP simples ou por TLS criptografada para mensagens confidenciais. A aplicação cliente publica mensagens em um tópico específico, enviando a mensagem e o tópico ao broker. Por fim, o broker encaminha a mensagem a todas as aplicações clientes que assinam esse tópico.

3.4.3.1 Eclipse Mosquitto MQTT Broker

O Eclipse Mosquitto é broker de mensagens, *software* livre que implementa o protocolo MQTT versões 5.0, 3.1.1 e 3.1. Esse é um software leve e adequado para uso em todos os dispositivos. Fornece uma biblioteca C para implementar clientes MQTT e os clientes MQTT de linha de comando `mosquitto_pub` (publisher) e `mosquitto_sub` (subscriber) ([MOSQUITTO, 2021](#)).

Esse software é responsável por garantir o funcionamento do protocolo MQTT, rodando localmente na máquina. Assim, ao ser publicada no mosquitto broker pelo Node-RED com um tópico, por exemplo, *Nova Imagem*, a imagem é capturada pelo MatLab. Essa conexão ocorre por meio da função *MQTT in MATLAB*, ([MATHWORKS, 2019](#)).

O MatLab por sua vez, estando um modo de espera, recebe a imagem, processa avaliando e classificando e publica no broker o resultado com outro tópico, o qual representa a resposta da requisição. Como o Node-RED está inscrito, ele recebe a mensagem e envia para o Bot para escrever a resposta ao usuário. No caso a resposta enviada é a classe e a taxa de acerto em relação a imagem. A representação desse processo é mostrada através do Algoritmo no 1.

Algoritmo 1: Comunicação para Acesso Remoto ao Classificador de Imagens

Entrada: Mensagem no Telegram-Bot.

Saída: Resposta no Telegram-Bot.

início

para *Cada mensagem recebida faça*

if *mensagem é string then*

1. Envia essa mensagem para o Node-RED
2. Verifica a Resposta adequada para a mensagem
3. Retorna a Mensagem para o Telegram
4. Mensagem exibida a Tela no Telegram-BOT

end

if *mensagem é imagem then*

1. Envia essa imagem para o Node-RED.
2. O Node-RED envia essa imagem para o broker.
3. O Matlab captura essa imagem e faz classificação.
4. Retorna a uma mensagem ao broker com o resultado da classificação.
5. Node-RED recebe essa mensagem do broker.
6. Node-RED manda essa mensagem para o Telegram.
7. Resultado é exibido na tela do usuário.

end

fim

fim

4 Resultados

Os experimentos para avaliação da classificação proposta são apresentados e discutidos neste capítulo. Para isso, foram realizadas avaliações visuais do aumento de dados, quantitativas dos resultados obtidos pelos métodos propostos e uso da aplicação de acesso remoto.

4.1 Classificação das Imagens Histológicas com CNN

Inicialmente, antes da aplicação dos modelos CNN, essa etapa foi realizada para o aumento de dados. Na Figura 17, é apresentado o resultado de uma imagem da classe normal submetida ao algoritmo de aumento de dados. Nessa figura podem ser observado a imagem original no centro e os exemplos das transformações empregadas. Ao final desse processo, para cada imagem, foram geradas outras 19 imagens "sintéticas" definidas pelas transformações propostas. Assim, de 74 imagens de cada classe foram gerados 1406 imagens com as operações de rotações, inserções de ruído e color shifting.

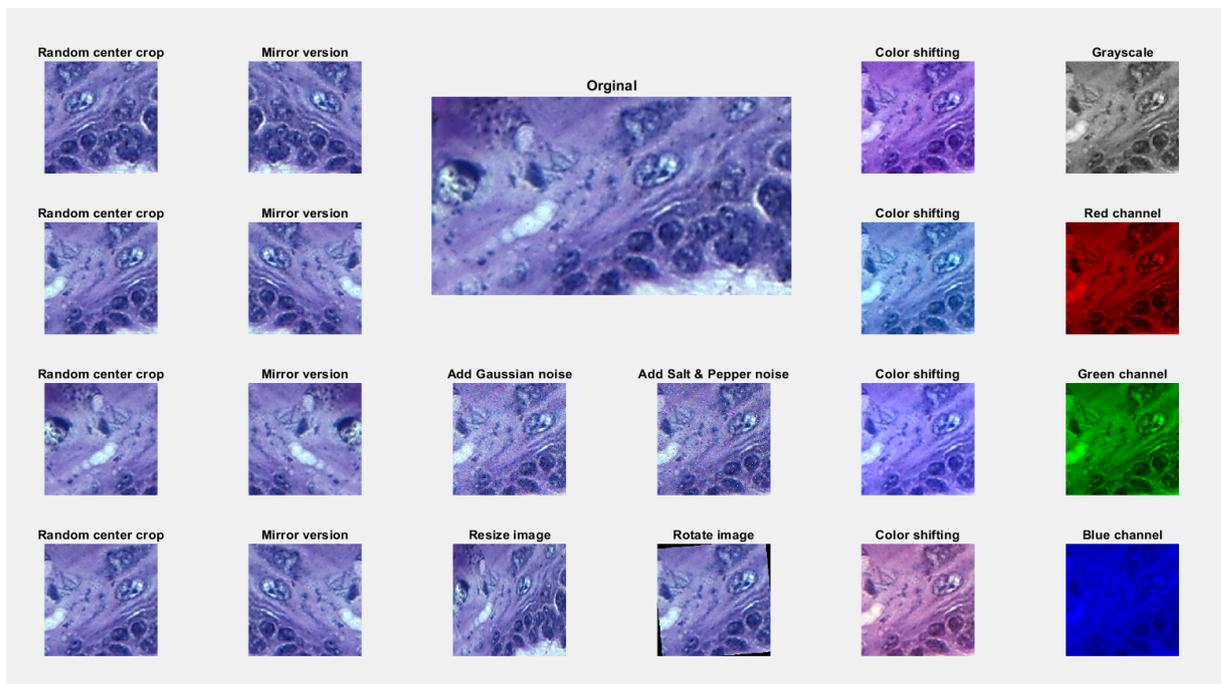


Figura 17 – Representação de imagens sintéticas geradas pela técnica de aumento de dados.

Feito o aumento de dados, de forma aleatória definiu-se um grupo de treinamento, 3.544 imagens, validação, 392 imagens, e teste, 1688 imagens. A fim de avaliar compara-

tivamente o resultado das diferentes redes neurais, AlexNet, VGG16 e Resnet18, e com a intenção de evitar o *overfitting*, definiu-se a mesma quantidade de épocas para o treinamento. Baseado em resultado empíricos usou-se 6 épocas e 354 iterações por época, totalizando 2124 iterações.

Nas Figuras 18, 19, 20 são apresentados os resultados da etapa de treinamento dos modelos. Para cada figura, o gráfico superior representa a acurácia do treinamento ao longo do tempo (épocas), enquanto a imagem inferior indica o valor da função de perda. A função de perda tem o objetivo de otimizar os valores dos parâmetros em um modelo de rede neural, indicando quão adequados são esses parâmetros para a tarefa da classificação. Essas Figuras mostram que a partir da terceira época todos os modelos forneceram uma acurácia superior a 85% para o grupo de validação (representado pela linha preta nos gráficos dessas figuras). Ainda ao final da sexta época, percebe-se que todas obtiveram uma resultado acima de 95% de acurácia para os modelos.

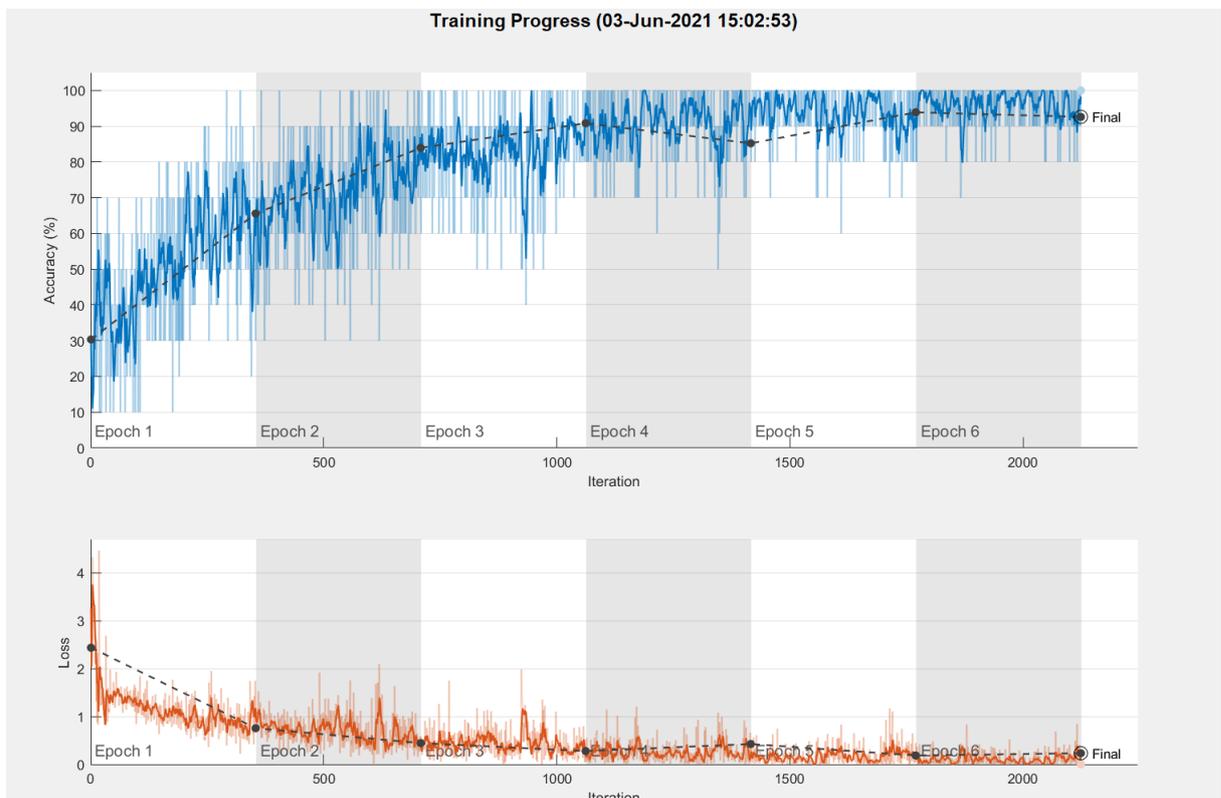


Figura 18 – Representação da etapa de treinamento utilizando o modelo AlexNet: avaliação da acurácia (imagem superior) e função de perda (imagem inferior).

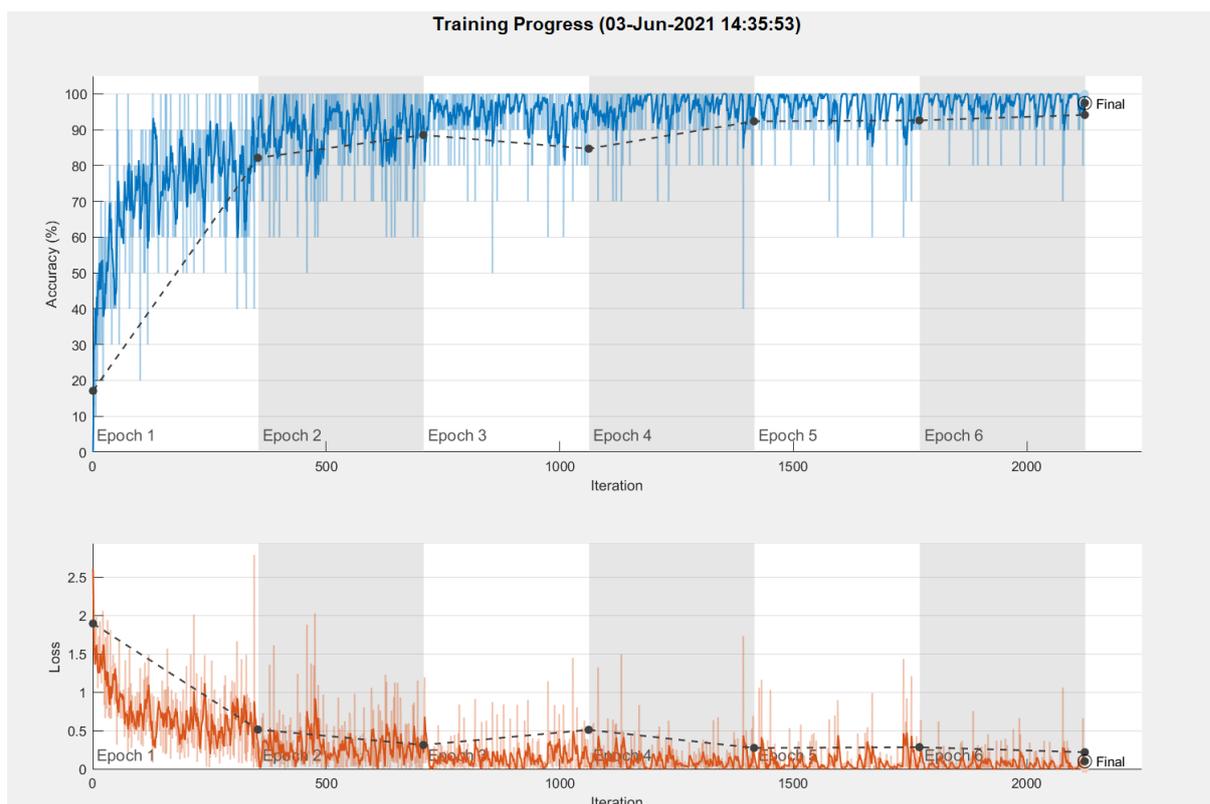


Figura 19 – Etapa de treinamento utilizando o modelo ResNet: avaliação da acurácia (imagem superior) e função de perda (imagem inferior).

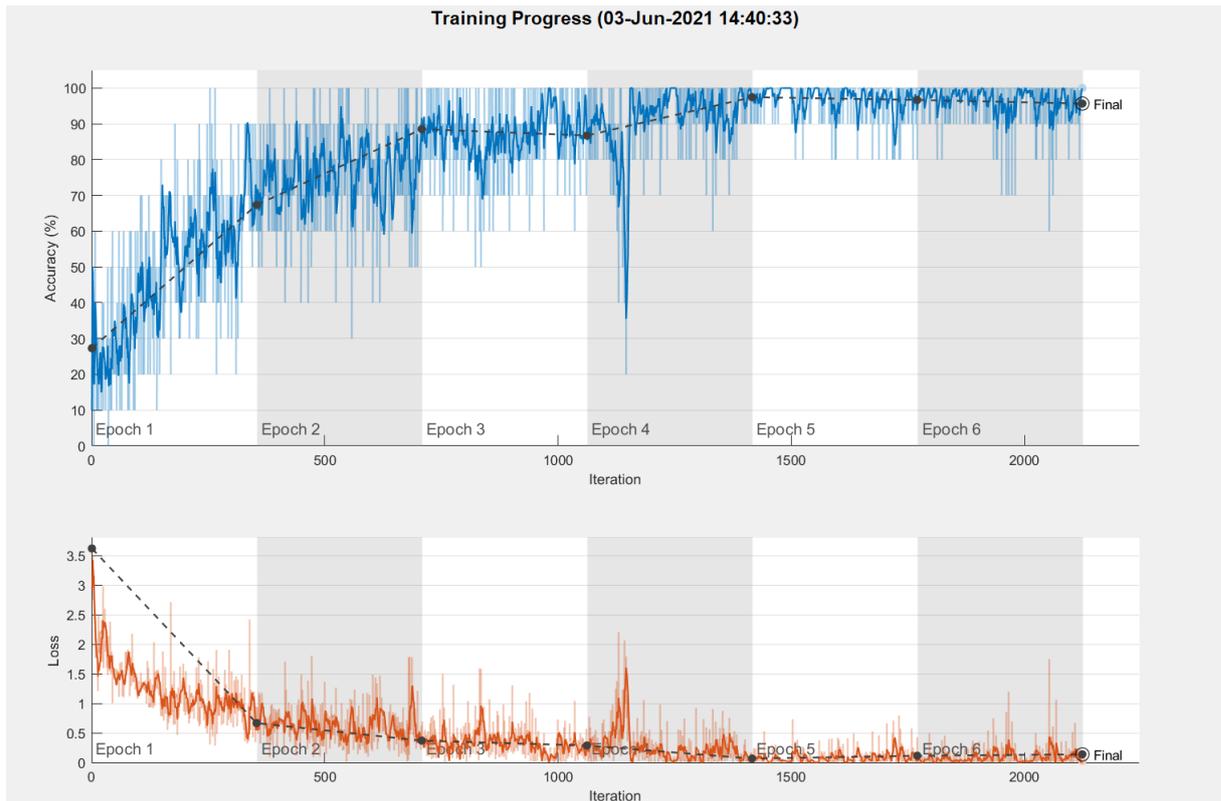


Figura 20 – O modelo da VGG na etapa de treinamento: a acurácia (imagem superior) e função de perda (imagem inferior).

A Tabela ?? mostra os resultados nas etapas de validação e teste e tempo de processamento durante a fase de treinamento para os três modelos investigados. Nessa tabela, nota-se que o tempo de treinamento varia bastante conforme a complexidade do modelo da rede investigada. O treinamento da Alexnet, por exemplo, durou apenas 1:45 minutos enquanto a VGG16 levou 12:37 minutos para treinar a classificador das imagens. Nota-se ainda que a ResNet obteve um tempo intermediário entre os modelos investigados nesse treinamento.

Em relação a classificação, a Resnet obteve uma acurácia de 96,56% para o grupo de teste, sendo o melhor resultado obtido em relação a métrica investigada. O modelo AlexNet foi o que apresentou o resultado mais baixo entre os modelos investigados (90,28%) para etapa de teste. No entanto, esses resultados médios obtidos com esses modelos são valores muito bons para um processo de classificação de diferentes grupos de lesões investigadas.

As redes de aprendizagem profunda são frequentemente consideradas "caixas pretas" que não oferecem nenhuma maneira de descobrir o que uma rede aprendeu. Porém, o mapeamento de ativação de classe (do inglês, *class activation maps* - CAM) é uma técnica usada para obter explicações visuais das previsões de redes neurais convolucionais e podem identificar tendências no conjunto de treinamento para aumentar a precisão do

modelo ao identificar alguns problemas que podem não caracterizar a imagem. Com esse objetivo, nessa etapa foram investigadas essas informações sobre o modelo que proporcionou o melhor desempenho na etapa de teste (ResNet). Para cada classe a ser classificada pela CNN, os mapas de ativação indicam quais partes da imagem são importantes na classificação. Então, depois de treinada e os pesos das características atualizados, um mapa de calor da classe prevista foi gerado ao criar o CAM. Nesse caso, essa informação foi capturada na última camada de ativação, que recebe o nome de *res5b_relu*.

Apenas o mapa de calor das classes permite entender quais regiões espaciais são efetivas na tomada de uma decisão certa ou errada da rede. Nas Figuras 21, 22, 23 e 24 são apresentados os resultados desse mapeamento para a Resnet18. Cada imagem apresenta um detalhe diferente para ser classificado. Na imagem normal (Figura 21), por exemplo, nota-se que a rede usou uma parte específica da imagem que possui um núcleo em destaque, enquanto que para a classe moderada (Figura 23), a rede usou a parte superior da imagem para classificá-la. Essas imagens mostram que a região interna dos núcleos estão sendo utilizadas para tomada de decisão, o que muitas vezes é uma das regiões observadas pelo especialista.

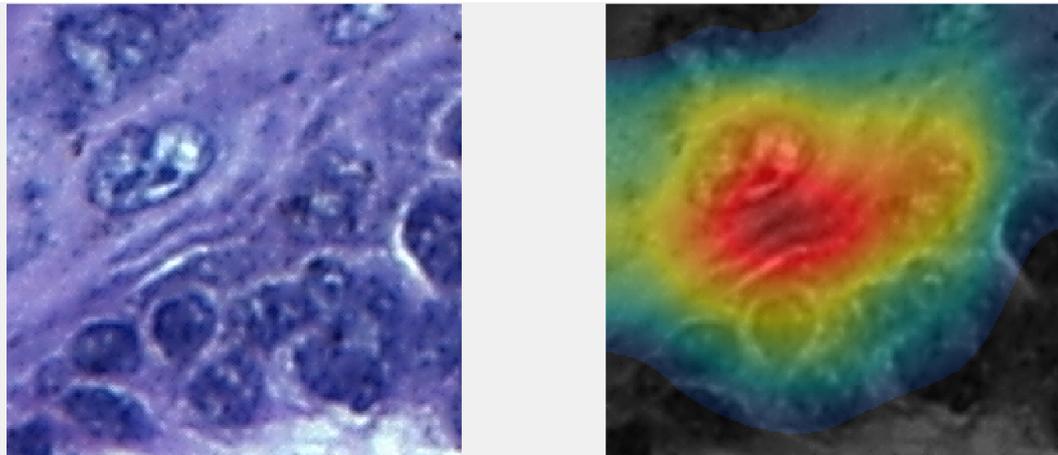


Figura 21 – A representação de CAM (direita) para uma imagem da classe normal (esquerda) empregando a Resnet18 para a classificação.

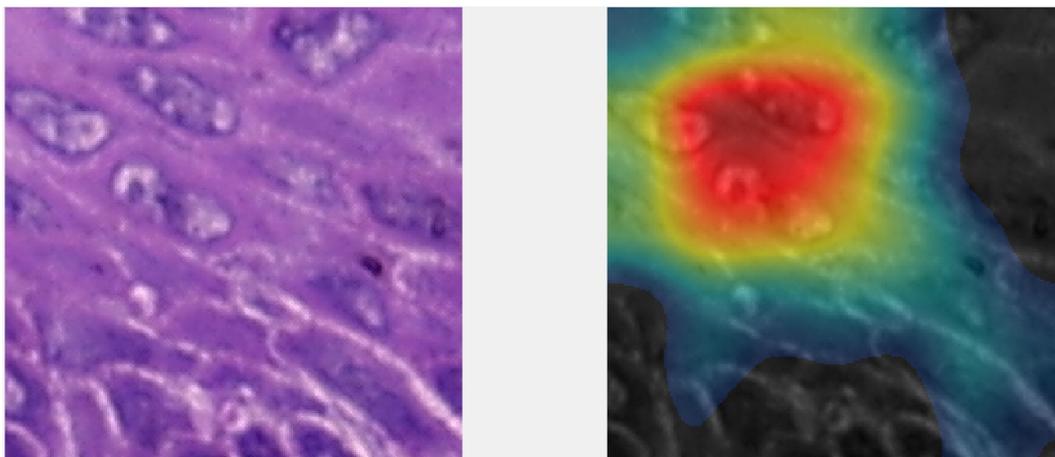


Figura 22 – Imagem do displasia leve (esquerda) e mapeamento de ativação de classe de uma imagem com displasia leve usando a Resnet18 (direita).

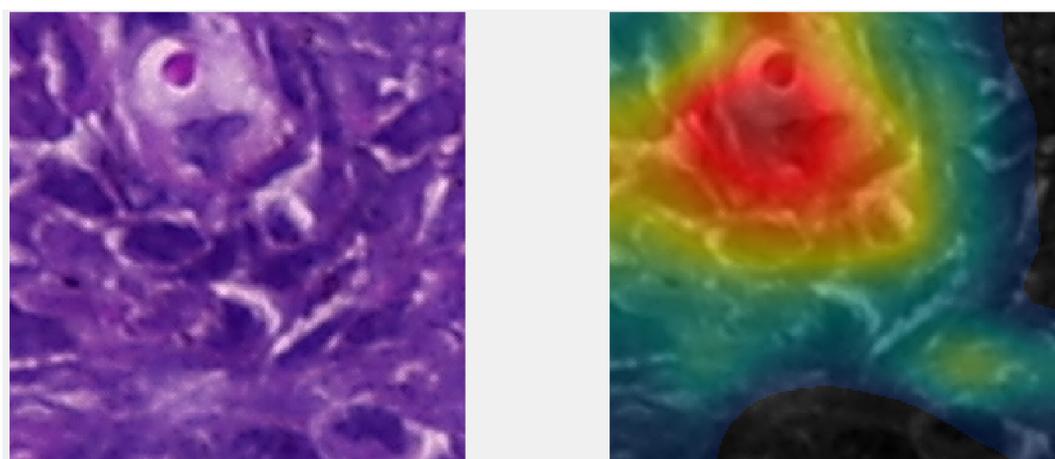


Figura 23 – Representação da CAM (esquerda) de uma imagem da categoria moderado (direita) usando a Resnet18.

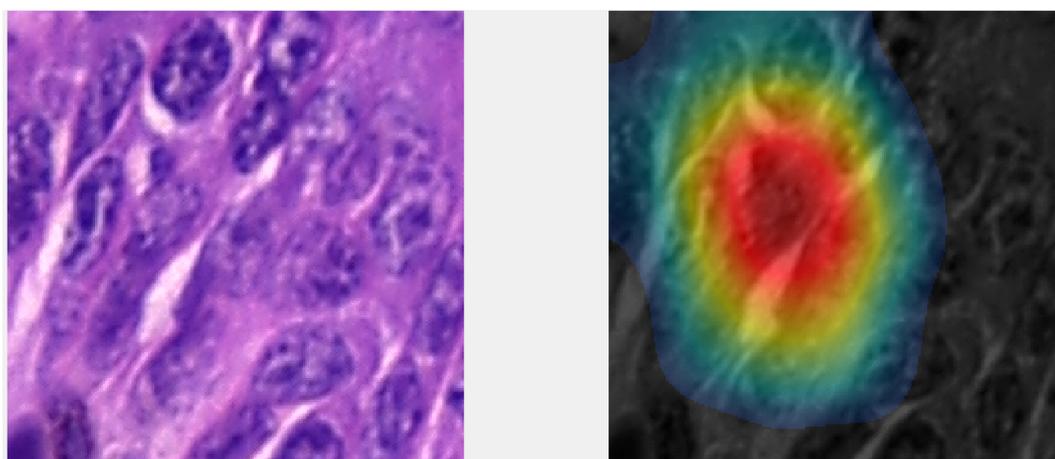
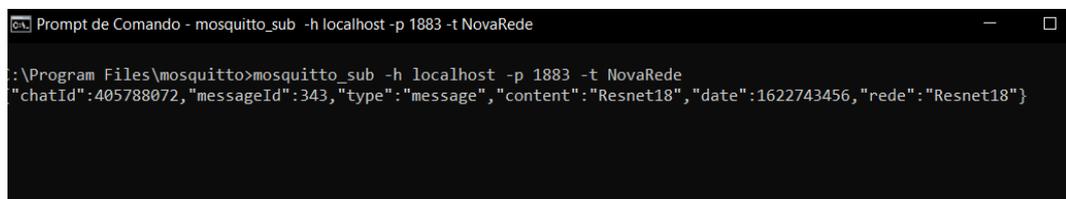


Figura 24 – Imagem de classe grave (esquerda) e mapeamento de ativação de classe (direita) usando o modelo Resnet18 para classificação.

4.2 Aplicação em Dispositivo Móvel

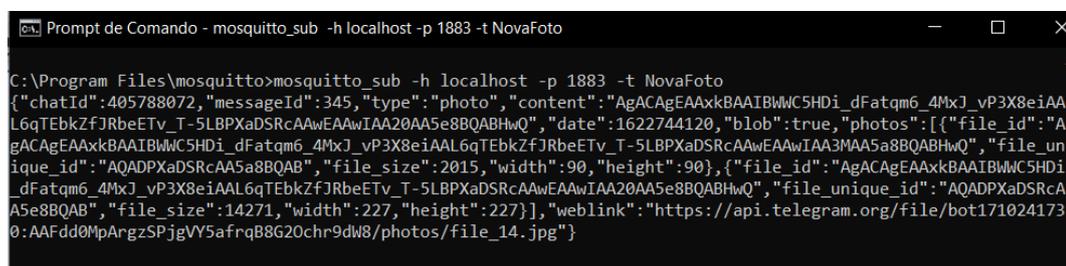
Com criação do bot, na Figura 27 é apresentado o fluxo construído no Node-RED para o tratamento das informações recebidas. Esse figura mostra que uma nova mensagem é recebida pelo *Node Telegram Receiver*, a mensagem entra em um comutador que verifica o tipo da mensagem. Caso a informação seja um texto, segue o caminho inferior, e ao contrário, uma imagem segue o caminho superior. Quando a informação é uma imagem, esses dados são enviados ao Broker com o tópico *NovaFoto*. Já o texto segue para outro lado para verificação do conteúdo da mensagem. Dependendo desse conteúdo, é definida a mensagem resposta e esta é enviada através do node *Telegram Sender*. Vale ressaltar que, se o conteúdo da mensagem for um dos modelos CNNs, essas informações são enviados ao Broker com o tópico *NovaRede*. O Broker foi definido com suas configurações padrões sem a utilização de senha para assinar os tópicos.

Na Figura 25 é ilustrado o resultado do funcionamento do Broker para o tópico *NovaRede*. Enquanto na Figura 26 é apresentado o resultado do funcionamento para o tópico *NovaFoto*. Nessas figuras pode-se observar que o conteúdo da mensagem enviada pelo Node-RED é uma json-string. Finalmente, o software Matlab recebe as informações do Broker e utiliza as CNNs pré-treinadas para classificar e enviar ao usuário a informação com o tópico *MatlabResposta*, conforme ilustrado na Figura 27.



```
Prompt de Comando - mosquito_sub -h localhost -p 1883 -t NovaRede
C:\Program Files\mosquito>mosquito_sub -h localhost -p 1883 -t NovaRede
{"chatId":405788072,"messageId":343,"type":"message","content":"Resnet18","date":1622743456,"rede":"Resnet18"}
```

Figura 25 – Exemplo de funcionamento do Broker com o tópico *NovaRede*



```
Prompt de Comando - mosquito_sub -h localhost -p 1883 -t NovaFoto
C:\Program Files\mosquito>mosquito_sub -h localhost -p 1883 -t NovaFoto
{"chatId":405788072,"messageId":345,"type":"photo","content":"AgACAgEAAxkBAABWwC5Hdi_dFatqm6_4MxJ_vP3X8eiAAL6qTEbkZfJRbeETv_T-5LBPXaDSRcAAwEAAwIAA20AA5e8BQABHwQ","date":1622744120,"blob":true,"photos":[{"file_id":"AgACAgEAAxkBAABWwC5Hdi_dFatqm6_4MxJ_vP3X8eiAAL6qTEbkZfJRbeETv_T-5LBPXaDSRcAAwEAAwIAA3MAA5a8BQABHwQ","file_unique_id":"AQADPXaDSRcAA5a8BQAB","file_size":2015,"width":90,"height":90},{"file_id":"AgACAgEAAxkBAABWwC5Hdi_dFatqm6_4MxJ_vP3X8eiAAL6qTEbkZfJRbeETv_T-5LBPXaDSRcAAwEAAwIAA20AA5e8BQABHwQ","file_unique_id":"AQADPXaDSRcAA5e8BQAB","file_size":14271,"width":227,"height":227}],weblink":"https://api.telegram.org/file/bot1710241730:AAFdd0MpArgzSPjgVY5afrrqB8G20chr9dW8/photos/file_14.jpg"}
```

Figura 26 – Operação do Broker com o tópico *NovaFoto*

Nas Figuras 28 e 29 são mostrados a interação do bot com o usuário bem como suas respostas. A Figura 28 apresenta a interação inicial do usuário com a escolha de operações para uma classificação das imagens. Já na Figura 29 tem-se o processo de submissão da imagem e a resposta em relação a classificação que ocorre no servidor com as CNNs.

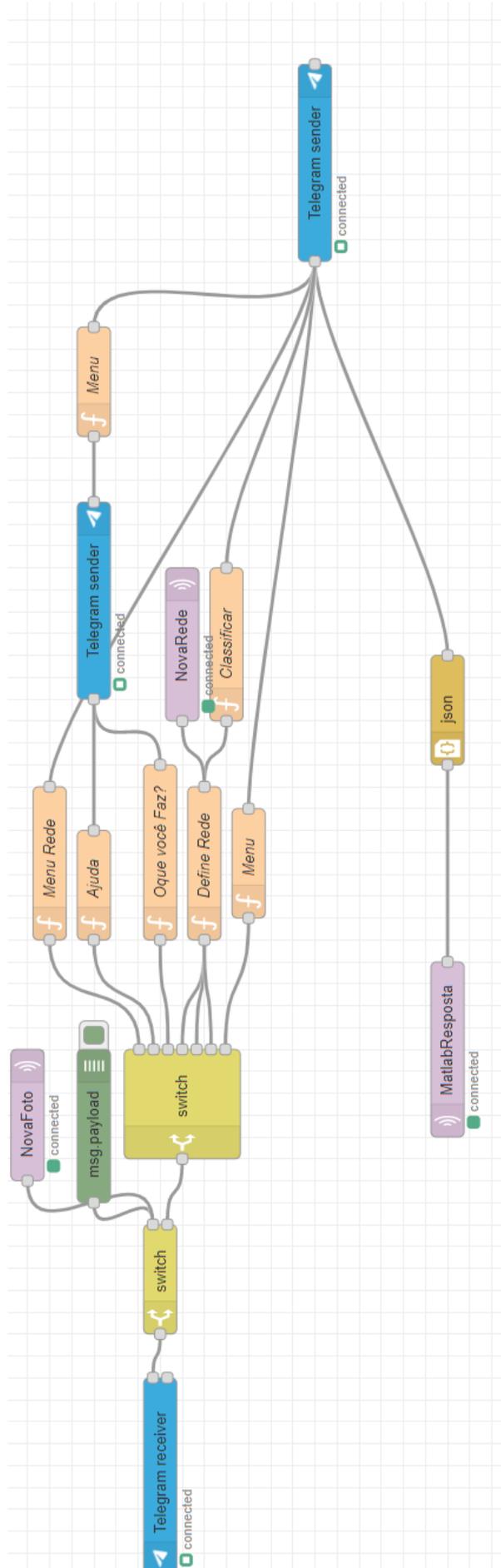


Figura 27 – Resultado do tratamento do fluxo das informações recebidas no Node-RED.

Nesse exemplo foi inserida uma imagem da classe moderada e o algoritmo classificou com taxa de acurácia de 99.99%. O método gasta 10 segundos aproximadamente para realizar toda a operação

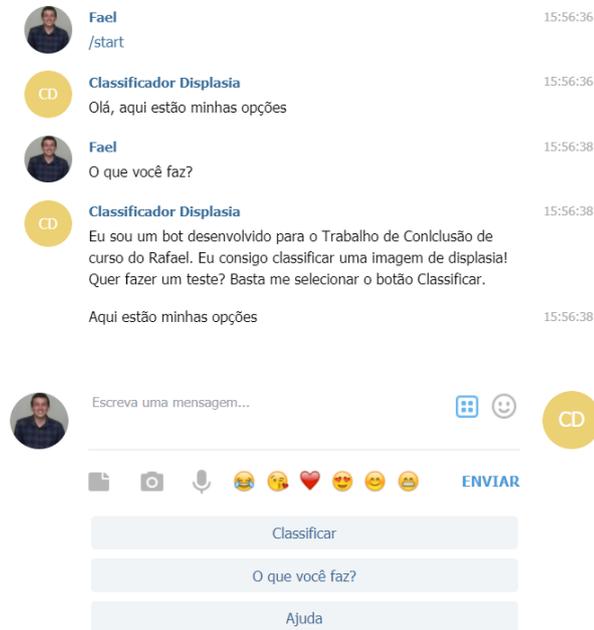


Figura 28 – Exemplo de funcionamento do bot na interação do usuário para escolha das opções.

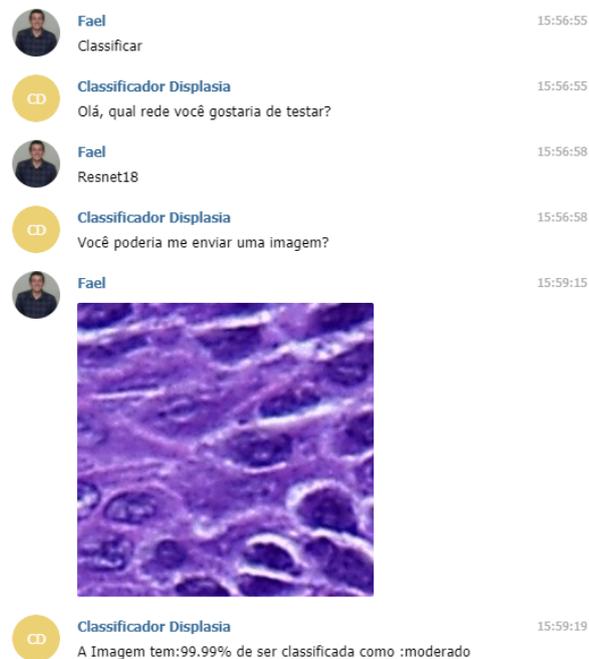


Figura 29 – Tela com informações da submissão de uma imagem e o processo de retorno do servidor com o valor de acurácia para uma imagem histológica.

5 Conclusão

Neste trabalho, foi desenvolvido um método computacional utilizando CNN e transferência de aprendizado para analisar imagens histológicas da cavidade bucal. As redes foram treinadas e avaliadas localmente em relação a métrica acurácia para a fase de teste. Ao final dessa etapa, foi desenvolvida uma aplicação para dispositivos móveis para submissão de imagem para análise e classificação das lesões histológicas. Definiu-se aplicar transferência de aprendizado para os três modelos CNN's bem conhecidos pelos seus ótimos resultados em análise de imagens com a base de imagens ImageNet. Para a solução dos dispositivos móveis, empregou-se um bot do aplicativo Telegram, comunicando com o Node-RED, que por sua vez usa-se do software mosquitto broker para se comunicar com o Matlab, o qual tinha as arquiteturas de redes implementadas para avaliação e classificação das lesões, permitindo estabelecer comunicação de envio e recebimento de mensagens.

No treinamento das redes, obteve-se resultados relevantes, atingindo 96,56% de acurácia com a ResNet18 e 94,33% com a VGG16. Todas as redes testadas tiveram resultados superiores a 90% para a fase de testes, o que pode contribuir com uma aplicação remota via dispositivos móveis para uso da área médica. Apesar dessa solução ser executada em um ambiente controlado, a solução do dispositivo móvel traz facilidade de acesso para a comunidade científica e acadêmica da UFU, uma vez que qualquer pessoa pode facilmente interagir com bot e enviar uma imagem para classificação. A velocidade de resposta de comunicação para análise da imagem é satisfatória, pois o algoritmo demanda apenas 10 segundos para iniciar a interação e devolver a resposta.

As principais metas do trabalho foram alcançadas por meio de um estudo introdutório à histologia, necessário para compreender o problema abordado. Além disso, o conteúdo de disciplinas do curso da Engenharia Mecatrônica contribuíram para essa finalidade, como a Programação Aplicada à Engenharia, que possibilitou o aprendizado da linguagem utilizada no Matlab, a Inteligência Computacional e Inteligência Artificial, que introduziram os modelos de redes neurais convolucionais, Sistemas Embarcados, que possibilitou o uso do Node-RED com o Telegram em automações residenciais e, a disciplina de Redes de Computadores, que introduziu o protocolo MQTT.

5.1 Desafios Encontrados

Durante o desenvolvimento do projeto, foram encontradas algumas dificuldades, das quais se destacam: o entendimento de tópicos relacionados a histologia, em razão da falta de experiência e conhecimento prévio da área; integração de todos os softwares utilizados, devido a falta de conhecimento das ferramentas.

Esses desafios serviram como grandes oportunidades para se reinventar e tornar a produto final utilizável. Isso permitiu ainda, em validar a importância de várias práticas e conceitos vistos ao decorrer do curso. Além disso, a superação dos desafios encontrados contribuiu para a consolidação de uma aplicação cada vez mais robusta, que pode colaborar para área interdisciplinar de biologia em conjunto com a computação.

5.2 Trabalhos Futuros

O método proposto apresentou resultados promissores para classificação das lesões. Em trabalhos futuros, pretende-se investigar novas arquiteturas CNN's e aprimorar a aplicação de dispositivos móveis para torná-la mais robusta e mais *user friendly* possível em um ambiente de domínio público para toda a comunidade UFU.

Referências

- ABBAS, A. K. et al. *Robbins Patologia Básica*. [S.l.]: Elsevier Brasil, 2008. Citado 2 vezes nas páginas 10 e 14.
- ADEL, D. et al. Oral epithelial dysplasia computer aided diagnostic approach. In: IEEE. *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. [S.l.], 2018. p. 313–318. Citado na página 24.
- ARAÚJO, T. et al. Classification of breast cancer histology images using convolutional neural networks. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 6, p. e0177544, 2017. Citado na página 11.
- AXÉLL, T. Occurrence of leukoplakia and some other oral white lesions among 20 333 adult swedish people. *Community dentistry and oral epidemiology*, Wiley Online Library, v. 15, n. 1, p. 46–51, 1987. Citado na página 10.
- BAIK, J. et al. Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms. *Cellular Oncology*, Springer, v. 37, n. 3, p. 193–202, 2014. Citado na página 24.
- BIANCHI, A. L. *Entendendo de vez a convolução: base para processamento de imagens*. 2020. Disponível em: <<https://tinyurl.com/4xzn3v73>>. Acesso em: 19/5/2021. Citado 3 vezes nas páginas 4, 19 e 20.
- BLOCK, H. A review of “perceptrons: An introduction to computational geometry. *Information and Control*, Academic Press, v. 17, n. 5, p. 501–522, 1970. Citado na página 18.
- CHOLLET, F. et al. *Deep learning with Python*. [S.l.]: Manning New York, 2018. v. 361. Citado 3 vezes nas páginas 4, 16 e 17.
- CRUZ-ROA, A. et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific reports*, Nature Publishing Group, v. 7, n. 1, p. 1–14, 2017. Citado na página 11.
- DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, Elsevier, v. 31, n. 4-5, p. 198–211, 2007. Citado na página 15.
- DOUDKINE, A. et al. Nuclear texture measurements in image cytometry. *Pathologica*, v. 87, n. 3, p. 286, 1995. Citado na página 24.
- FARIA, P. R. de et al. Absence of galectin-3 does not affect the development of experimental tongue carcinomas in mice. *Experimental and molecular pathology*, Elsevier, v. 90, n. 2, p. 189–193, 2011. Citado na página 26.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <<http://www.deeplearningbook.org>>. Citado na página 19.

- GURUCHARAN, M. K. *Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network*. 2020. Disponível em: <<https://www.upgrad.com/blog/basic-cnn-architecture/>>. Acesso em: 25/5/2021. Citado 2 vezes nas páginas 4 e 22.
- HASSAN, M. ul. *VGG16 – Convolutional Network for Classification and Detection*. 2018. Disponível em: <<https://neurohive.io/en/popular-networks/vgg16/>>. Acesso em: 25/5/2021. Citado 2 vezes nas páginas 4 e 30.
- HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>. Citado 3 vezes nas páginas 4, 30 e 31.
- 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 2 vezes nas páginas 11 e 12.
- HUSSAIN, M.; BIRD, J. J.; FARIA, D. R. A study on cnn transfer learning for image classification. In: SPRINGER. *UK Workshop on computational Intelligence*. [S.l.], 2018. p. 191–202. Citado na página 19.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: PMLR. *International conference on machine learning*. [S.l.], 2015. p. 448–456. Citado na página 21.
- KRISHNAN, M. M. R. et al. Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis—an svm based approach. *Computers in biology and medicine*, Elsevier, v. 39, n. 12, p. 1096–1104, 2009. Citado na página 24.
- KRISHNAN, M. M. R. et al. Textural characterization of histopathological images for oral sub-mucous fibrosis detection. *Tissue and Cell*, Elsevier, v. 43, n. 5, p. 318–330, 2011. Citado na página 24.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, v. 25, p. 1097–1105, 2012. Citado 3 vezes nas páginas 11, 12 e 28.
- KUJAN, O. et al. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. *Oral oncology*, Elsevier, v. 42, n. 10, p. 987–993, 2006. Citado na página 10.
- 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 19.
- LECUN, Y. A. et al. Efficient backprop. In: *Neural networks: Tricks of the trade*. [S.l.]: Springer, 2012. p. 9–48. Citado na página 18.
- LEI, X.; PAN, H.; HUANG, X. A dilated cnn model for image classification. *IEEE Access*, IEEE, v. 7, p. 124087–124095, 2019. Citado na página 19.
- LU, S.; LU, Z.; ZHANG, Y.-D. Pathological brain detection based on alexnet and transfer learning. *Journal of computational science*, Elsevier, v. 30, p. 41–47, 2019. Citado 4 vezes nas páginas 4, 11, 23 e 31.

- MATHWORKS. *MQTT in MATLAB*. 2019. Disponível em: <<https://www.mathworks.com/matlabcentral/fileexchange/64303-mqtt-in-matlab>>. Acesso em: 26/5/2021. Citado na página 34.
- MOSQUITTO. *HomePage do Mosquitto*. 2021. Disponível em: <<https://mosquitto.org/>>. Acesso em: 26/5/2021. Citado na página 34.
- NODE-RED. *HomePage do NODE-RED*. 2021. Disponível em: <<https://telegram.org/>>. Acesso em: 26/5/2021. Citado na página 33.
- ROSEBROCK, A. *deep learning for computer vision with python*. [S.l.]: PYIMAGESEARCH, 2017. v. 1. Citado 7 vezes nas páginas 4, 16, 17, 18, 19, 20 e 21.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado na página 18.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Citado na página 18.
- SANT’ANA, J. M. D. A. et al. Activation of the wnt/beta-catenin signaling pathway during oral carcinogenesis process is not influenced by the absence of galectin-3 in mice. *Anticancer research*, International Institute of Anticancer Research, v. 31, n. 9, p. 2805–2811, 2011. Citado na página 26.
- SAVYAKHOSLA. *CNN | Introduction to Pooling Layer*. 2019. Disponível em: <<https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/>>. Acesso em: 19/5/2021. Citado 2 vezes nas páginas 4 e 21.
- SCULLY, C. Challenges in predicting which oral mucosal potentially malignant disease will progress to neoplasia. *Oral diseases*, Wiley Online Library, v. 20, n. 1, p. 1–5, 2014. Citado na página 10.
- SERTEL, O. et al. An image analysis approach for detecting malignant cells in digitized h&e-stained histology images of follicular lymphoma. in pattern recognition (icpr). *2010 20th International Conference on Pattern Recognition*, p. 273–276, 2010. Citado na página 11.
- SHANMUGAMANI, R. *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras*. [S.l.]: Packt Publishing Ltd, 2018. Citado 4 vezes nas páginas 4, 17, 20 e 22.
- SILVA, A. B. et al. Métodos computacionais para análise e classificação de displasias em imagens da cavidade bucal. Universidade Federal de Uberlândia, 2019. Citado 4 vezes nas páginas 4, 15, 26 e 27.
- SILVA, A. B. et al. Segmentation of oral epithelial dysplasias employing mask r-cnn and color normalization. In: IEEE. *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.], 2020. p. 2818–2824. Citado na página 24.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 29.

- SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. Citado 2 vezes nas páginas 11 e 12.
- SINGH, S. P. *Fully Connected Layer: The brute force layer of a Machine Learning model*. 2018. Disponível em: <<https://iq.opengenus.org/fully-connected-layer/>>. Acesso em: 25/5/2021. Citado na página 21.
- SMITH, J. et al. Biomarkers in dysplasia of the oral cavity: a systematic review. *Oral oncology*, Elsevier, v. 45, n. 8, p. 647–653, 2009. Citado 2 vezes nas páginas 10 e 14.
- TAN, C. et al. *A Survey on Deep Transfer Learning*. 2018. Citado na página 23.
- TANG, X.-H. et al. Oral cavity and esophageal carcinogenesis modeled in carcinogen-treated mice. *Clinical Cancer Research*, AACR, v. 10, n. 1, p. 301–313, 2004. Citado na página 26.
- TELEGRAM. *HomePage do Telegram*. 2021. Disponível em: <<https://nodered.org/>>. Acesso em: 26/5/2021. Citado na página 32.
- WARNAKULASURIYA, S. et al. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. *Journal of Oral Pathology & Medicine*, Wiley Online Library, v. 37, n. 3, p. 127–133, 2008. Citado 3 vezes nas páginas 6, 14 e 15.
- WRIGHT, D.; ADDIS, B. J.; LEONG, A. S. *Diagnostic lymph node pathology*. [S.l.]: CRC Press, 2011. Citado na página 10.
- YOSINSKI, J. et al. How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*, 2014. Citado 2 vezes nas páginas 11 e 23.