

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
DEPARTAMENTO DE COMPUTAÇÃO E MATEMÁTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Diagnóstico de câncer auxiliado pela
aplicação de redes convolucionais
profundas sobre imagens histopatológicas**

Guilherme Martiniano de Oliveira

MONOGRAFIA FINAL

PROJETO DE
GRADUAÇÃO

Supervisor: Prof. Dr. Joaquim Cezar Felipe

Ribeirão Preto - São Paulo
2023

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Dedico este trabalho a todos que ajudaram ao longo desta caminhada.

Quando a sabedoria e a emoção são direcionadas para um mesmo objetivo, não há o que possa evitar a sua realização, pois não existe nada mais forte no mundo.

— Venerável Mestre Hsing Yün

Resumo

Guilherme Martiniano de Oliveira. **Diagnóstico de câncer auxiliado pela aplicação de redes convolucionais profundas sobre imagens histopatológicas**. Monografia (Bacharelado). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Departamento de Computação e Matemática, Universidade de São Paulo, São Paulo, 2023.

Com a integração tecnológica sendo cada vez maior na medicina, surge uma demanda de sistemas que auxiliem profissionais da saúde a realizarem suas tarefas e tomarem decisões. Dessa forma, precisa-se garantir a precisão e acurácia de tais ferramentas por lidarem diretamente com vidas humanas. Redes neurais têm sido uma solução para ensinar sistemas a tomarem decisões da forma similar a que um ser humano tomaria. No campo da oncologia o uso de imagens é altamente difundido para que os profissionais diagnostiquem seus pacientes. Assim, o próximo passo é capacitar softwares a interpretar imagens e tomarem decisões referentes ao conteúdo das mesmas. Isso é possível por meio de redes neurais convolucionais que extraem informações de imagens, podendo classificá-las pelo contexto ali encontrado. Essa pesquisa busca avaliar três arquiteturas de redes neurais convolucionais: ResNet, DenseNet e EfficientNet. Utilizando aquela com melhor performance para um estudo de alterações estruturais da rede. Os dados da pesquisa são amostras celulares de lâminas de tecido da mucosa oral e da mama, os quais serão os alvos para aperfeiçoamento da arquitetura selecionada. Conforme estudado, a arquitetura da ResNet tem a melhor performance, velocidade e maleabilidade, e portanto, foi escolhida para as alterações estruturais. No final, foi obtido um modelo com acurácia acima de 96% para detecção de tumores em células de tecido de mucosa oral ou mama.

Palavras-chave: Rede Neural Convolucional. Imagem Histopatológicas. Diagnóstico Auxiliado por Computador.

Abstract

Guilherme Martiniano de Oliveira. **Diagnosis of Cancer Aided by the Application of Deep Convolutional Networks on Histopathological Images**. Bachelor's Thesis (Bachelor). Faculty of Philosophy, Sciences and Letters at Ribeirão Preto, Department of Computing and Mathematics, University of São Paulo, São Paulo, 2023.

With technological integration increasingly prevalent in medicine, there arises a demand for systems that assist healthcare professionals in performing their tasks and making decisions. Thus, it is essential to ensure the precision and accuracy of such tools as they directly deal with human lives. Neural networks have emerged as a solution to teach systems to make decisions in a manner similar to how a human would. In the field of oncology, the use of images is widely spread to enable professionals to diagnose their patients. Therefore, the next step is to empower software to interpret images and make decisions regarding their content. This is achievable through convolutional neural networks that extract information from images, enabling them to classify based on the context found therein. This research aims to assess three architectures of convolutional neural networks: ResNet, DenseNet, and EfficientNet. Choosing the one with the best performance for a study on structural alterations of the network. The research data consists of cellular samples from slides of oral mucosa and breast tissue, which will be the targets for refining the selected architecture. As studied, the ResNet architecture demonstrates the best performance, speed, and flexibility and, therefore, has been chosen for structural alterations. In the end, a model with an accuracy exceeding 96% for the detection of tumors in oral mucosa or breast tissue cells was achieved.

Keywords: Convolutional Neural Network. Histopathological Images. Computer-Aided Diagnosis.

Lista de Figuras

2.1	Gráfico de pesos	5
3.1	Amostras mucosa oral	8
3.2	Amostras mama	8
3.3	Filtro de uma camada de convolução	9
3.4	Bloco residual	10
3.5	Bloco denso	11
3.6	Escalonamento de arquitetura com EfficientNet	12
3.7	Comparação tonalidade	14
3.8	Comparação desfoque	14
3.9	Comparação transformação geométrica	15
3.10	Comparação adição de ruído	16

Lista de Tabelas

2.1	Pesos de ranqueamento dos artigos	5
3.1	Variantes da ResNet	10
3.2	Variantes da DenseNet	11
3.3	Arquitetura da EfficientNet-B0	12
3.4	Intervalos transformação de tonalidade	13
3.5	Intervalos transformação de desfoque	14
3.6	Intervalos transformação geométrica	15

4.1	Hiperparâmetros avaliação inicial	19
4.2	Hiperparâmetros <i>transfer learning</i>	20
4.3	Hiperparâmetros aprendizado do zero	20
5.1	Resultados treino arquiteturas	21
5.2	Resultados <i>transfer learning</i> aprimoramento	22
5.3	Resultados aprendizado do zero aprimoramento	22
5.4	Resultados aprendizado do zero mucosa oral	23

Sumário

1	Introdução	1
1.1	Contextualização	2
1.2	Objetivo	3
2	Trabalhos Correlatos	4
3	Materiais e Métodos	7
3.1	Base de dados	7
3.2	Arquiteturas	9
3.3	Transformações	12
3.4	Customizações da arquitetura	16
4	Metodologia Proposta	18
4.1	Avaliação inicial das arquiteturas	18
4.2	Comparação de diferentes abordagens	19
5	Resultados e discussão	21
5.1	Seleção da arquitetura	21
5.2	Aprimoramento da arquitetura	21
5.3	Demais experimentos	22
6	Conclusões	24
	Referências	25

Capítulo 1

Introdução

A incidência global de câncer tem testemunhado um aumento constante (FERLAY *et al.*, 2020), desafiando a comunidade médica a encontrar abordagens mais eficazes para o diagnóstico e tratamento dessa complexa variedade de doenças. A heterogeneidade intrínseca aos diferentes tipos de câncer, combinada com a rápida evolução das terapias oncológicas, demanda uma compreensão cada vez mais refinada para personalizar os cuidados ao paciente. Dessa maneira, soluções tecnológicas avançadas podem ser implantadas para apoiar os profissionais, e é nesse cenário que a integração de softwares de apoio à decisão desempenha um papel crucial, alguns casos de uso sendo: a análise rápida de grandes conjuntos de dados, a identificação de padrões relevantes e a oferta de recomendações personalizadas que podem melhorar significativamente os resultados clínicos.

À medida que avançamos na era da medicina personalizada, a interseção entre câncer e tecnologias de informação assume uma relevância cada vez maior (BRAGA *et al.*, 2019). Este projeto de graduação busca explorar modelos de inteligência artificial (*artificial intelligence* ou AI) como sendo um ponto de apoio à decisão no contexto oncológico (LYSAGHT *et al.*, 2019) avaliando como essas ferramentas conseguem classificar amostras de lâminas de biópsia. Ao examinar a interação entre a complexidade do câncer e a capacidade dos softwares de apoio à decisão em fornecer *insights* relevantes, esta pesquisa destaca as oportunidades e desafios associados a essa abordagem inovadora, visando aprimorar a prática clínica e, consequentemente, melhorar os resultados para os pacientes enfrentando essa doença devastadora.

Os modelos que foram avaliados são de arquiteturas de aprendizado profundo (*deep learning* ou DL), uma subárea da AI, que tem revolucionado diversos domínios, destacando-se de maneira particular na visão computacional (*computer vision* ou CV) (SCHNEIDER e VLACHOS, 2023). Esta representa uma área fascinante e altamente promissora, ao capacitar os sistemas a compreenderem e interpretarem imagens e vídeos de maneira semelhante à cognição humana. Exemplos de uso podem variar desde a detecção de objetos em tempo real até a segmentação precisa de imagens médicas, os avanços nesta interseção entre DL e CV têm redefinido os limites do que é possível em termos de automação visual.

Os dados são imagens celulares obtidas de biopsias de tecidos tumorais das regiões: mucosa oral e mama. As amostras são compostas de tecido saudável e tumoral corados

de hematoxilina e eosina (HE), nos quais os modelos foram treinados para classificação binária das amostras tumorais. A origem das etiquetas usadas para treino destas vem das seguintes fontes: as amostras de tecido mamário tem os rótulos obtidos pelo fornecedor do conjunto, já as da mucosa oral por patologistas colaboradores dessa pesquisa. Mais detalhes na seção 3.1.

1.1 Contextualização

Com a motivação de utilizar o conjunto de dados sem alterar estruturalmente o tipo deles, atualmente os classificadores mais adequados a lidar com imagens são as redes neurais convolucionais (*convolutional neural networks* ou CNNs). Além disso, são amplamente utilizadas para o diagnóstico de câncer (MRIDHA *et al.*, 2021) e tem se segmentado em demais campos como métodos de alta performance desde de seus primórdios com a LeNet (LECUN *et al.*, 1998). Estas emergem como uma subárea vital dentro do campo de CV, destacando-se por sua capacidade única de processar e extrair padrões complexos de dados visuais. Projetadas especificamente para lidar com informações de grade, as CNNs são estruturadas de maneira a simular o processo visual humano, permitindo uma análise hierárquica e contextual de imagens. Essas redes são capazes de identificar características específicas em diferentes níveis de abstração, possibilitando a classificação precisa em imagens médicas. Sua eficácia na aprendizagem de representações visuais torna as CNNs essenciais para avanços significativos na visão computacional, tornando possível a automação inteligente e a interpretação precisa de dados visuais em uma ampla gama de aplicações práticas. Neste trabalho foi comparado o desempenho de três arquiteturas de CNNs, para a classificação de células cancerígenas e foi proposta uma customização naquela que teve a melhor performance. Além disso, foi aplicada uma série de métodos de pré-processamento (conhecido como *data augmentation*) nas imagens usadas no treinamento das redes.

Como os dados vêm de fontes e tecidos distintos, a pesquisa buscou encontrar a arquitetura que melhor identifique tumores celulares, independente da origem celular. Além disso, por ser uma quantidade de dados pequena, quando comparada a treinos de grandes modelos de DL, as funções de *data augmentation* tornam possível o aumento artificial das amostras por meio de transformações geométricas, mudança de tonalidade e adição de ruído (SHORTEN e KHOSHGOFTAAR, 2019).

Atualmente, outras arquiteturas mais complexas de CNN se tornaram o estado da arte (*state-of-the-art* ou SOTA) em diversos conjuntos de teste (*benchmarks*) de CV. Entretanto, muitas seguem uma mesma base de rede, a ResNet (HE *et al.*, 2015), que foi revolucionária pela significativa diminuição do problema do sumiço do gradiente (*vanishing gradient problem*) que se torna mais relevante com o aumento da complexidade destas. Em seguida houve duas arquiteturas (ambas usando a anterior como base) que se segmentaram como SOTA por longos períodos de tempo foram: a DenseNet (HUANG *et al.*, 2017) e a EfficientNet (TAN e LE, 2019). Essas são as arquiteturas utilizadas nessa pesquisa. Os detalhes de cada uma aparece na seção 3.2.

1.2 Objetivo

O trabalho visou avaliar qual arquitetura de CNN pode ser de maior interesse para o desenvolvimento de um sistema de auxílio a tomada de decisão, especificamente para a análise histológica, que sirva como uma segunda opinião para patologistas terem maior segurança em seus diagnósticos. O segundo objetivo foi avaliar alterações estruturais e de hiperparâmetros da melhor rede para a análise específica de imagens de tecido de mucosa oral e de mama. Adicionalmente este trabalho visa fomentar a pesquisa feita para aplicações de saúde que utilizem de ferramentas baseadas em sistemas de aprendizado de máquina.

Capítulo 2

Trabalhos Correlatos

Ao explorar estudos previamente conduzidos e investigar os avanços e desafios enfrentados por outros pesquisadores, esta seção proporciona um contexto valioso para a presente pesquisa. Neste sentido, analisamos trabalhos que abordam temáticas semelhantes, avaliando suas metodologias, descobertas e lacunas identificadas. A análise destes trabalhos correlatos não só contribuirá para a construção de uma base sólida para o presente estudo, mas também possibilita a identificação de áreas específicas onde inovações e contribuições significativas podem ser efetuadas.

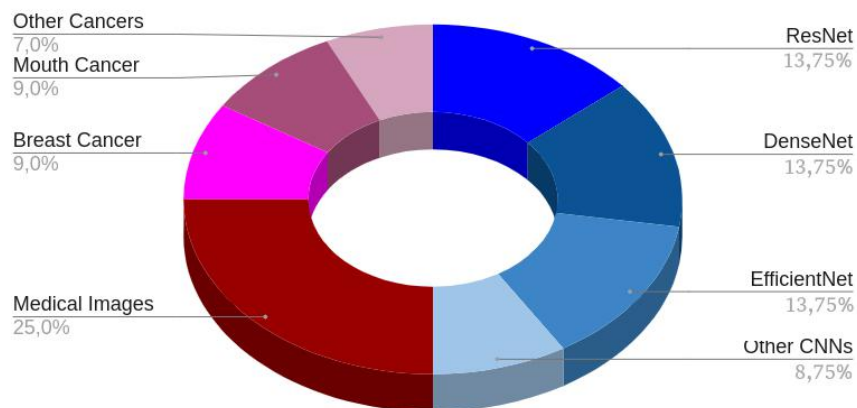
Os artigos aqui selecionados foram cuidadosamente escolhidos por meio de uma pesquisa precisa realizada nas seguintes bases acadêmicas: Periódico CAPES, Google Scholar, Web of Science, PubMed e SCOPUS. A consulta foi conduzida através de termos-chave contendo uma combinação das seguintes palavras (podendo estas terem sido utilizadas em inglês): inteligência artificial, aprendizado de máquina, aprendizado profundo, rede neural, rede neural convolucional, ResNet, DenseNet, EfficientNet, câncer, imagem, imagens médicas, imagens histológicas e imagens histopatológicas. Com essa metodologia foram encontrados mais de 600 artigos relevantes, destes houve uma filtragem a partir do ano de publicação para somente artigos de 2017 em diante e que tenham ao menos 5 dos termos-chave presentes no artigo (obrigatoriamente dois deles tinham de ser: câncer e rede neural convolucional). Assim, somente 17 artigos remanesceram e para melhor identificar aqueles que mais se relacionam com a pesquisa foi feita uma métrica que relaciona os termos-chaves de cada artigo com as características deste trabalho. Essa métrica, denominada, Relação, é composta pelos pesos apresentados na tabela 2.1, também representados graficamente na figura 2.1. Assim foram selecionados os cinco melhores artigos dada essa métrica de relação como artigos a serem analisados.

Classification of H&E images exploring ensemble learning with two-stage feature selection (TENGUAM *et al.*, 2022) (Relação=0,51). Os autores trazem nesse trabalho utilizar diversas arquiteturas de rede neural combinando-as em um classificador unido (*ensemble*). Este será formado pelas seguintes já consolidadas arquiteturas: ResNet (HE *et al.*, 2015), DenseNet (HUANG *et al.*, 2017), EfficientNet (TAN e LE, 2019), VGG (SIMONYAN e ZISSERMAN, 2015) e Inception (SZEGEDY, LIU *et al.*, 2014), além de conter extratores de atributos por eles próprios desenvolvidos. Os dados foram utilizados para treinamentos também incluem, imagem histopatológicas de câncer de mama e colorretal,

Tabela 2.1: Pesos aplicados para ranquear os artigos

Termo-Chave	Peso (%)
ResNet	13,75
DenseNet	13,75
EfficientNet	13,75
Outra CNN	8,75
Imagem Médica	25
Câncer de Mama	9
Câncer de Mucosa Oral	9
Outro tecido com Câncer	7
Penalização ^a	[0, 40]

^a Penalização de artigos que contenham mais de um modelo de CNN pois estes normalmente são ensembles ou artigos de *survey*, que fogem do escopo desta pesquisa. Valor da penalização é dez por cento multiplicado pela a soma de quantas CNNs estão presentes no artigo multiplicado, até no máximo 40%

**Figura 2.1:** Gráfico tipo torta, mostrando três grupos de pesos: redes convolucionais (50%), imagens médicas (25%) e tipo de câncer (25%)

além de amostras de fígado e displasia oral. Seu modelo de classificação tem uma performance que varia de 93.1 à 100 por cento de acurácia. Apesar desse alto desempenho não é evidenciado nos resultados por eles apresentados como poderia ser feita uma avaliação de quais dos extratores de atributos são os mais relevantes para podermos otimizar o *ensemble* e somente usar aqueles extratores mais relevantes. Concluindo, este trabalho representa como diferentes arquiteturas podem ser conectadas em um sistema homogêneo e ainda elevar a capacidade de previsão destes classificadores.

Classification of Hematoxylin and Eosin-Stained Breast Cancer Histology Microscopy Images Using Transfer Learning with EfficientNets (MUNIE, VIRIRI *et al.*, 2021) (Relação=0,415). Este estudo aborda os desafios no diagnóstico de câncer de mama. Ele constata as dificuldades e incertezas existentes no processo de diagnóstico convencional baseado em tecido de biópsia, que podem resultar em potenciais conflitos no diagnóstico final. Para superar esses problemas, sistemas de diagnóstico assistido por computador foram desenvolvidos. No entanto, há uma necessidade contínua de aprimoramento no desempenho para garantir a confiabilidade dos diagnósticos determinados.

A presente pesquisa foca na utilização da arquitetura EfficientNet (TAN e LE, 2019) para classificar imagens histológicas de câncer do conjunto de dados ICIAR2018. Sete variantes da rede foram ajustadas e avaliadas, demonstrando a capacidade do modelo de classificar imagens de mais de uma classe. O modelo EfficientNet-B2 alcançou notável precisão e sensibilidade (98,33% e 96,67%, respectivamente). Esses resultados destacam o potencial de transferência de características genéricas de imagens naturais para médicas por meio do ajuste fino de EfficientNets, proporcionando avanços promissores no diagnóstico confiável de câncer de mama.

Classification of breast cancer histopathological image with deep residual learning (Hu et al., 2021) (Relação=0,3775). Neste artigo, descreve-se uma metodologia para classificação de câncer de mama com base em imagens histopatológicas usando DL. Um modelo de rede neural convolucional baseado em aprendizado residual, chamado myResNet-34, foi projetado para classificação binária. Além disso, propõe-se um algoritmo para a geração automática de imagens-alvo para *stain normalization*, eliminando o viés causado pela seleção manual da imagem de referência. A distorção elástica foi introduzida e combinada com transformação afim para aumento artificial do conjunto de dados. Experimentos foram conduzidos no conjunto de dados BreakHis com o *framework* proposto, tendo resultados promissores, com uma precisão média de classificação em torno de 91% na classificação em nível de imagem. Os resultados indicaram que tanto a técnica de aumento de dados quanto a normalização de corantes contribuíram efetivamente para melhorar a precisão da classificação em 2-3%. Entretanto, não é evidenciado como foi feita a separação dos conjuntos de treinamento, podendo este estar enviesado com informações de mesmo paciente aparecendo em treino quanto em teste.

Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network (ALOM et al., 2019) (Relação=0,365). O trabalho combina três arquiteturas em um único modelo de classificação, elas sendo, InceptionV4 (SZEGEDY, IOFFE et al., 2016), ResNet (HE et al., 2015) e *Recurrent convolutional neural network* (RCNN). Além disso, seu estudo é focado em câncer de mama utilizando de dois conjunto de dados públicos, sendo eles o BreakHis e o *Breast Cancer (BC) classification challenge 2015*. Para aumentar ainda mais a quantidade de dados para treinamento, eles utilizam de ferramentas de *data augmentation* como: translações, rotações, escalonamento, recorte (*crop*), giro verticais e horizontais e etc. Um diferencial desta pesquisa é que o classificador desenvolvido pode ser utilizado para mais de uma classe, neste caso classificando até quatro tipos de tumores cancerígenos. Seus resultados ultrapassam 96% de acurácia sem *data augmentation* e 97% quando utilizado, porém os demais modelos que foram utilizados como comparação são modelos não considerados como SOTA atualmente. Exemplo disso é a seleção da arquitetura de CNN, AlexNet (KRIZHEVSKY et al., 2012), que foi inovadora quando lançada em 2012.

Capítulo 3

Materiais e Métodos

Esta seção elabora os materiais que foram usados na pesquisa, de modo a dar uma introdução à tecnologia e explicar como foi aplicada.

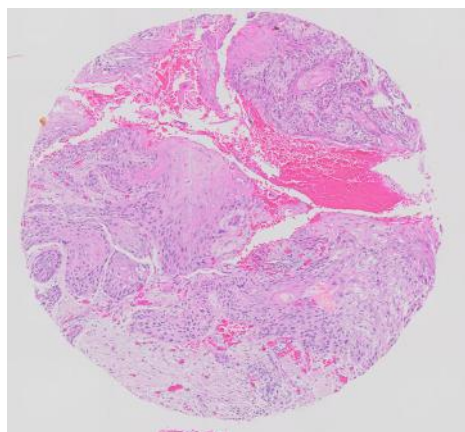
3.1 Base de dados

O conjunto de dados é composto por imagens médicas celulares de dois tipos de tecido: mucosa oral e mamário. Ambos utilizam coloração HE para os componentes celulares serem melhor evidenciados ao olho humano quando vistos por um microscópio digital.

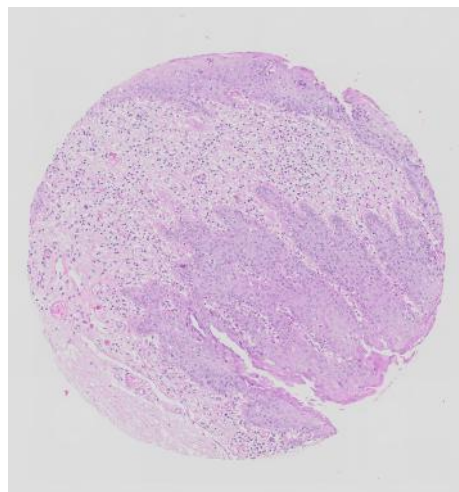
O primeiro tem sua origem atrelada ao estudo realizado em uma pesquisa conjunta entre faculdades brasileiras de medicina ([MATOS *et al.*, 2020](#)), onde, ao todo, foram coletadas amostras de 37 pacientes (26 com tumor superficial ou profundo e 11 saudáveis). Para uma análise mais precisa das células, patologistas foram consultados para evidenciar regiões onde eles conseguem diferenciar as células cancerígenas das saudáveis. Com isso, também podem ser identificadas regiões de interesse (*region of interest* ou ROIs) demarcadas para que os classificadores possam ser treinados em um estudo futuro. Exemplo de amostras desse tecido estão presentes na figura [3.1](#).

Já o segundo foi adquirido de uma instituição dos Estados Unidos, chamada *Tissue Array TA*, [2022](#), que fornece vários conjuntos de dados médicos já rotulados. Especificamente, este conjunto contém 80 cortes celulares (40 cancerígenos e 40 de tecido adjacente não cancerígeno) de 40 pacientes. Assim, é disponibilizada uma tabela com a descrição de cada amostra, onde é possível localizar a classificação celular, como maligna ou tecido adjacente (conjunto de controle), e outras informações, tais como grau do tumor e diagnóstico patológico. Exemplo de amostras desse tecido estão presentes na figura [3.2](#).

Juntando ambos conjuntos temos um total de 77 pacientes, somando 234 amostras (153 do primeiro conjunto e 80 do segundo). Este conjunto amostral é pouco para o treino de arquiteturas de DL do zero, portanto duas abordagens foram tomadas: a primeira seria avaliar como as redes se adequariam utilizando de pesos pré treinados com o treino focado

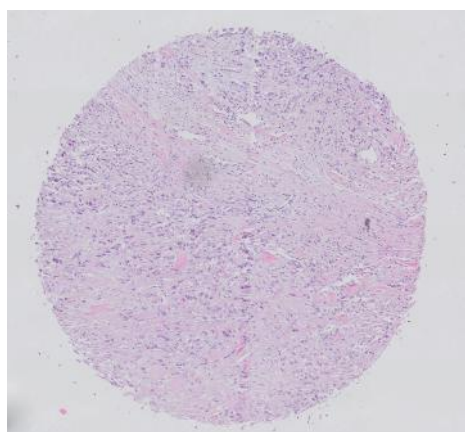


(a) Amostra com tumor

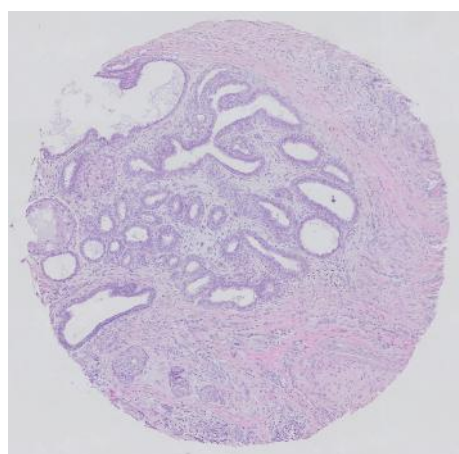


(b) Amostra sem tumor

Figura 3.1: Comparação das amostras do tecido mucosa oral



(a) Amostra com tumor



(b) Amostra sem tumor

Figura 3.2: Comparação das amostras do tecido mamário

na *head* de classificação e em seguida aumentar artificialmente a quantidade de dados para avaliarmos como o modelo se comportaria com um treino de pesos aleatorizados.

3.2 Arquiteturas

Com a ampla difusão e diferentes versões de cada rede utilizada, foi necessário utilizar um *framework* de aprendizado de máquina que tenha todas disponíveis em sua versão base, como divulgada por seus autores originais. Assim, o *framework* **Tensorflow** foi escolhido por ser uma ferramenta poderosa de manipulação de redes neurais com fácil desenvolvimento de pipelines de aprendizado.

As redes convolucionais são caracterizadas pelo alto uso de camadas de convolução, estas são filtros (*kernels*) que são passados sob o *input* da respectiva camada. O diferencial é que diferente de filtros tradicionais, como um filtro de bordas que tem seus valores fixos, em uma rede neural os valores destes são aprendidos conforme o modelo aprende sob seu conjunto de dados. Assim, é permitido que sejam criados filtros que melhor interpretem as informações de interesse para cada problema. A figura 3.3 mostra um *kernel* que deverá ter seus valores atualizados. Além disso, a estrutura da rede é comumente referida como a entrada e suas camadas próximas sendo a seção inferior (poucos mapas de atributos (também conhecidos como *feature maps*) e perto da *head* de classificação a seção superior tendo o comportamento inverso (muitos mapas de atributos).

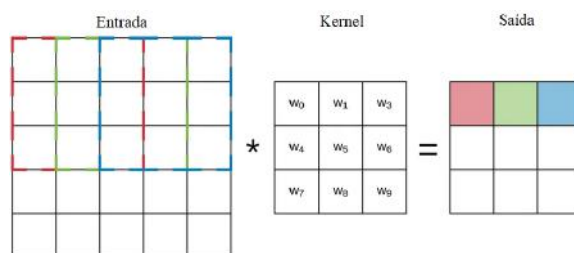


Figura 3.3: Filtro de uma camada de convolução com seus valores a serem definidos

3.2.1 ResNet

Selecionada pela sua alta relevância dentre as arquiteturas de CNNs, trazendo uma das maiores revoluções no campo. Originalmente de 2015 esta foi desenvolvida com objetivo de possibilitar treinos de redes mais profundas, dado que o impeditivo maior é o problema da dissipação do gradiente (*vanishing gradient problem*). Este ocorre naturalmente com o aumento de profundidade da rede caso esta não tenha algum mecanismo que facilite

Técnica conhecida como *Transfer Learning*

Técnica conhecida como *Data Augmentation*

<https://www.tensorflow.org/>

Versão 2.10.0

Resultado de uma camada convolucional que representa características específicas na imagem de entrada ou no mapa de atributos.

Profundidade de rede equivale a quantidade de camadas em série.

a transmissão de informação obtida a cada iteração de treino. A situação faz com que as camadas inferiores recebam informação inversamente proporcional a profundidade da rede, ou seja elas terão seu treino prejudicado, necessitando de um esforço maior ao se comparar com as camadas superiores. A solução proposta é utilizar um mecanismo que transmite a informação de entrada de um bloco de convolução a sua saída, fazendo que pelo menos parcialmente a informação original tenha mais relevância no resultado daquele bloco. Além disso, a perda ao ser propagada pela rede terá mais impacto nas camadas iniciais, aumentando a velocidade do aprendizado. Essa estrutura foi denominada de bloco residual, a figura 3.4 ilustra seu funcionamento.

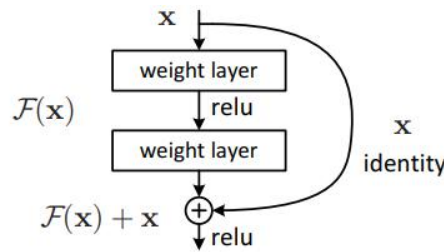


Figura 3.4: Bloco residual. Fonte: HE et al., 2015

Quando publicada a arquitetura da rede apresenta cinco variantes relacionadas ao seu tamanho: ResNet-18, ResNet-34, ResNet-50, ResNet-101 e ResNet152. A arquitetura de cada uma é descrita na tabela 3.1. A variante selecionada foi a ResNet-50 por ter um tamanho similar as demais arquiteturas utilizadas.

Tabela 3.1: Arquiteturas das variantes da ResNet. Fonte: HE et al., 2015

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

3.2.2 DenseNet

Escolhida pois sua arquitetura aplica o conceito do aprendizado residual (HE et al., 2015) em massa sob todas as camadas da rede, tendo como característica principal a alta conectividade (densidade). Originalmente de 2016, seus pesquisadores à desenvolveram para continuar minimizando o problema da dissipação do gradiente, aumentar a propagação dos atributos de cada camada e otimizar o treino. Assim foi feito o bloco denso, ilustrado pela

figura 3.5, estruturado para que seja transmitido para cada camada seguinte os atributos que foram obtidos como resultado das operações de convolução da camada atual.

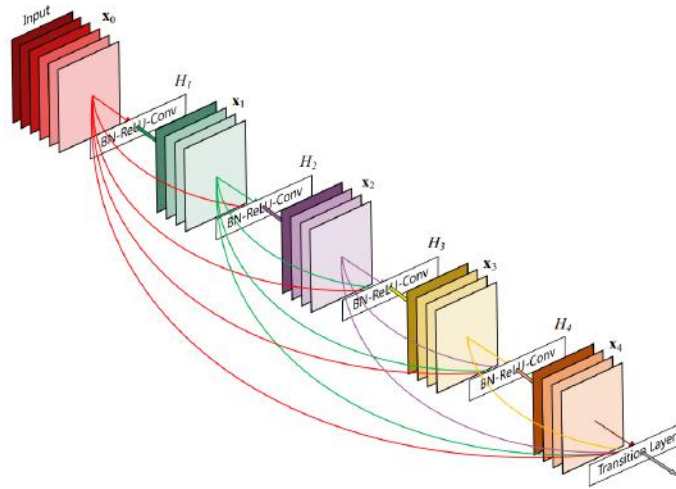


Figura 3.5: Bloco denso de 5 camadas. Fonte: HUANG et al., 2017

Assim como a anterior, quando publicada a rede tinha quatro variantes: DenseNet-121, DenseNet-169, DenseNet-201 e DenseNet-264. A arquitetura de cada uma é descrita na tabela 3.2. A variante selecionada foi a DenseNet-121 por ser a variante com tamanho mais próximo as demais arquiteturas utilizadas.

Tabela 3.2: Arquiteturas das variantes da DenseNet. Fonte: HUANG et al., 2017

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

3.2.3 EfficientNet

A terceira rede estudada foi selecionada por trazer uma nova abordagem na elaboração de CNNs, na qual deve ser feito uma arquitetura de base onde serão avaliadas diversas mudanças estruturais (profundidade, largura, resolução ou uma composição destas). Ilustrado

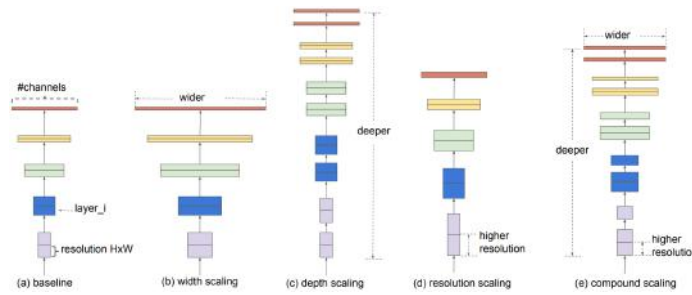


Figura 3.6: Escalonamento de arquitetura com EfficientNet, sob diferentes eixos de escalonamento.
 Fonte: TAN e LE, 2019

na figura 3.6 as diferentes abordagens para se alterar estruturalmente uma arquitetura de base.

Além disso, sob o conjunto de dados NCT-CRC-HE-100K (KATHER *et al.*, 2018) este é o modelo considerado como SOTA em acurácia com uma taxa de 95,59% de acerto. Dado que este é um conjunto de imagens celulares com câncer, sua alta performance indicou essa arquitetura como uma forte candidata para investigação. Adicionalmente, assim como os demais modelos aqui estudados, este em sua publicação original dispôs de 8 variantes (uma *baseline* e 7 customizações dessa). A variante selecionada foi a EfficientNet-B0 (*baseline*) por ser a versão SOTA no *benchmark* do conjunto de dados aqui citado, sua arquitetura é descrita na tabela 3.3.

Tabela 3.3: Arquitetura da EfficientNet-B0. Fonte: TAN e LE, 2019

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

3.3 Transformações

Devida a baixa quantidade de amostras para realizar o treinamento de uma rede profunda, foi utilizada a técnica de *data augmentation* e um gerador de dados dinâmico

Benchmarks de diversos modelos nesse conjunto de dados foram encontrados no site [https://paperswithcode.com/\(META_AI_RESEARCH,2018\)](https://paperswithcode.com/(META_AI_RESEARCH,2018)).

para artificialmente aumentar o conjunto de dados. Para isso foi utilizado o pacote, **Albumentations**, da linguagem de programação *Python*, que traz em sua disposição diversas transformações (desfoque, recorte, geométricas, tonalidade e outras). Além disso, ela permite um controle do desenvolvedor para fixar intervalos de limite de intensidade das transformações aplicadas e probabilidades para que essas ocorram, tornando-a uma ferramenta extremamente potente na formação de novos dados que evite réplicas. Nesta pesquisa as seguintes foram utilizadas:

3.3.1 ColorJitter

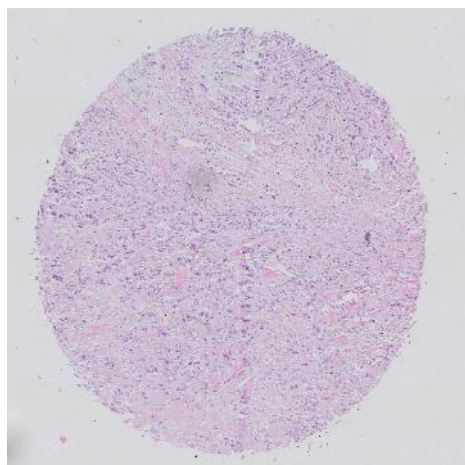
Essa transformação modifica as tonalidades da imagem, podendo alterar os seguintes atributos: brilho, contraste, saturação e *hue* (matiz). A transformação foi selecionada para reforçar o modelo a aprender características e padrões estruturais das imagens, tirando o viés das cores originais da imagem. Cada atributo possui um intervalo a ser definido pelo desenvolvedor de limite de alteração do mesmo, da qual a função aleatoriamente selecionara um valor desse intervalo. Para o brilho, contraste e saturação caso seja selecionado 1.0 aquele atributo não modificará a imagem, já para o atributo *hue* esse valor é 0.0. A figura 3.7 ilustra o resultado dessa transformação. Para essa pesquisa foram utilizados os intervalos mostrados na tabela 3.4.

Tabela 3.4: Intervalos para os atributos da transformação de tonalidades

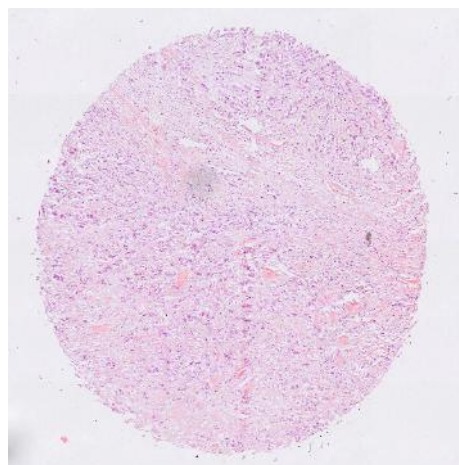
Atributo	Intervalo
Brilho	[0.8, 1.2]
Contraste	[0.7, 1.3]
Saturação	[0.7, 1.3]
<i>Hue</i>	[-0.1, 0.1]
Probabilidade da transformação	80%

3.3.2 GaussianBlur

Essa transformação modifica os valores de cada pixel da imagem, aproximando-os de seus vizinhos para a diminuição da nitidez ao utilizar a função probabilística Gaussiana. Dessa forma, é gerado o fenômeno de desfoque onde a intensidade de cada pixel é menos contrastada com os demais. A transformação foi selecionada para reforçar o modelo a aprender as características da imagem mesmo se esta tiver sofrido um desvio de nitidez, adicionando maior generalidade ao modelo para detectar as características de interesse. A figura 3.8 ilustra o resultado dessa transformação. O pacote permite a alteração do tamanho do filtro e o desvio padrão a ser utilizado, porém foram usados os valores padrões provenientes da função, encontrados na tabela 3.5.



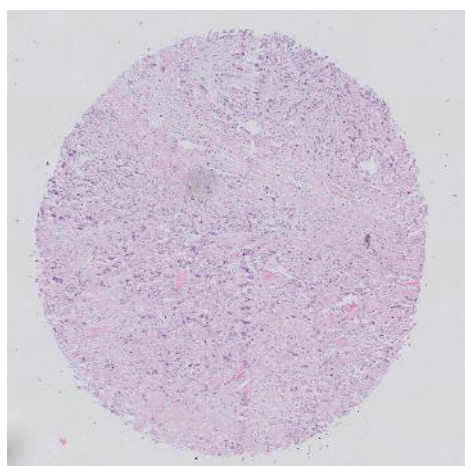
(a) Imagem original



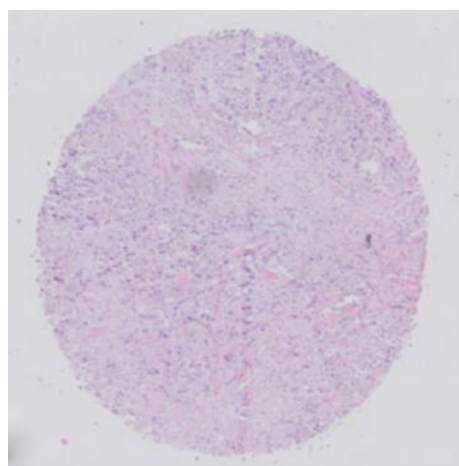
(b) Imagem transformada

Figura 3.7: Comparação da imagem original com a resultante da transformação de tonalidade**Tabela 3.5:** Intervalos para os atributos da transformação de desfoque

Atributo	Conjunto
Tamanho do <i>kernel</i>	{3x3, 5x5, 7x7}
Desvio padrão	{0.8, 1.1, 1.4}
Probabilidade da transformação	80%



(a) Imagem original



(b) Imagem transformada

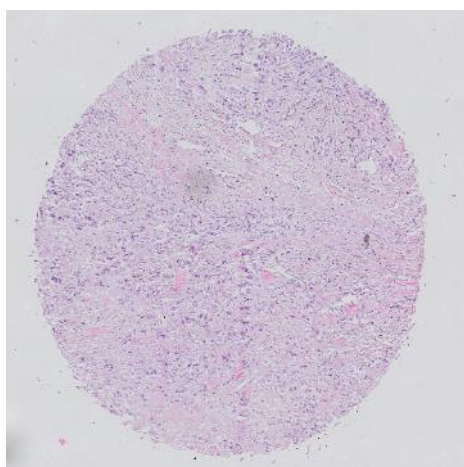
Figura 3.8: Comparação da imagem original com a resultante da transformação de desfoque

3.3.3 ShiftScaleRotate

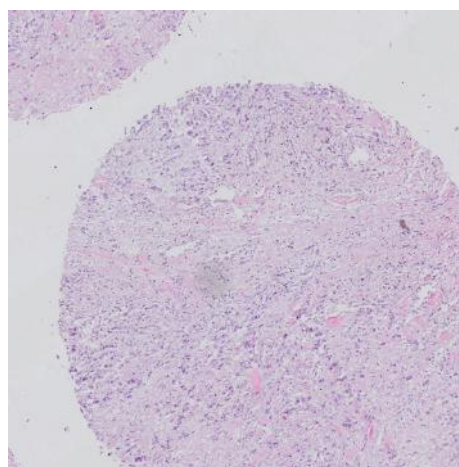
Essa transformação modifica a posição de cada pixel da imagem, utilizando de funções geométricas para transladar, escalonar e rotacionar os pixels. Dessa forma, é feita uma combinação (com uma ou mais) destas funções para gerar estruturas novas (não centradas, cortes, maior ou menor escala e etc) que o modelo não veria com os dados originais, já que estes vêm padronizados. A transformação foi selecionada para remover o viés do modelo para seguir os padrões estruturais do conjunto de dados original, tornando o modelo mais robusto para detectar as características de interesse. Para a escala caso seja selecionado 1.0 aquele atributo não modificará a imagem, para a rotação este deverá ser 0°, já para o atributo translação essa sempre causará algum tipo de deslocamento na imagem caso a transformação ocorra. A figura 3.9 ilustra o resultado dessa transformação. Para essa pesquisa foram utilizados os intervalos mostrados na tabela 3.6.

Tabela 3.6: Intervalos para os atributos da transformação geométrica

Atributo	Intervalo
Translação (vertical e horizontal)	[0.05, 0.2]
Escala	[0.9, 1.5]
Rotação	[-45°, 45°]
Probabilidade da transformação	90%



(a) Imagem original



(b) Imagem transformada

Figura 3.9: Comparação da imagem original com a resultante da transformação geométrica

3.3.4 RandomSnow

Essa transformação modifica o valor de alguns pixels da imagem, alterando seus valores para branco. Dessa forma, aleatoriamente ocorre a "remoção" de alguns valores da imagem original, dando um efeito de neve sob algumas partes da amostra. A transformação foi selecionada para adicionar ruído ao conjunto de dados, tornando o modelo mais resiliente

Isso foi imposto para forçar que pelo menos um atributo sempre seja alterado, impedindo que a imagem não se modifique mesmo quando ela passe pela função de transformação.

a pequenas alterações de valor na imagem. Essa transformação se usada em alta escala poderia comprometer o aprendizado do modelo, assim foi selecionado uma probabilidade de ocorrência desta de 30%, já os demais parâmetros da função foram mantidos seus valores padrões. A figura 3.10 ilustra o resultado dessa transformação.

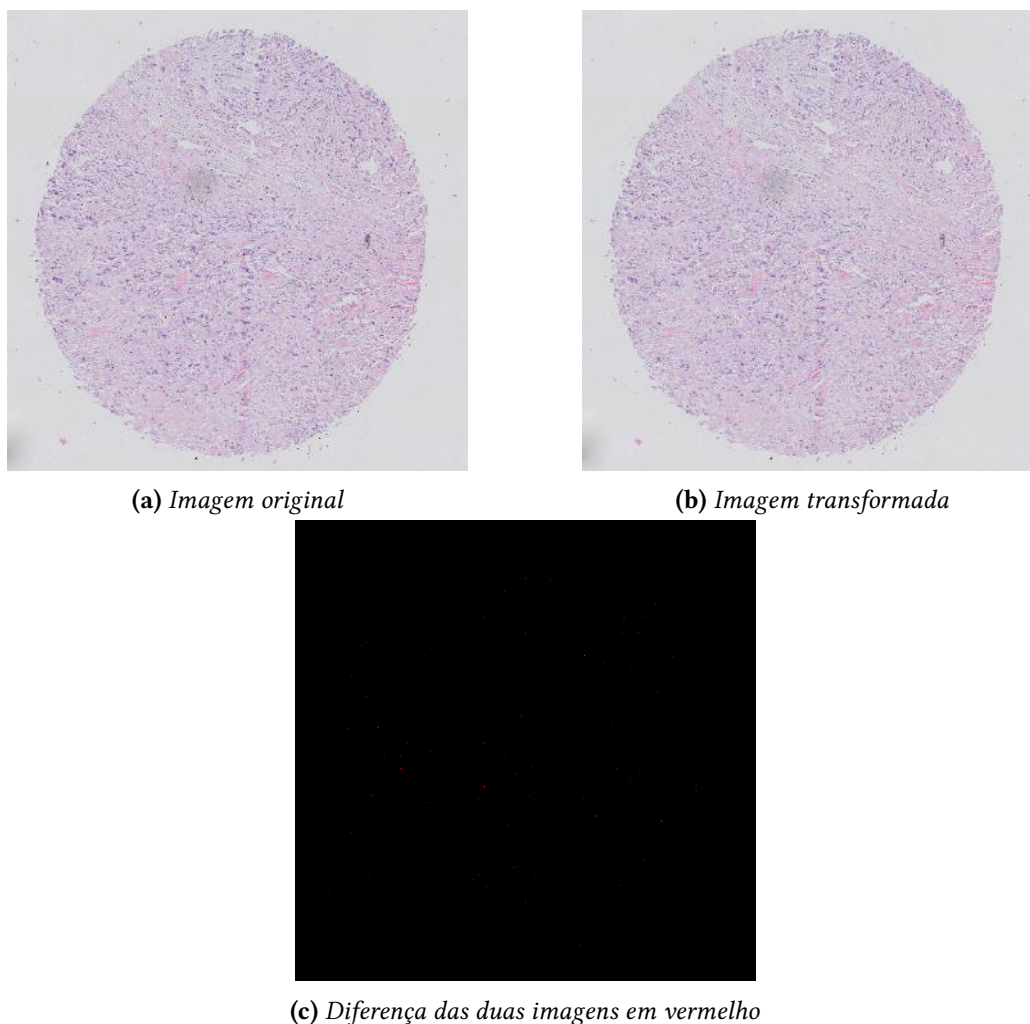


Figura 3.10: Comparação da imagem original com a resultante da transformação de adição de ruído

3.4 Customizações da arquitetura

A metodologia proposta para as customizações da arquitetura consiste em adicionar ou alterar camadas e/ou hiperparâmetros da rede. As customizações realizadas visaram não alterar ou deformar as capacidades de interpretação dos modelos treinados, somente customizando de modo que retenham as características principais da rede original. A adição de uma camada de *dropout* aqui sugerida não foi implementada em nenhuma das variantes em suas arquiteturas originais. Porém esta customização vêm tomando popularidade em demais áreas de DL ajudando a melhor selecionar neurônios relevantes na camada de classificação da rede. Já as outras duas, são práticas bem estabelecidas no desenvolvimento de CNNs, pois são passos naturais na exploração de balancear métricas

de avaliação (acurácia, precisão, F1-Score, entre outros...) com o desempenho (velocidade de treino, velocidade de inferência e etc...) do modelo. Além disso, para adicionar mais robustez ao treino, nessa etapa foram aplicadas as transformações citadas na seção 3.3 para aumentar o conjunto de dados artificialmente, no objetivo de treinar um modelo que seja menos propício a quaisquer vieses.

3.4.1 Dropout

A camada de *dropout* tem como efeito selecionar aleatoriamente, a cada iteração, alguns neurônios da camada anterior e zerar seus pesos (ignorando a saída daquele neurônio). Portanto, ao longo do treino uns neurônios serão ativados e desativados dando uma perspectiva de quais tiveram maior impacto no processo de treinamento. Esta camada foi aplicada na *head* de classificação, pois essa etapa concentra o maior número de neurônios que uma camada terá ao longo da rede. Isso facilitará o treino reduzindo a quantidade de neurônios que devem ser atualizados a cada época (HINTON *et al.*, 2012). O único parâmetro de configuração é a taxa de neurônios para serem desativados, no estudo foi usado o valor de 50%.

3.4.2 Aumento de resolução da imagem de entrada

O aumento da resolução da imagem de entrada tem sérios efeitos no custo computacional de treino da rede pois significa que cada matriz gerada se torna maior, necessitando de mais operações matemáticas. Como este trabalho envolve redes profundas, esses custos são maiores já que temos mais camadas de convoluções que uma rede convolucional não profunda. Mas esse aumento traz os seguintes benefícios: maior quantidade de informação para o modelo aprender, melhor representação de contextos e redução da perda de informação. A resolução de imagem escolhido foi de 640x640, valor bem acima ao utilizado no treino dos autores original das rede no conjunto de dados, *ImageNet*, de 224x224. Para testes em ambientes de uso em larga escala ou de muitas chamadas de inferência o tempo é um fator importante para avaliação do modelo, mas como esse foi uma pesquisa acadêmica de arquitetura de redes neurais tais considerações não foram feitas.

3.4.3 Adição de camadas de convolução

Adicionar mais camadas de convolução a uma CNN pode resultar em melhorias significativas no aprendizado e na representação de características em dados visuais. Cada camada de convolução permite que a rede aprenda representações hierárquicas mais abstratas à medida que avança na profundidade. As primeiras camadas são capazes de capturar características simples, como bordas e texturas, enquanto camadas mais profundas conseguem extrair características mais complexas e específicas da tarefa em questão. Isso aumenta a capacidade da CNN de discernir padrões sutis e nuances em dados visuais, tornando-a mais robusta e capaz de lidar com a complexidade inerente a muitas tarefas de visão computacional. No entanto, é importante equilibrar o aumento da profundidade com considerações de eficiência computacional e evitar *overfitting*, garantindo que o modelo generalize bem para dados não vistos.

É feito um rebalanceamento dos pesos remanescentes para compensar os neurônios faltantes

Capítulo 4

Metodologia Proposta

Após a avaliação das três arquiteturas estudadas foram feitas, sob a melhor, diversas alterações para analisarmos possíveis melhorias em busca de otimizar o treino do modelo no conjunto de dados desse estudo.

O estudo foi dividido em duas etapas de treino. A primeira consta da avaliação das três arquiteturas propostas comparando-as entre si para identificar a que melhor performar, com base na métrica de acurácia. Para agilizar o processo de comparação, nesta etapa não foram utilizadas as técnicas de *data augmentation* descritas na seção 3.3 e as arquiteturas utilizaram de pesos pre treinados do conjunto de dados, *ImageNet*. Ou seja, foram utilizadas as 234 amostras na íntegra, separando-as em treino (60%|140 amostras), validação (30%|71 amostras) e teste (10%|23 amostras) e reduzindo seus tamanhos para (244, 244).

Já a segunda etapa é composta dos seguintes passos: elaboração do conjunto de dados dinâmico; avaliação da arquitetura selecionada sob diferentes tipos de treino e análise dos resultados. Aqui a dinâmica de treino será mais alongado, já que buscamos encontrar minuciosamente avaliar a evolução de cada treino, constatar *insights* para os diferentes resultados obtidos e otimizar o treino do modelo no conjunto de dados desse estudo. Para isso, os dados originais não serão suficiente, então, as funções de transformação agora foram aplicadas para aumentarem em 10x a quantidade de amostras, trazendo heterogeneidade aos dados. Assim, ao todo ficaram 2340 amostras, separadas em treino (70%|1643 amostras), validação (20%|463 amostras) e teste (10%|234 amostras) e aumentadas seus tamanhos para (480, 480).

Para ambas etapas a biblioteca, **Comet**, auxiliou no controle da evolução do treino, também como na geração de gráficos para visualizar a performance dos modelos.

4.1 Avaliação inicial das arquiteturas

Para a avaliação inicial, as características do treino para todas as arquiteturas são mostrados na tabela 4.1.

Tabela 4.1: Hiperparâmetros da avaliação inicial das arquiteturas: ResNet, DenseNet e EfficientNet

Hiperparâmetros	Valores
Taxa de aprendizado (<i>learning rate</i>)	10^{-3}
Máximo de épocas	200
Otimizador (<i>optimizer</i>)	ADAM ^a
Peso pré treinado	ImageNet
Tipo de treino	Transfer learning ^b

^a Adaptive Moment Estimation

^b Descrição do conceito na seção 4.2

O peso utilizado no treino de *transfer learning*, é fornecido pelo *framework* sendo este treinado sob 1.000 classes do conjunto *ImageNet* e portanto, não deve ser usado no topo da rede, onde ocorre a classificação, pois este trabalho lida com classes binárias. Dessa maneira, a *head* de classificação foi padronizada para todas como sendo um neurônio que recebe um vetor de atributos após o último mapa de atributos passar pela camada de *GlobalAveragePooling*.

4.2 Comparação de diferentes abordagens

Nesta etapa, foram realizados 2 tipos de treino: *transfer learning* e aprendizado do zero. Seus hiperparâmetros são descritos nas tabelas 4.2 e 4.3, respectivamente, para cada experimento de uma alteração nova sob a arquitetura escolhida. Os tipos de treino podem ser descritos da seguinte forma:

Transfer learning. Essa estratégia de treino faz com que as camadas de convolução da rede não sejam treináveis, portanto o treino ocorre na *head* de classificação e as demais camadas são utilizadas como extratores de atributos.

Aprendizado do zero. Essa é a estratégia padrão para se treinar uma modelo de aprendizado de máquina, são gerados aleatoriamente pesos que preenchem os nós do modelo. Porém, para arquiteturas de DL como as trabalhadas nessa pesquisa, é necessária uma grande quantidade de dados para resultar em um treino efetivo. Com o conjunto de dados dinâmico podemos simular mais dados para tentarmos alcançar bons treinos do zero.

A arquitetura escolhida foi submetida aos experimentos descritos na seção 3.4 e um novo treino sem alterações arquiteturais para ser utilizado como *baseline*. Portanto, foram feitos 4 experimentos sob 2 diferentes tipos de treino, totalizando 8 modelos a serem analisados. Porém, no caso da adição de camadas de convoluções, o treino de *transfer learning* não é válido por seu peso pré treinado não ser compatível com arquiteturas distintas a que este foi treinado. Assim, obtivemos 7 resultados para análise.

Tabela 4.2: *Hiperparâmetros da segunda etapa, para o tipo de treino transfer learning*

<i>Transfer learning</i>	
Hiperparâmetros	Valores
Taxa de aprendizado (<i>learning rate</i>)	10^{-3}
Máximo de épocas	200
Otimizador (<i>optimizer</i>)	ADAM
Peso pré treinado	<i>ImageNet</i>

Tabela 4.3: *Hiperparâmetros da segunda etapa, para o tipo de treino aprendizado do zero*

Aprendizado do zero	
Hiperparâmetros	Valores
Taxa de aprendizado (<i>learning rate</i>)	10^{-2}
Máximo de épocas	300
Otimizador (<i>optimizer</i>)	ADAM
Peso pré treinado	Não se aplica, pesos inicializados aleatoriamente

Capítulo 5

Resultados e discussão

5.1 Seleção da arquitetura

A primeira etapa deste trabalho, a comparação das três arquiteturas, concluiu com a arquitetura ResNet (HE *et al.*, 2015) tendo a melhor performance, tomando-se como base a acurácia de teste, conforme evidenciado pela tabela 5.1. Sua incrível velocidade de treino e melhor acurácia sob o conjunto de teste, a fizeram se destacar. A arquitetura se provou altamente maleável para se adaptar com o conjunto de dados da pesquisa, mesmo nessa etapa utilizando somente 211 amostras para aprender as características de um tumor. Ainda assim, vale destacar que apesar da demora para estabilizar, o treino da EfficientNet-B0 (TAN e LE, 2019) obteve a maior acurácia no conjunto de treino, empatou na acurácia de validação e *loss* provando-se uma arquitetura eficiente, podendo até performar melhor que a selecionada dado os recursos computacionais e tempo de treino maiores.

Tabela 5.1: Resultados do treino das três arquiteturas

Arquitetura	Acurácia de teste (%)	Acurácia de treino (%)	Acurácia de validação (%)	Loss de treino final	Épocas ^a
ResNet-50	75	82,64	68,75	0,38	23
DenseNet-121	68,75	74,22	76,56	0,49	96
EfficientNet-B0	62,5	88,2	76,56	0,38	200

^a Quantidade de épocas necessárias para o modelo estabilizar

5.2 Aprimoramento da arquitetura

Conforme descrito na seção 3.4, foram feitas alterações na arquitetura da ResNet-50 (HE *et al.*, 2015) com intuito de melhorar ainda mais sua acurácia sob o conjunto de dados da pesquisa. Para esta segunda etapa foi feito um aumento da resolução das imagens em todas os experimentos realizados, de (244, 244) para (480, 480), com objetivo de passar o máximo de informação possível ao modelo. Além disso, passa a ser utilizado as funções de transformações citadas na seção 3.3 para aumentar o conjunto de dados. As customizações de *dropout* e aumento da resolução da imagem de entrada foram aplicadas seguindo suas descrições citadas nas seções 3.4.1 e 3.4.2, já a customização de arquitetura foi feita seguindo

as características principais da ResNet (He *et al.*, 2015), utilizando blocos residuais. Foram adicionados dois blocos residuais a arquitetura original na camada superior da rede, com 1024 filtros cada um. Essa alteração foi feita para serem obtidos mapas de atributos mais complexos que possam precisamente diferenciar a região de um tumor de uma região saudável, com a esperança de conseguir diferenciar melhor pequenos detalhes de cada amostra.

Tabela 5.2: Resultados do treino tipo *transfer learning* para aprimoramento da arquitetura da ResNet-50

Transfer learning			
Modelo	Acurácia de teste (%)	Acurácia de treino (%)	Acurácia de validação (%)
ResNet-50 <i>baseline</i>	93,75	96,73	83,72
ResNet-50 + Dropout	81,25	93,61	83,55
ResNet-50 + Aumento Resolução	96,87	97,23	85,20

Analisando a tabela 5.2 referente ao treino de *transfer learning*, pode-se concluir que como esperado ao utilizar uma resolução maior para imagem de entrada, o modelo terá maior facilidade em interpretar as características que diferem as células saudáveis daquelas com tumor. Mas, o modelo de base teve uma performance semelhante em relação ao com aumento de resolução. Isso é evidenciado pela diferença de somente 3,12% de acurácia sob o conjunto de teste, mostrando novamente como a arquitetura é potente e maleável.

Tabela 5.3: Resultados do treino tipo *aprendizado do zero* para aprimoramento da arquitetura da ResNet-50

Aprendizado do zero			
Modelo	Acurácia de teste (%)	Acurácia de treino (%)	Acurácia de validação (%)
ResNet-50 <i>baseline</i>	62,5	78,34	73,68
ResNet-50 + Dropout	46,87	30,33	42,76
ResNet-50 + Aumento Resolução	37,5	21,59	26,32
ResNet Customizada ^a	62,5	82,95	73,68

^a Nesse treino o otimizador ADAM, não estava conseguindo calcular uma loss para o treino. Por isso, esse foi substituído pelo otimizador SGD, com o parâmetro de *clipvalue* sendo igual a 1.

Com os treinos de pesos aleatorizados (resultados do treino na tabela 5.3) o cenário já muda, tanto a arquitetura de base quanto a arquitetura customizada performaram melhor. Indo mais além, na acurácia de treino a arquitetura maior se destacou por conseguir aprimorar as características do conjunto de treino melhor em seu classificador, mostrando como que as camadas superiores podem auxiliar na interpretação de formas complexas.

5.3 Demais experimentos

Houve outros experimentos que foram realizados durante a pesquisa, mas que infelizmente não tiveram resultado condizente ou inferior aos aqui apresentados. Nesta seção serão apresentados alguns destes, que podem levar a investigações futuras.

5.3.1 Fine tuning

Ao incorporar um conjunto de pesos já pré treinados a um modelo, existe a possibilidade de tentar aperfeiçoá-los a um conjunto de dados de interesse. Isso ocorre da seguinte forma:

após importados os pesos são liberados para poderem ser modificados novamente por uma rotina de treino, porém esse treino deve ser feito com cautela para preservar as características intrínsecas dos pesos. Ou seja, o treino deve ser realizado com uma taxa de aprendizado reduzido e sob poucas épocas. Essa estratégia de treino faz com que as camadas de convoluções da rede se tornem treináveis, podendo então fazer pequenas alterações no extratores de atributos para melhor se adequarem ao conjunto de dados desejado. Além das demais, esta também foi utilizada para treino na segunda etapa da pesquisa. Porém, em todos as variantes do modelo da ResNet-50 (He *et al.*, 2015), treinadas dessa maneira, resultaram com métricas piores do que as mesmas vinda do treino de *transfer learning* por uma margem de 25% em média.

5.3.2 Separação dos tecidos

Foi testado separar o conjunto de dados em seus respectivos tecidos. Porém com o tecido mamário as métricas apontavam para uma indicação de *overfitting* muito forte, tendo modelos chegando em 100% de acurácia no conjunto de teste. Embora, no conjunto de tecido de mucosa oral, as métricas tiveram comportamento normal. A tabela 5.4 mostra os resultados da segunda etapa da pesquisa somente no conjunto de mucosa.

Tabela 5.4: Resultados do treino tipo aprendizado do zero para aprimoramento da arquitetura da ResNet-50 no conjunto de mucosa oral

Aprendizado do zero			
Modelo	Acurácia de teste (%)	Acurácia de treino (%)	Acurácia de validação (%)
ResNet-50 <i>baseline</i>	75	73,69	93,05
ResNet-50 + Dropout	75	67,63	59,03
ResNet-50 + Aumento Resolução	25	38,06	40,97
ResNet Customizada	87,5	70,05	89,93

5.3.3 Validação Cruzada

Foram feitos testes para avaliar os experimentos gerados sob o mesmo conjunto de dados, porém de maneira a serem treinadas com validação cruzada para eliminar o possível viés da ordem dos conjuntos de dados. Porém, os resultados obtidos indicam a necessidade de aprimoramento da implementação. Assim, esta abordagem será apresentada em estudos futuros.

Capítulo 6

Conclusões

O estudo trouxe diversas perspectivas sobre as arquiteturas aqui trabalhadas, também no uso das transformações para geração de dados artificiais e sobre os métodos de treinamento de modelo de DL. Primeiramente, foi evidenciado que a arquitetura da ResNet (HE *et al.*, 2015) continua relevante tanto utilizando seu peso pré treinado que provê um ótimo extrator de atributos, quanto ao utilizar dos blocos residuais para complementar a arquitetura e torná-la mais potente para a extração de informações mais complexas. Além desta, a arquitetura EfficientNet (TAN e LE, 2019) trouxe surpresas na primeira etapa da pesquisa, já que performou bem apesar de ter sido mais lenta que a outra. Mas traz uma perspectiva de escalabilidade que pode ser aplicada a outras arquiteturas.

As transformações utilizadas produziram um conjunto de dados mais representativos e variados, que apresentou um ganho de robustez para os modelos treinados. A conclusão disso, é que essas foram essenciais para que as arquiteturas pudessem aprender e ao mesmo tempo não tendo problemas de *overfitting*. Além de ter evidenciado, que elas acrescentam heterogeneidade nos conjuntos de dados e não alterem a estrutura das imagens onde poderiam perder as características de interesse. Concluindo, também foram evidenciados os diversos comportamentos e requerimentos que cada modo de treino de rede neural fornece. Podem ser evidenciadas características do método de *transfer learning* sendo, alta adaptabilidade e capacidade de convergir o treino rapidamente, porém há a restrição de manter o modelo com a mesma arquitetura em que os pesos foram treinados. Já treinos de arquiteturas do zero, são custosos em relação a quantidade de dados necessários para o modelo poder generalizar o conteúdo de interesse, mas quando disponível tendem a melhor performar que modelos pré treinados com pesos de diferentes contextos.

Foram resultados deste trabalho: um modelo de rede neural convolucional treinado para classificação de imagens celulares de tecidos de mucosa oral e mamário, um conjunto de transformações que generalizem a informação de tumores celular e uma avaliação das alterações realizadas na arquitetura selecionada. Por fim, podemos entender este projeto como um auxílio no desenvolvimento de ferramentas de apoio ao diagnóstico médico, já que qualquer modelo com mais de 90% de acurácia poderá servir como uma interpretação que auxilie o médico no diagnóstico de tumores.

Referências

- [ALOM *et al.* 2019] Md Zahangir ALOM, Chris YAKOPCIC, Mst Shamima NASRIN, Tarek M TAHA e Vijayan K ASARI. “Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network”. Em: *Journal of digital imaging* 32 (2019), pgs. 605–617 (citado na pg. 6).
- [BRAGA *et al.* 2019] Ana Vitória BRAGA *et al.* “Machine learning: o uso da inteligência artificial na medicina/ machine learning: the use of artificial intelligence in medicine”. Em: *Brazilian Journal of Development* 5.9 (set. de 2019), pgs. 16407–16413. DOI: [10.34117/bjdv5n9-190](https://doi.org/10.34117/bjdv5n9-190). URL: <https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/3437> (citado na pg. 1).
- [FERLAY *et al.* 2020] J FERLAY *et al.* “Global cancer observatory: cancer today”. Em: (2020). [Accessed 14-11-2023] (citado na pg. 1).
- [HE *et al.* 2015] Kaiming HE, Xiangyu ZHANG, Shaoqing REN e Jian SUN. *Deep Residual Learning for Image Recognition*. 2015. arXiv: [1512.03385](https://arxiv.org/abs/1512.03385) [[cs.CV](https://arxiv.org/abs/1512.03385)] (citado nas pgs. 2, 4, 6, 10, 21–24).
- [HINTON *et al.* 2012] Geoffrey E. HINTON, Nitish SRIVASTAVA, Alex KRIZHEVSKY, Ilya SUTSKEVER e Ruslan SALAKHUTDINOV. “Improving neural networks by preventing co-adaptation of feature detectors”. Em: *CoRR* abs/1207.0580 (2012). arXiv: [1207.0580](https://arxiv.org/abs/1207.0580). URL: <http://arxiv.org/abs/1207.0580> (citado na pg. 17).
- [HU *et al.* 2021] Chuhan HU, Xiaoyan SUN, Zhenming YUAN e Yingfei WU. “Classification of breast cancer histopathological image with deep residual learning”. Em: *International Journal of Imaging Systems and Technology* 31.3 (2021), pgs. 1583–1594 (citado na pg. 6).
- [HUANG *et al.* 2017] Gao HUANG, Zhuang LIU, Laurens VAN DER MAATEN e Kilian Q WEINBERGER. “Densely connected convolutional networks”. Em: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pgs. 4700–4708 (citado nas pgs. 2, 4, 11).
- [KATHER *et al.* 2018] Jakob Nikolas KATHER, Niels HALAMA e Alexander MARX. *100,000 histological images of human colorectal cancer and healthy tissue*. Versão v0.1. Zenodo, mai. de 2018. DOI: [10.5281/zenodo.1214456](https://doi.org/10.5281/zenodo.1214456). URL: <https://doi.org/10.5281/zenodo.1214456> (citado na pg. 12).

- [KRIZHEVSKY *et al.* 2012] Alex KRIZHEVSKY, Ilya SUTSKEVER e Geoffrey E HINTON. “Imagenet classification with deep convolutional neural networks”. Em: *Advances in Neural Information Processing Systems*. Ed. por F. PEREIRA, C.J. BURGESS, L. BOTTOU e K.Q. WEINBERGER. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf (citado na pg. 6).
- [LECUN *et al.* 1998] Yann LECUN, Léon BOTTOU, Yoshua BENGIO e Patrick HAFNER. “Gradient-based learning applied to document recognition”. Em: *Proceedings of the IEEE* 86.11 (1998), pgs. 2278–2324 (citado na pg. 2).
- [LYSAGHT *et al.* 2019] Tamra LYSAGHT, Hannah LIM, Vicki XAFIS e Kee NGIAM. “Ai-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research”. Em: *Asian Bioethics Review* 11 (set. de 2019). DOI: [10.1007/s41649-019-00096-0](https://doi.org/10.1007/s41649-019-00096-0) (citado na pg. 1).
- [MATOS *et al.* 2020] Leandro Luongo MATOS *et al.* “Cancer-associated fibroblast regulation by micrnas promotes invasion of oral squamous cell carcinoma”. Em: *Oral Oncology* 110 (2020), pg. 104909. ISSN: 1368-8375. DOI: <https://doi.org/10.1016/j.oraloncology.2020.104909>. URL: <https://www.sciencedirect.com/science/article/pii/S1368837520303456> (citado na pg. 7).
- [META AI RESEARCH 2018] META AI RESEARCH. Papers with Code. <https://paperswithcode.com/> [Accessed: Nov 20, 2023]. 2018 (citado na pg. 12).
- [MRIDHA *et al.* 2021] Muhammad Firoz MRIDHA *et al.* “A comprehensive survey on deep-learning-based breast cancer diagnosis”. Em: *Cancers* 13.23 (2021), pg. 6116 (citado na pg. 2).
- [MUNIEN, VIRIRI *et al.* 2021] Chanaleä MUNIEN, Serestina VIRIRI *et al.* “Classification of hematoxylin and eosin-stained breast cancer histology microscopy images using transfer learning with efficientnets”. Em: *Computational Intelligence and Neuroscience* 2021 (2021) (citado na pg. 5).
- [SCHNEIDER e VLACHOS 2023] Johannes SCHNEIDER e Michalis VLACHOS. *A Survey of Deep Learning: From Activations to Transformers*. 2023. arXiv: [2302.00722](https://arxiv.org/abs/2302.00722) [cs.LG] (citado na pg. 1).
- [SHORTEN e KHOSHGOFTAAR 2019] Connor SHORTEN e Taghi M KHOSHGOFTAAR. “A survey on image data augmentation for deep learning”. Em: *Journal of big data* 6.1 (2019), pgs. 1–48 (citado na pg. 2).
- [SIMONYAN e ZISSERMAN 2015] Karen SIMONYAN e Andrew ZISSERMAN. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: [1409.1556](https://arxiv.org/abs/1409.1556) [cs.CV] (citado na pg. 4).

REFERÊNCIAS

- [SZEGEDY, IOFFE *et al.* 2016] Christian SZEGEDY, Sergey IOFFE, Vincent VANHOUCKE e Alex ALEMI. *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*. 2016. arXiv: 1602.07261 [cs.CV] (citado na pg. 6).
- [SZEGEDY, LIU *et al.* 2014] Christian SZEGEDY, Wei LIU *et al.* “Going deeper with convolutions”. Em: *CoRR* abs/1409.4842 (2014). arXiv: 1409.4842. URL: <http://arxiv.org/abs/1409.4842> (citado na pg. 4).
- [TA 2022] TA. *TissueArray.Com* — [tissuearray.com](https://www.tissuearray.com/). <https://www.tissuearray.com/>. [Accessed 20-08-2023]. 2022 (citado na pg. 7).
- [TAN e LE 2019] Mingxing TAN e Quoc LE. “Efficientnet: rethinking model scaling for convolutional neural networks”. Em: *International conference on machine learning*. PMLR. 2019, pgs. 6105–6114 (citado nas pgs. 2, 4, 6, 12, 21, 24).
- [TENGUAM *et al.* 2022] Jaqueline Junko TENGUAM *et al.* “Classification of h&e images exploring ensemble learning with two-stage feature selection”. Em: *2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE. 2022, pgs. 1–4 (citado na pg. 4).