



UNIVERSIDADE FEDERAL DO MARANHÃO

Curso de Ciência da Computação

José Ribamar Durand Rodrigues Junior

Estudo de Deep Fine Tuning Aplicado à Classificação de Imagens Histológicas

São Luís

2023

José Ribamar Durand Rodrigues Junior

Estudo de Deep Fine Tuning Aplicado à Classificação de Imagens Histológicas

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Geraldo Braz Junior

São Luís

2023

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Rodrigues Junior, José Ribamar Durand.

Estudo de deep fine tuning aplicado à classificação de
imagens histológicas / José Ribamar Durand Rodrigues
Junior. - 2023.

46 p.

Orientador(a): Geraldo Braz Junior.

Curso de Ciência da Computação, Universidade Federal do
Maranhão, São Luís, 2023.

1. Classificação de imagens histopatológicas. 2. Deep
fine tuning. 3. Deep learning. 4. Transfer learning. I.
Braz Junior, Geraldo. II. Título.

José Ribamar Durand Rodrigues Junior

Estudo de Deep Fine Tuning Aplicado à Classificação de Imagens Histológicas

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho aprovado em São Luís, 04 de julho de 2023:

Prof. Dr. Geraldo Braz Junior
Orientador

Profa. Dra. Simara Vieira da Rocha
Examinadora

Prof. Dr. Italo Franyles Santos da Silva
Examinador

São Luís
2023

Agradecimentos

A minha mãe Maria de Nazaré, gostaria de agradecer pelo constante incentivo aos meus estudos, por valorizar minha educação desde o início, e por ser meu maior exemplo de coragem, força, perseverança e caráter. Você me ensinou o valor do esforço, da honestidade, da determinação e todos os outros valores que carrego e que levarei para o resto da minha vida.

Meu profundo agradecimento ao meu estimado orientador Geraldo Braz Junior pelo seu apoio e orientação ao longo deste trabalho. Sou grato por ter a oportunidade de aprender com você e de contar com seu apoio ao longo deste trabalho. Muito obrigado por sua paciência e por ser um orientador excepcional.

A minha irmã Anna Beatriz, por todo apoio durante todos esses anos de graduação, foi fundamental para que conseguir alcançar meus objetivos na graduação, sou muito grato a você.

A minha namorada Gilvana Siqueira, por todo carinho e companheirismo nestes 5 anos de relacionamento, obrigado pelo apoio querida.

A todos os amigos que fiz no caminho, não foram muitos, mas foram os melhores possíveis, obrigado a Antonio, Giovanna, Hannah, Mayrla, Alana, Beker, Emily, Fernando, Rebecca, Igor e todos os outros que participaram dessa jornada, vocês foram muito importantes, sempre vou lembrar dos grupos de estudo após o almoço no PETComp, das revisões rápidas nos últimos 10 minutos antes das provas, das madrugadas escrevendo artigos pro PET, do grupo gigantesco em direção ao RU.

Gostaria de agradecer ao PETComp, ao FNDE, e aos meus tutores durante meu tempo como petiano, Prof. Luis Rivero e Prof. Geraldo, meu orientador. Apesar da bolsa não ser grande coisa, ganhei muita experiência, e muitos amigos nesse período, acredito que sem o PETComp teria desistido da graduação, muito obrigado.

Sou grato a UNASUS/DTED, trabalhei por lá por mais de 2 anos, e foi fundamental para financiar meus custos durante minha graduação, gostaria de agradecer especialmente a prof. Mario Meireles, Thacyla Lima, Jone Correia, Vinicius Augusto, Vinicius Costa, Luigi Dovera, Rayanne Silveira e todos os outros colegas a qual tenho muita estima.

Aos meus professores que foram fundamentais na minha formação, em especial a Simara, Mario, Anselmo, Riveiro, Salles, Francisco e novamente ao meu orientador prof. Geraldo. Sou grato por ter a oportunidade de aprender com vocês.

Por último, mas não menos importante, gostaria de agradecer a minha cachorrinha Cacaú que infelizmente faleceu de câncer este ano. Você estava ao meu lado desde os meus

7 anos, sempre ao meu lado nos momentos mais difíceis, sempre me acompanhando deitada ao pé da minha mesinha de estudos, dormindo ao meu lado enquanto fazia os trabalhos do ensino médio, enquanto estudava para o vestibular, enquanto revisava os conteúdos das disciplinas da graduação, enquanto programava até as 2 horas da madrugada para cumprir os prazos do trabalho, enquanto escrevia parte desta pesquisa, infelizmente você não está aqui para comemorar comigo a conclusão deste trabalho, mas você sempre esteve ao meu lado, nunca vou esquecer de você. Muito obrigado Cacau.

"Uma nova, vasta e poderosa linguagem é desenvolvida para o futuro das análises, na qual é possível utilizá-la para aplicações práticas para os objetivos da humanidade mais rápidas e precisas do que os meios até então em nossa posse tornaram possível."

Ada Lovelace - A primeira programadora

Resumo

São esperados 704 mil novos casos de câncer no Brasil entre 2023 e 2025, sendo a segunda maior causa de morte precoce no país e encontra-se entre as três maiores em mais de 127 países. A expectativa é que esses números aumentem com o envelhecimento populacional e o declínio de outras causas de morte prematuras que anteriormente lideravam as estatísticas como acidente vascular cerebral e doença cardíaca coronária. Diante da relevância desta doença para saúde pública global, métodos automatizados de diagnóstico e estadiamento se tornam cada vez mais necessários. Dentre os métodos considerados padrão ouro, está a análise histopatológica, permitindo determinar o tipo de câncer e o estágio de progressão da doença por meio de alterações morfológicas nas amostras de tecido. Portanto, o objetivo deste trabalho consiste em avaliar arquiteturas convolucionais e a utilização de *deep fine tuning* no contexto de análise de imagens histopatológicas. Serão analisadas arquiteturas CNN recentes com e sem *deep fine tuning*, visando analisar os impactos desta técnica neste contexto. A aplicação da técnica permitiu resultados mais robustos, estáveis, e com maior velocidade de convergência, e dentre as arquiteturas analisadas a HR-Net mostrou-se a mais eficaz.

Palavras-chave: Classificação de imagens histopatológicas, Deep learning, Deep fine tuning, Transfer learning.

Abstract

704,000 new cases of cancer are expected in Brazil between 2023 and 2025, and it is the second leading cause of early death in the country and among the top three in more than 127 countries. The expectation is that these numbers will increase with the aging population and the decline of other causes of premature death that previously led the statistics as stroke and coronary heart disease. Given the relevance of this disease for global public health, automated methods of diagnosis and staging become increasingly necessary, among the methods that are considered the gold standard is histopathological analysis, allowing to determination of the type of cancer and the stage of disease progression through morphological changes in tissue samples. Therefore, the goal of this work is to evaluate convolutional architectures and the use of deep fine-tuning in the context of histopathological image analysis. Recent CNN architectures with and without deep fine tuning will be analyzed, aiming to analyze the impacts of this technique in this context. The application of the technique allowed more robust, stable, and faster convergence results, and among the analyzed architectures the HR-Net proved to be the most effective.

Keywords: Histopathological image classification, Deep learning, Deep fine tuning, Transfer learning.

Lista de ilustrações

Figura 1 – Imagem de Slide Inteiro	15
Figura 2 – Stain Normalization Macenko	16
Figura 3 – Arquitetura Alex-Net	18
Figura 4 – Bloco ResNext e ResNet	19
Figura 5 – Diagrama do bloco SE	20
Figura 6 – Bloco SE (Squeeze-and-Excitation)	21
Figura 7 – Bloco ConvNeXt	22
Figura 8 – Arquitetura HRNet	23
Figura 9 – Head HRNet	23
Figura 10 – Transfer Learning	24
Figura 11 – Metodologia.	26
Figura 12 – subtipos histológicos.	27
Figura 13 – Stain-normalization.	28
Figura 14 – Validação Cruzada	31
Figura 15 – Treino ResNeXt-101 Split 5	34
Figura 16 – Treino seresnet-101 Split 5	35
Figura 17 – Treino ConvNeXt Split 5	36
Figura 18 – Treino ConvNeXt Splits com treinamento do zero	37
Figura 19 – Treino ConvNeXt Splits com DFT	38
Figura 20 – Treino HRNet-W30 Split 5	39

Lista de tabelas

Tabela 1 – Detalhes das arquiteturas	29
Tabela 2 – Parâmetros de treinamento	30
Tabela 3 – Resultados obtidos nos experimentos considerando as variações de arquiteturas e forma de treinamento, onde DFT = Deep Fine Tunning, DP = Desvio Padrão.	33

Sumário

1	INTRODUÇÃO	12
1.1	Trabalhos Relacionados	13
1.2	Objetivos	14
1.2.1	Objetivos Específicos	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Imagens Histopatológicas	15
2.2	Stain normalization	16
2.2.1	Macenko	17
2.3	Redes Neurais Convolucionais	18
2.3.1	ResNext	19
2.3.2	SE-Resnet	20
2.3.3	ConvNeXt	21
2.3.4	HRNet	22
2.4	Transfer Learning	24
2.4.1	<i>Fine Tuning</i>	25
3	METODOLOGIA	26
3.1	Aquisição das Imagens	26
3.2	Pré-processamento	27
3.3	Modelos Avaliados	28
3.4	Transfer Learning	29
3.4.1	Hiperparâmetros de treinamento	29
3.5	Avaliação	30
4	RESULTADOS	33
4.1	Análise do modelo ResNeXt 101	34
4.2	Análise do modelo SEResNet 101	35
4.3	Análise do modelo ConvNeXt Base	35
4.4	Análise do modelo HRNet	37
4.5	Análise dos experimentos com DFT	38
5	CONCLUSÃO	40
	REFERÊNCIAS	42

1 Introdução

Apenas no Brasil, no triênio 2023-2025 são esperados 704 mil novos casos de câncer, considerando apenas as 21 variedades de câncer com maior incidência no país. Apenas de câncer de pele não melanoma, ocorrerão 483 mil casos novos, de câncer de mama feminino são esperados 73 mil novos casos, seguindo de próstata com 71 mil, cólon e reto (45 mil), pulmão (32 mil), estômago (21 mil) e o câncer do colo do útero (17 mil), e outros menos comuns (SANTOS et al., 2023).

O câncer é um problema de saúde mundial, representando uma das principais causas de morte prematura, estando entre as três primeiras causas em mais de 127 países, constituindo assim uma barreira para o aumento da expectativa de vida (BRAY et al., 2021). O aumento da mortalidade por câncer se deve pelo envelhecimento populacional, e tem ganhado destaque pelo declínio acentuado nas taxas de mortalidade por acidente vascular cerebral e doença cardíaca coronária, outras duas doenças que lideram as estatísticas de morte prematura (SUNG et al., 2021).

A descoberta e diagnóstico de câncer pode ser feita por diversos exames de imagens como raios-x, tomografia computadorizada (TC), ressonância magnética (RM), ultrassonografia (LI et al., 2021a). Biomarcadores também podem ser utilizados para este fim, analisando a presença de moléculas produzidas pelas células tumorais ou por outras células do corpo em resposta ao tumor, com assinatura molecular exclusiva e características identificáveis, como níveis ou atividades de uma infinidade de genes, proteínas ou outras características moleculares facilitando a definição molecular do câncer (KAMEL; AL-AMODI, 2016).

Mesmo com diversas opções de exames, a análise histopatológica é considerada o padrão ouro para estadiamento da maioria dos tipos de câncer (RUBIN et al., 2008). Ela envolve a identificação e a análise de alterações morfológicas e características em amostras de tecido, ajudando a determinar com precisão o grau, o estágio e o tipo de câncer. Para a detecção de nódulos pulmonares e para o diagnóstico de fibrose pulmonar idiopática a análise histopatológica é considerada padrão ouro (POLETTI et al., 2013), (SHAFFIE, 2021). Também é considerado padrão ouro para estadiamento de cânceres do cólon proximal ao reto (MALMSTRØM et al., 2016).

Dentre os diversos tipos de câncer que a análise histopatológica é considerada o padrão ouro em estadiamento e/ou diagnóstico destaca-se: fibrose pulmonar idiopática (POLETTI et al., 2013), nódulos pulmonares (SHAFFIE, 2021), câncer do cólon (MALMSTRØM et al., 2016), câncer de próstata (KWAK et al., 2011), câncer de mama

(SWATHI; KRISHNA; RAMESH, 2019), carcinoma testicular (FAZAL et al., 2022), câncer cerebral (KANTHIKAR; NIKUMBH; DRAVID, 2017).

Embora haja uma escassez crescente de patologistas, suas cargas de trabalho estão aumentando com o aumento do número de casos, e a exigência de diagnósticos mais abrangentes para identificar o tratamento mais adequado para os pacientes (LAAK; LITJENS; CIOMPI, 2021). A análise histopatológica é subjetiva e limitada pela variação entre diferentes patologistas visto que podem ocorrer discordância entre observadores, e falhas nos diagnósticos, além de ser uma atividade demorada (GHANATE et al., 2011).

A análise histopatológica é especialmente atraente para o uso de IA (Inteligência Artificial), que atualmente exige que patologistas avaliem cuidadosamente imagens de alta resolução em busca de diagnosticar doenças com base nas características visuais de tecidos, portanto a utilização de ferramentas de IA permitiria auxiliar os processos de diagnóstico, tornando-os mais eficientes e reproduzíveis, e aumentando a exatidão e a precisão (LAAK; LITJENS; CIOMPI, 2021).

1.1 Trabalhos Relacionados

Dentro do campo de análise de imagens histopatológicas, as metodologias de extração de características por meio de *deep learning*, está cada vez mais comum, com resultados consistentes, e como menor esforço em relação a engenharia de características.

Bardou, Zhang e Ahmad (2018) apresenta um estudo comparativo de features oriundas de SIFT e SURF com uma rede CNN. A rede CNN proposta mostrou-se eficiente para extração de características uteis para a tarefa de classificação, superando os demais métodos.

Bayramoglu, Kannala e Heikkilä (2016) propõem uma rede convolucional para classificar imagens histopatológicas em malignas e benignas e o grau de ampliação das imagens. O objetivo de adicionar o grau de ampliação nos rótulos das imagens é forçar a CNN a aprender características de magnificação, tencionando contribuir para a tarefa de classificação principal. A metodologia se mostrou eficiente em apenas um grau de ampliação, com os demais apresentando resultados inferiores em comparação com a CNN treinada apenas com os rótulos de doença.

Gour, Jain e Kumar (2020) desenvolvem uma arquitetura residual baseada nas ResNet-50, com modificações nas últimas camadas convolucionais. As primeiras 50 camadas convolucionais são idênticas ao modelo original, sendo inicializadas com os pesos pré-treinados. As características extraídas são utilizadas em outros classificadores, com a SVM apresentando os melhores resultados. Yan et al. (2020) elaboraram uma metodologia para classificar imagens grandes do dataset Aresta et al. (2019) mediante uma abordagem híbrida

de CNN e LSTM. A abordagem apresentou resultados melhores que o voto majoritário da classificação dos patches e da classificação por SVM a partir de características extraídos dos patches.

Zerouaoui e Idri (2022) implementa uma arquitetura para classificação de imagens histopatológicas utilizando redes neurais convolucionais pré-treinadas para extração de características utilizadas em alguns classificadores avaliados. As arquiteturas de CNN analisadas foram VGG16, VGG19, Inception V3, DenseNet 201, ResNet 50, Inception ResNet V2 e MobileNet V2. Os classificadores investigados foram MLP, SVM, DT e KNN.

Com o desenvolvimento de novas arquiteturas é fundamental analisar seu desempenho em contextos específicos, uma vez que cada problema tem suas características particulares, e cada arquitetura pode apresentar resultados diferentes de acordo com essas particularidades.

1.2 Objetivos

O objetivo geral desse trabalho é avaliar o desempenho de arquiteturas convolucionais utilizando fine tuning para a classificação de imagens histopatológicas de câncer de mama.

1.2.1 Objetivos Específicos

Destacam-se como objetivos específicos deste trabalho:

- Avaliar o desempenho das arquiteturas delimitadas para a tarefa de classificação de imagens histopatológicas;
- Avaliar os impactos da utilização de fine tuning na tarefa de classificação de imagens histopatológicas nas arquiteturas delimitadas.

2 Fundamentação Teórica

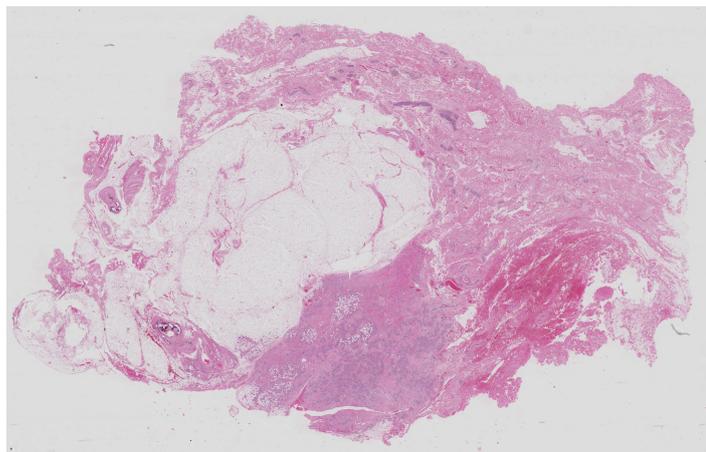
Neste capítulo estão expostos os principais fundamentos teóricos utilizados na metodologia proposta, desde o método de pré-processamento utilizado, até as arquiteturas e métodos de treinamento utilizados no estudo.

2.1 Imagens Histopatológicas

As imagens Histopatológicas são imagens de amostras de tecido que foram coradas e examinadas em um microscópio para identificar e diagnosticar doenças como o câncer. Essas imagens são obtidas por um processo que envolve a coleta de uma amostra de tecido, que deve ser cortada em seções finas, e coradas com corantes para destacar características do tecido, feito isso a imagem pode ser analisada em um microscópio, ou um scanner de lâminas.

As Lâminas fornecem imagens com uma visão mais abrangente da doença e seu efeito nos tecidos, pois o processo de preparação preserva a arquitetura do tecido subjacente. A estrutura adicional nessas imagens, ao mesmo tempo, em que fornece uma riqueza de informações, também apresenta um novo conjunto de desafios de uma perspectiva de análise automatizada de imagens. Espera-se que o aproveitamento adequado dessa informação espacial permita caracterizações mais específicas das imagens de uma perspectiva diagnóstica (GURCAN et al., 2009).

Figura 1 – Imagem de Slide Inteiro



Fonte: ICIAR 2018 BACH Challenger

As Lâminas histopatológicas podem conter diferentes artefatos que dificultam a análise qualitativa dessas imagens. Esses artefatos podem ser adicionados durante o processo de aquisição dessas lâminas, podendo conter dobras no tecido ou poeira nas

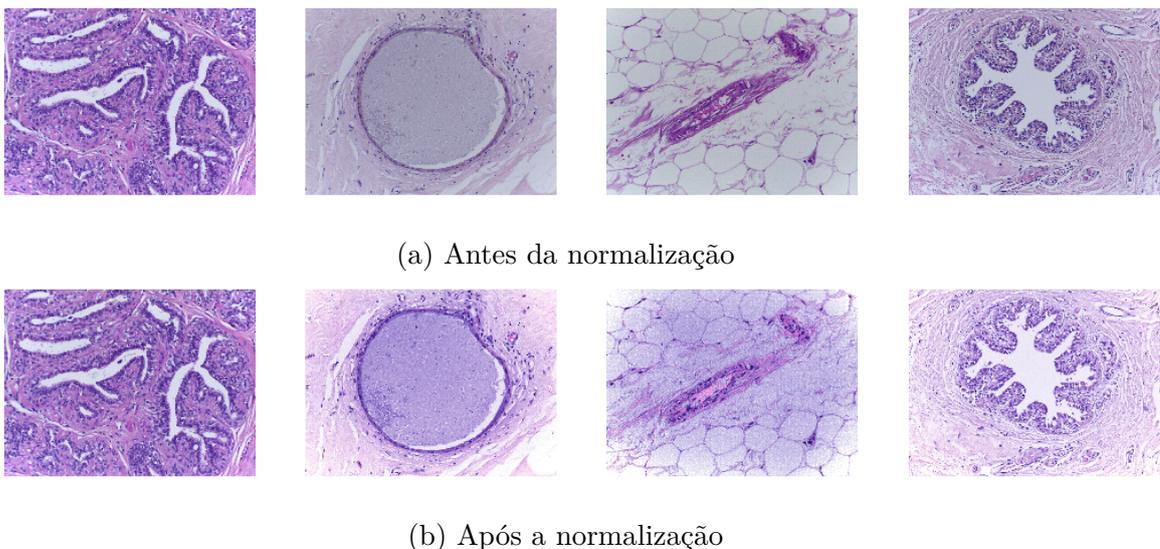
lâminas, além de variação de cores que podem ocorrer por diferentes lotes e fabricantes de corantes, diferentes espessuras nos cortes de tecido e até diferentes modelos de scanner (KOMURA; ISHIKAWA, 2018).

2.2 Stain normalization

A variação de cor no tecido histopatológico se deve a um conjunto fatores, como uso de diferentes scanners, durante a preparação de lâminas, diferentes equipamentos usados, diferentes colorações de corantes, atividade de diferentes fabricantes e lotes de corantes. Essa variação de cores nas imagens histopatológicas causa problemas.

Na preparação da imagem histopatológica, a imagem é corada por um corante ou tinta que geralmente aparece em uma cor diferente. Se processarmos as imagens sem pré-processamento, o resultado obtido pode estar sendo submetido a um diagnóstico incorreto (ROY et al., 2018).

Figura 2 – Stain Normalization Macenko



Fonte: (WANGHAO, 2020)

Uma das maneiras de reduzir a variação de cores na imagem histopatológica é converter a imagem RGB em escala de cinza, mas, na escala de cinza, algumas das informações são perdidas. Portanto, temos que usar técnicas de normalização de cores. A normalização de cores é o processo em que fazemos a transformação da cor média de uma imagem para outra (ROY et al., 2018).

As técnicas de normalização de cor para imagens histopatológicas são chamadas de normalização de mancha, ou stain normalization. Diferentes métodos de stain normalization foram desenvolvidos, desde métodos inicialmente criados para transferência de cor e de estilo como Reinhard (REINHARD et al., 2001), e técnicas desenvolvidas especialmente

para normalização de manchas histológicas como Macenko (MACENKO et al., 2009), Khan (KHAN et al., 2014) e Vahadane (VAHADANE et al., 2016), além de abordagem mais robusta como a de Salehi e Chalechale (2020) utilizando redes neurais geradoras adversárias (GANs).

2.2.1 Macenko

Um dos métodos propostos para superar a grande variância na coloração de imagens histológicas é o método de Macenko, criado para reduzir essas inconsistências e permitir uma análise quantitativa aprimorada. Primeiramente o conjunto de píxeis da imagem são convertidos do espaço RGB para densidade óptica (OD), através da Equação 2.1.

$$OD = -\log_{10}(I) \quad (2.1)$$

A transformação em densidade óptica fornece um espaço onde uma combinação linear de manchas resultará em uma combinação linear de valores de OD. Uma vez definido os vetores corretos o esquema de deconvolução de cores, exibido na Equação 2.2 é usado para transformar os valores de cores em valores quantitativos de interesse. Na equação OD é a densidade óptica observada, V e S são as matrizes dos vetores de mancha e as saturações de cada uma das manchas, respectivamente.

$$OD = VS \implies S = V^{-1}OD \quad (2.2)$$

Os pixels com baixa densidade óptica são eliminados, por motivos de estabilidade, para isso foi definido um limiar de corte. Um limiar de corte alto pode remover dados importante, por outro lado, um limiar muito baixo, pode comprometer a estabilidade do método. Após análise empírica o limiar sugerido foi de $\beta = 0,15$.

Para encontrar os pontos finais do vetor de manchas o método precisa aproximar os vetores, para isso é necessário calcular o plano que os vetores formam. Isso é feito formando um plano a partir dos dois vetores correspondentes aos dois maiores valores singulares da decomposição em valores singulares (SVD) dos píxeis OD transformados.

Todos esses píxeis transformados em OD são então projetados neste plano e subsequentemente normalizados para o comprimento da unidade. O ângulo em relação à primeira direção SVD é calculado para cada ponto, mapeando assim as direções no plano para um escalar.

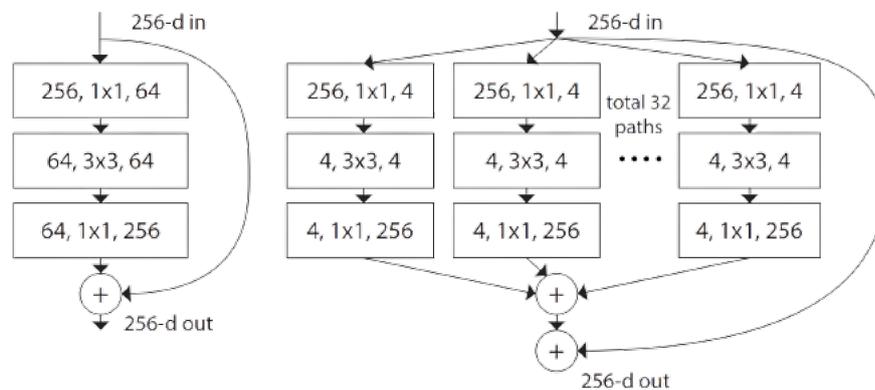
Para cada mancha em questão, são calculados os histogramas de intensidade para todos os píxeis que possuem a maioria dessa mancha. Em seguida, encontramos o percentil 99 desses valores de intensidade e usamos isso como uma aproximação do máximo. Esses valores são convertidos novamente para densidade óptica (OD).

A Alex-Net (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) é considerada a primeira rede neural convolucional profunda de sucesso, conseguiu superar significativamente os modelos anteriores de classificação de imagens ao utilizar uma rede neural profunda com várias camadas convolucionais e um grande número de parâmetros. Ele foi treinado no conjunto de dados ImageNet, e obteve uma taxa de erro top-5 de 15,3%, o que representou uma melhoria significativa em relação aos modelos anteriores. A Alex-Net abriu caminho para o desenvolvimento de redes neurais convolucionais cada vez mais profundas e complexas, que desde então se tornaram o padrão diversas tarefas de visão computacional.

2.3.1 ResNext

A ResNext (XIE et al., 2017) é uma rede baseada na ResNet (HE et al., 2016), com mudanças nos blocos residuais para adicionar uma nova dimensão (Figura 4). A ResNext adiciona o conceito de cardinalidade a rede, ao invés de usar uma camada convolucional 1x1 no início do bloco residual com uma entrada de 256 e 64 canais, a ResNext usa blocos em paralelo (caminhos), com uma camada convolucional com 256 de entrada e uma quantidade de canais que varia conforme a cardinalidade desta rede.

Figura 4 – Bloco ResNext e ResNet



Fonte: (XIE et al., 2017)

A cardinalidade é inspirada na estratégia *split-transform-merge*, presente nos módulos Inception, mas ao invés de combinar filtros convolucionais de diferentes núcleos mesclados por concatenação, a ResNeXt usa filtros convolucionais de mesmo núcleo. O objetivo é que esse bloco se aproxime do poder representacional de camadas grandes e densas, mas com uma complexidade computacional consideravelmente menor.

Com a adição de uma nova dimensão ao bloco, é proposto um novo método de agregação da camada residual para considerar a cardinalidade da rede, substituindo a transformação elementar presente na Equação 2.3 por uma função mais genérica que por si

só também pode ser uma rede, expandindo ao longo de uma nova dimensão representada pela equação 2.4.

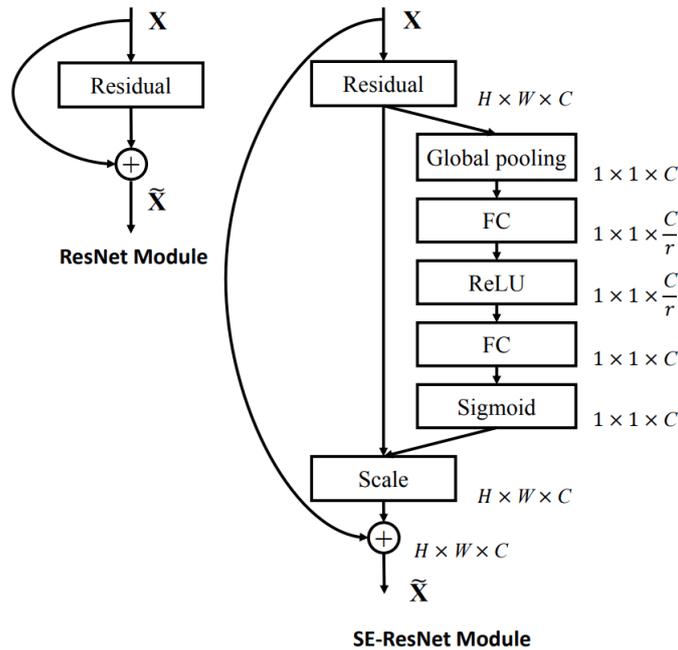
$$\sum_{i=1}^D w_i \cdot x_i \tag{2.3}$$

$$F(x) = \sum_{i=1}^C T_i(x) \tag{2.4}$$

2.3.2 SE-Resnet

A arquitetura SE-Resnet (HU; SHEN; SUN, 2018), assim como a anterior, é uma rede baseada na ResNet, que adiciona mecanismos para melhorar o desempenho geral da rede, neste caso o objetivo foi modelar as interdependências entre os canais convolucionais, para melhorar a qualidade das representações produzidas pela rede, com esse fim foi criado um novo bloco, *Squeeze-and-Excitation* (SE).

Figura 5 – Diagrama do bloco SE

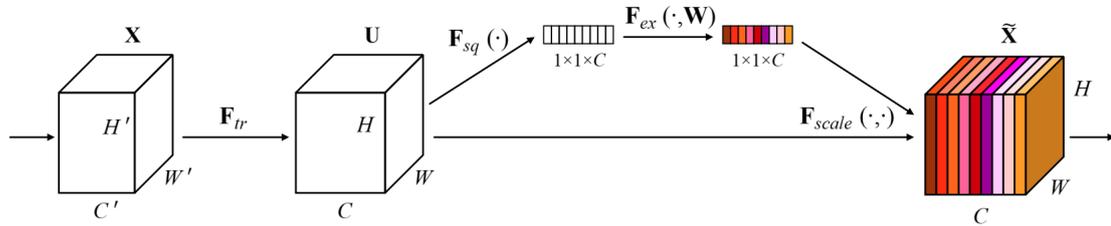


Fonte: (HU; SHEN; SUN, 2018)

A ideia base é adicionar parâmetros associados a cada canal de um bloco convolucional para a rede poder ajustar adaptativamente a ponderação de cada mapa de características. O bloco SE representado na Figura 5, possui uma camada de pooling para reduzir cada canal em um valor numérico, isso resulta em um vetor com tamanho n, sendo n o número de canais de saída do bloco convolucional, esse vetor é conectado a uma camada totalmente conectada, seguido de uma camada de ativação RELU, e outra

totalmente conectada, e finalmente uma camada de ativação sigmoide para obter os pesos de ponderação das características do mapa convolucional.

Figura 6 – Bloco SE (Squeeze-and-Excitation)



Fonte: (HU; SHEN; SUN, 2018)

Essa técnica pode ser adicionada em qualquer rede convolucional, com um acréscimo ínfimo no custo da rede. A adição dos blocos SE em uma ResNet-152 fizeram o custo computacional da rede aumentar de 11,30 GFlops para 11,32 GFlops, menos de 1% (HU; SHEN; SUN, 2018).

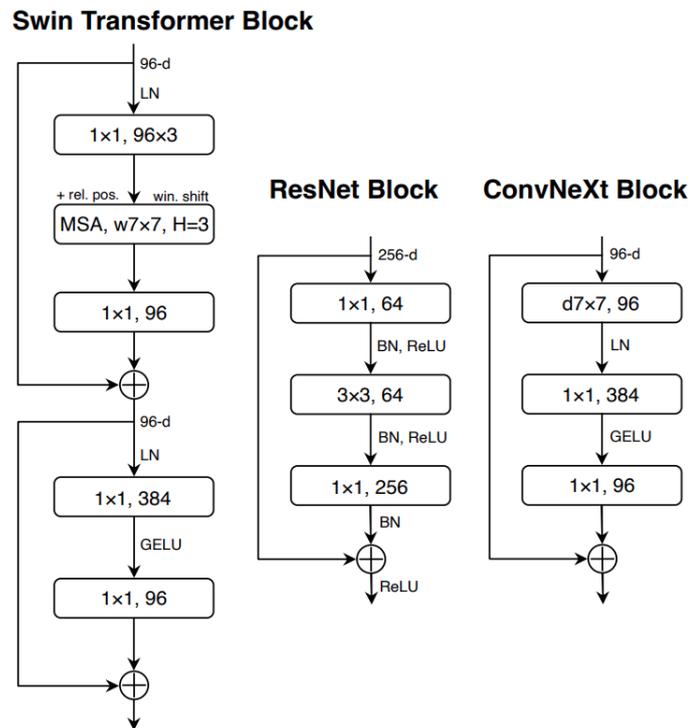
2.3.3 ConvNeXt

A ConvNeXt (LIU et al., 2022) é um modelo proposto visando criar uma rede convolucional que seja capaz de competir com as redes baseadas em transformers, para essa finalidade a arquitetura ResNet padrão é "modernizada" gradualmente. Nesse processo é descoberto um conjunto de componentes-chave para melhorar o desempenho da rede, algumas dessas mudanças até inspiradas nas redes transformers. As modificações realizadas na ConvNeXt são tanto em arquitetura (Figura 7) quanto em metodologia de treinamento.

Dentre as modificações adotadas na ConvNeXt podemos citar a troca do otimizador SGD pelo AdamW, troca da abordagem *Stem* no bloco inicial para uma rápida redução da resolução dos mapas, por uma estratégia *patchify* no bloco inicial da rede com uma camada convolucional de kernel 4x4 e *stride* 4. Também foram adotados convoluções em profundidade e *Inverted Bottleneck*, onde a dimensão interna do bloco é quatro vezes maior que a dimensão de saída, inspirada na dimensão oculta do bloco MLP em redes transformes.

Outra mudança é o uso de *kernels* de convolução maiores permitindo um campo receptivo mais abrangente, saindo dos tradicionais 3x3 para os 7x7 associados a convolução em profundidade. Substituição das camadas de ativação ReLU por GELU, amplamente utilizada em transformers, além da substituição da normalização em *batch* por normalização em camada também amplamente utilizada em redes baseadas em transformers.

Figura 7 – Bloco ConvNeXt



Fonte: (LIU et al., 2022)

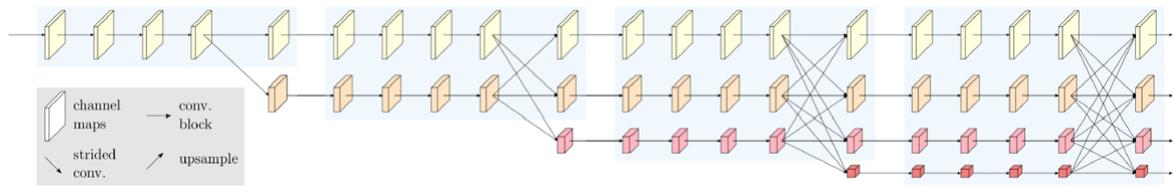
As modificações realizadas na ResNet para a criação da ConvNeXt são fortemente inspiradas em modelos baseados em transformers, tendo como resultado uma arquitetura completamente convolucional com desempenho comparável as transformers.

2.3.4 HRNet

As HRNets (WANG et al., 2020), são redes criadas para manter uma representação de alta resolução em toda extensão da rede, ao contrário da maioria das arquiteturas que gradualmente reduzem a resolução a medida que a rede se aprofunda, conectando convoluções de alta resolução e baixa resolução em série, que levam para uma representação de baixa resolução para ser usada em classificação. As HRNets foram desenvolvidas para tarefas que são sensíveis à posição como segmentação semântica, estimação de pose humana e detecção de objetos.

A arquitetura resultante consegue manter representações de alta resolução durante todo o processo, inicialmente com um fluxo de convolução de alta resolução, são gradualmente adicionados fluxos de convolução de alta para baixa resolução, e esses fluxos são mantidos em paralelos, com pontos de interconexão entre os fluxos, trocando informações entre os fluxos paralelos. A arquitetura é representado pela Figura 8

Figura 8 – Arquitetura HRNet

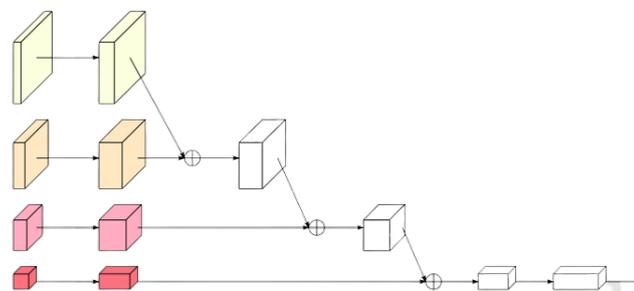


Fonte: (WANG et al., 2020)

As representações aprendidas pela HRNets, não são apenas semanticamente fortes, mas também espacialmente precisas. Isso ocorre por dois fatores: (I) A abordagem conecta fluxos de convolução em paralelo e não em série. Assim, a arquitetura consegue manter a resolução alta, ao invés de recuperar as características de alta resolução a partir de baixa resolução. (II) A maioria das arquiteturas agrega representações de alta resolução e baixa resolução com *upsampling*, na HRNet é repetido o processo de interconexão entre fluxos para aumentar as representações de alta resolução com a ajuda das representações de baixa resolução, e vice-versa. Como resultado, todas as representações de alta para baixa resolução são semanticamente fortes.

A arquitetura pode ser adaptada para classificação com a adição de um *head*, com convoluções para reduzir a resolução das características, representado na Figura 9. Inicialmente os canais de saída são aumentados de C , $2C$, $4C$, e $8C$, para 128, 256, 512 e 1024, respectivamente. Em seguida, reduzimos a amostragem de representação de alta resolução por meio de uma convolução 3×3 de duas camadas com saída e adicionamos à representação da segunda resolução mais alta ($2C$). Esse processo é repetido duas vezes para obter 1024 canais na menor resolução, seguido por uma convolução 1×1 com 2048 canais de saída e uma operação de *pooling* médio global. A representação de 2048 dimensões é inserida no classificador.

Figura 9 – Head HRNet

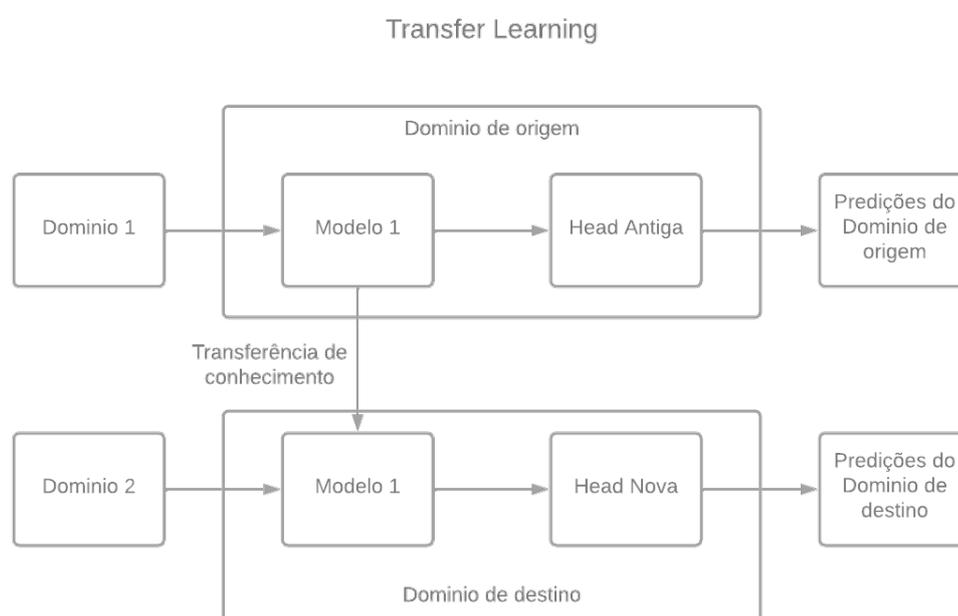


Fonte: (WANG et al., 2020)

2.4 Transfer Learning

A dependência de dados é um dos problemas mais sérios da aprendizagem profunda. A aprendizagem profunda tem uma forte dependência de dados de treinamento maciços em comparação com os métodos tradicionais de aprendizagem de máquina, porque precisa de uma grande quantidade de dados para entender seus padrões. A insuficiência de dados de treinamento é um problema inevitável em alguns domínios especiais, onde a coleta de dados é complexa e cara, o que torna extremamente difícil a criação de um conjunto de dados anotados de alta qualidade e em grande escala (TAN et al., 2018).

Figura 10 – Transfer Learning



Fonte: Autor

O conceito de *transfer learning* pode vir inicialmente da psicologia educacional. Segundo a teoria da generalização da transferência, proposta pelo psicólogo CH Judd, aprender a transferir é o resultado da generalização da experiência. É possível realizar a transferência de uma situação para outra, desde que a pessoa generalize sua experiência. De acordo com esta teoria, o pré-requisito da transferência é que deve haver uma conexão entre duas atividades de aprendizagem. Na prática, uma pessoa que aprendeu violino pode aprender piano mais rápido do que outras, pois tanto o violino quanto o piano são instrumentos musicais e podem compartilhar alguns conhecimentos comuns (ZHUANG et al., 2020).

o *transfer learning* presume que existam semelhanças e correlações entre domínios diferentes, se considerarmos modelos de classificação de imagem, uma textura, um contorno, uma forma aprendida em um domínio pode ser útil em outros, necessitando de poucos ajustes no modelo final, reduzindo a complexidade de desenvolvimento e treinamento, além

de possivelmente aumentar a capacidade de generalização e a velocidade de convergência do modelo.

2.4.1 *Fine Tuning*

O *Fine Tuning* é uma abordagem amplamente usada para o *transfer learning* ao trabalhar com modelos de aprendizagem profunda. Ela começa com um modelo pré-treinado no domínio de origem e o treina ainda mais no domínio de destino (GUO et al., 2019). Existem diferentes tipos de abordagens de *Fine Tuning*, incluindo o *Shallow Tuning* e *Deep Tuning*.

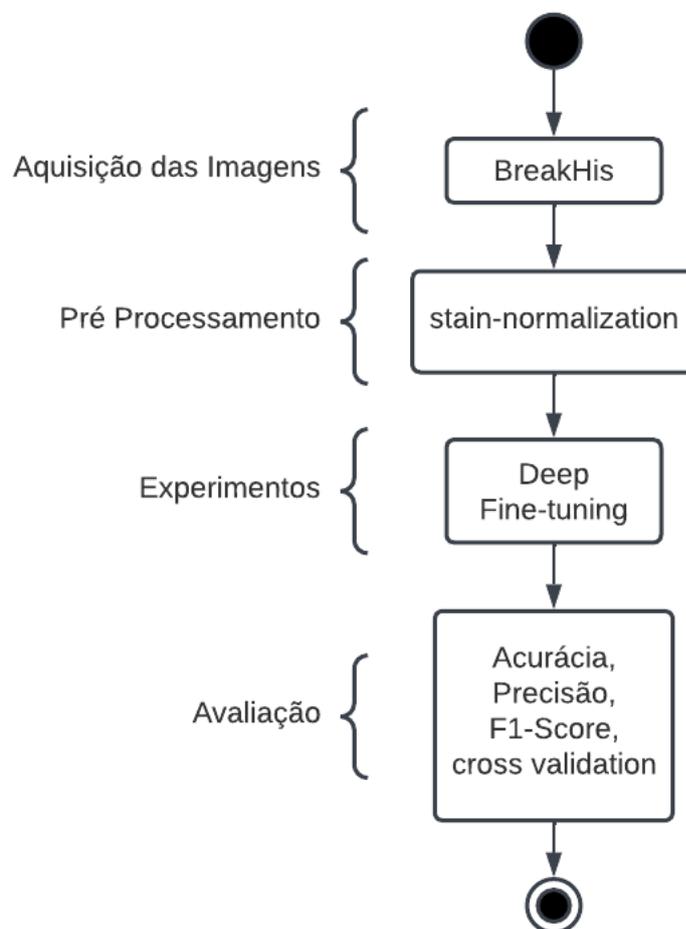
O *Shallow Tuning* é o método de *Fine Tuning* onde uma rede neural pré-treinada passa por um processo de treinamento apenas das camadas superiores e mantendo as camadas inferiores congeladas. Esta abordagem é útil quando a rede pré-treinada já aprendeu características de baixo nível relevantes para a nova tarefa, e apenas as características de nível superior precisam de ser ajustadas. No caso do *Deep Tuning* a rede inteira é retreinada a partir do modelo pré-treinado, aprendendo as características de baixo nível pertinentes para o problema.

Considerando redes neurais convolucionais aplicadas a imagens medicas, o uso de Fine Tuning demonstraram um desempenho superior, ou no pior dos casos, tão bom quanto às CNNs treinadas do princípio, e as CNNs com fine-tuning tiveram desempenhos mais robustos considerando o tamanho do conjunto de treino em relação as CNNs treinadas do princípio (TAJBAKSHI et al., 2016).

3 Metodologia

Neste capítulo é especificada a metodologia utilizada, a aquisição das imagens, o pré-processamento das imagens, as arquiteturas, os experimentos realizados e avaliação dos modelos. O fluxo das atividades realizadas é apresentado na Figura 11.

Figura 11 – Metodologia.



Fonte: Autor.

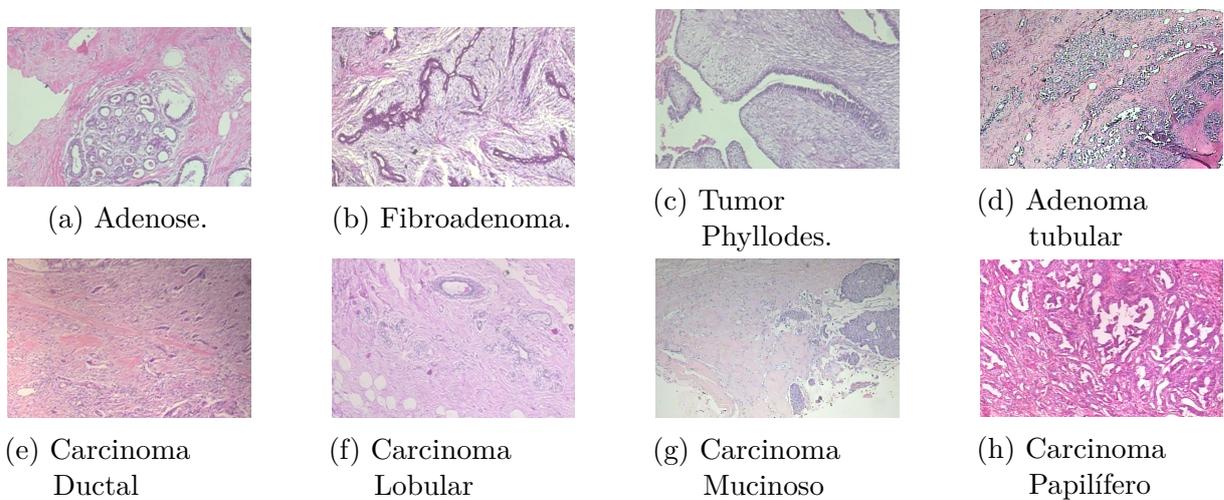
3.1 Aquisição das Imagens

Durante o trabalho foram utilizadas imagens da base de dados BreakHis (SPANHOL et al., 2015). Esta base de dados possui imagens de lâminas histológicas de tecido tumoral de mama, obtidos de 82 pacientes em um estudo clínico de janeiro de 2014 a dezembro

de 2014 no Laboratório de P&D, Paraná, Brasil. As imagens foram adquiridas usando diferentes fatores de ampliação (40X, 100X, 200X e 400X), fornecidas na resolução de 700×460 pixels, RGB de 3 canais, profundidade de 8 bits em cada canal, formato PNG.

O dataset conta com 7.909 imagens, sendo 2.480 amostras benignas e 5.429 malignas. Cada amostra possui ainda a anotação para os tipos histológicos de tumor, 4 benignos (adenose, fibroadenoma, tumor filodes e adenoma tubular) e 4 tipos malignos (carcinoma, carcinoma lobular, carcinoma mucinoso e carcinoma papilífero). A Figura 12 apresenta todos os subtipos histológicos presentes no dataset.

Figura 12 – subtipos histológicos.



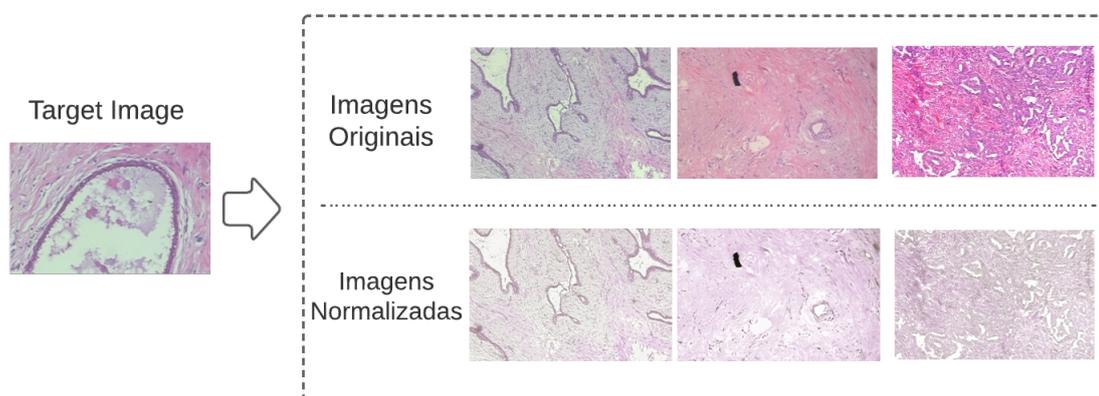
Para o trabalho foi usado apenas as imagens de 40x. Uma das vantagens desse fator de ampliação em relação aos demais é a visualização de estruturas macros das células. Como limitação, ele perde informações de textura menores.

3.2 Pré-processamento

Após a aquisição das imagens, elas foram submetidas a stain-normalization, visando normalizar o padrão de cores do dataset, visto que diferentes protocolos de coloração, equipamentos de captura, matéria-prima do corante, variação na concentração do corante de uma lâmina para outra, ou na distribuição do mesmo em uma única lâmina.

Neste estudo foi utilizado a normalização Macenko (MACENKO et al., 2009), método criado especificamente para superar inconsistência na coloração de lâminas histológicas, fazendo com que lâminas processadas e armazenadas em condições diferentes sejam processadas e transformadas para um espaço normalizado comum, visando aprimorar a análise quantitativa. O método de Macenko é utilizado em outros estudos (ALBARQOUNI et al., 2016), (ROY et al., 2019).

Figura 13 – Stain-normalization.



Fonte: Autor.

3.3 Modelos Avaliados

Este trabalho investiga a aplicação de arquiteturas convolucionais recentes para o reconhecimento em bases histopatológicas, para isso foram escolhidas arquiteturas convolucionais, mesmo que algumas características presentes nelas sejam inspiradas em modelos baseados em Transformer.

As arquiteturas SE-Resnet (HU; SHEN; SUN, 2018) e ResNext (XIE et al., 2017) foram adicionadas ao estudo por agregarem novos recursos as redes neurais convolucionais ao permitirem um desempenho superior em relação ao modelo em que foi baseado. A SE-Resnet utiliza uma MLP associada a saída do bloco convolucional para ponderar os mapas de características. A REsNeXt usa da estratégia *split-transform-merge* para aumentar a capacidade representativa dos blocos com uma complexidade computacional menor.

A ConvNext (LIU et al., 2022) foi adicionada ao estudo pelo seu alto desempenho em relação a arquiteturas baseadas em Transformer, consideradas o estado da arte para tarefas de classificação de imagens (DOSOVITSKIY et al., 2020).

A arquitetura HRNet (WANG et al., 2019) foi adicionada ao estudo por conta da sua capacidade de manter uma representação de alta resolução em toda extensão da rede, o que supomos que pode auxiliar nas tarefas de classificação de imagens histológicas, uma vez que características de textura são muito importantes para a tarefa.

Apresentamos na Tabela 1 a quantidade de parâmetros (pesos) e operações das arquiteturas analisadas. As implementações foram obtidas da biblioteca MMClassification (CONTRIBUTORS, 2020).

Tabela 1 – Detalhes das arquiteturas

Arquiteturas	Params (M)	Flops (G)
convnext base	88.59	15.36
resnext 101	88.79	16.50
seresnet 101	49.33	7.86
HRNet-W30	37.71	8.17

3.4 Transfer Learning

Ao utilizar o Transfer Learning, partimos de um conjunto de pesos pré-treinados benéficos para os modelos. Um desses benefícios é a capacidade de diminuir a quantidade de dados necessária para treinar um novo modelo. Isso é particularmente vantajoso quando há poucos dados disponíveis. Além disso, o Transfer Learning pode aumentar a eficiência do novo modelo, visto que o modelo pré-treinado já foi ensinado a reconhecer padrões complexos nos dados. Por fim, este permite que o modelo ajuste adequadamente os pesos pré-existentes a um novo domínio, o que pode reforçar ainda mais o desempenho geral do modelo.

Os experimentos foram realizados com transfer learning usando deep fine tuning: neste modelo a rede é passada com os pesos pré-treinados, mas todas as camadas continuam o treinamento com a nova base de dados.

Para fins de comparação, também conduzimos experimentos com as mesmas arquiteturas sem pré-treinamento. O objetivo é compreender como a técnica afeta o desempenho da rede para o problema em questão. O domínio de origem foi o Russakovsky et al. (2015) no qual tem bons resultados mesmo em imagens medicas (TAJBAKHSI et al., 2016).

3.4.1 Hiperparâmetros de treinamento

Todos os experimentos utilizaram hiperparâmetros padrão fornecidos nos arquivos de configuração dos modelos presentes na própria biblioteca (CONTRIBUTORS, 2020), com exceção da taxa de aprendizado, otimizador, decaimento de peso, tamanho do lote e número de épocas que mantemos fixos para todos os experimentos. Esses valores são apresentados na Tabela 2

O otimizador utilizado no estudo foi o gradiente estocástico descendente, o qual é um algoritmo clássico de otimização de primeira ordem, amplamente usado em problemas de aprendizado de máquina em grande escala devido ao seu baixo custo computacional (HE et al., 2020). A taxa de aprendizado foi fixada em 0,001 para todas as arquiteturas estudadas.

Tabela 2 – Parâmetros de treinamento

Parâmetro	Valor
Optimizador	SDG
Taxa de aprendizado	0,001
Momentum	0,9
Decaimento de peso	0,0001
Numero de épocas	120
Tamanho de lote (batch size)	32

O *Momentum* em associação com o otimizador SGD acelera a descida de gradiente, aumentando a velocidade de convergência do modelo, além de aumentar as chances de escapar de mínimos locais (WANG; LIN; ABERNETHY, 2021). Também foi adicionado decaimento de peso, uma importante técnica de regularização no otimizador de descida de gradiente estocástico (SGD). Ele adiciona um termo de penalidade à função de perda que incentiva o modelo a ter pesos menores, ajudando a prevenir o *overfitting* e melhora o desempenho de generalização. O fator de decaimento de peso determina a força desta punição, e é normalmente definido para um valor pequeno, como 0,0001 (LOSHCHILOV; HUTTER, 2017), (LI et al., 2021b), (ITO et al., 2022), (YAO et al., 2021), (ZHANG; ZHANG, 2020).

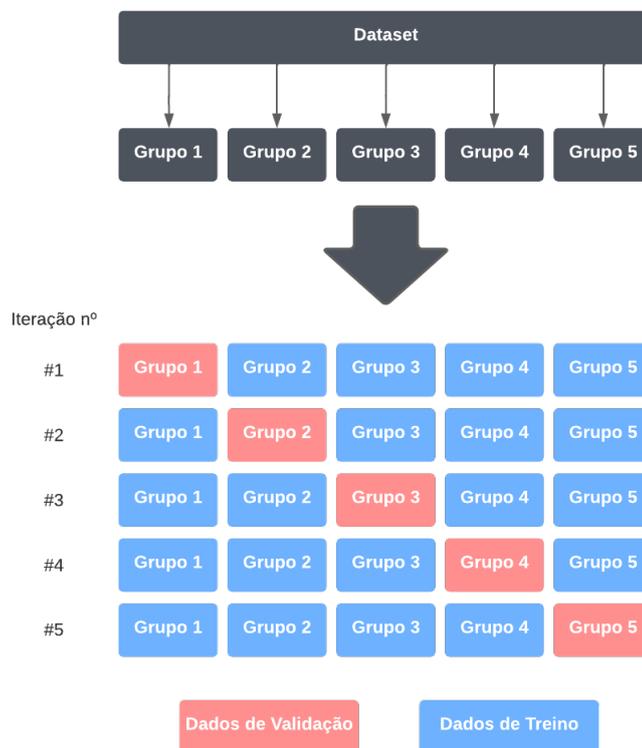
3.5 Avaliação

A validação cruzada é uma técnica amplamente usada no aprendizado de máquina para avaliar o desempenho dos modelos e evitar o *overfitting* (PRECHELT, 1998), e fornece uma maneira confiável de avaliar os modelos (MORET et al., 2012). A validação cruzada consiste em dividir o conjunto de dados em grupos (folds), a cada iteração um grupo é selecionado para avaliação e o restante para treinamento até que todos os grupos sejam avaliados.

Uma vantagem da avaliação cruzada é o uso de todas as instâncias como grupo de avaliação, tornando a avaliação menos dependente de como a divisão treino-teste foi realizada. Para avaliação da metodologia aplicada neste estudo foi empregado a validação cruzada utilizando o método k-folds com 5 grupos exemplificado na figura 14.

Para medir a eficiência dos modelos, usamos as métricas de acurácia, precisão, recall e F1-Score. A acurácia é uma métrica amplamente usada no aprendizado de máquina para avaliar o desempenho de um modelo. Ela mede a porcentagem de instâncias classificadas corretamente. Uma vantagem da acurácia como métrica é que ela é fácil de entender e interpretar. Ela fornece uma medida clara do desempenho de um modelo. No entanto, a acurácia possui desvantagens. Ela pode sugerir resultados enganosos quando usado em datasets desbalanceados, nestes casos, é possível obter uma alta acurácia prevendo a classe

Figura 14 – Validação Cruzada



Fonte: Autor

majoritária enquanto possui um desempenho ruim na classe minoritária (EBENUWA et al., 2019).

A precisão é uma métrica usada para medir a proporção de verdadeiros positivos entre todas as previsões positivas feitas por um modelo. Uma vantagem da precisão é que ela fornece uma medida da capacidade do modelo de evitar falsos positivos, o que pode ser particularmente importante em aplicações que falsos positivos podem ter consequências graves, como aplicações de diagnóstico médico.

O *F1-Score* é uma métrica comumente usada para avaliar o desempenho de um modelo em problemas de classificação binária. Ela é a média harmônica da precisão e da sensibilidade, o que a torna uma boa métrica para conjuntos de dados desequilibrados em que uma classe é muito mais predominante do que a outra. Uma vantagem da pontuação F1 é que ele fornece um único valor que resume o desempenho do modelo, facilitando a comparação dos resultados. As equações das métricas, descritas usando a matriz de confusão, são apresentadas a seguir.

$$\text{Acuracia} = \frac{VP + VN}{VP + FP + VN + FN} \quad (3.1)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (3.2)$$

$$\text{Recall} = \frac{VP}{VP + FN} \quad (3.3)$$

$$\text{F1 Score} = 2 * \frac{\text{Precisão} * \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (3.4)$$

onde: VP = Verdadeiros Positivos, VN = Verdadeiros Negativos, FP = Falsos Positivos e FN = Falsos Negativos.

4 Resultados

Neste capítulo são apresentados os resultados dos experimentos, visualizações e diagramas obtidos através da metodologia proposta neste trabalho.

Para melhor confiabilidade dos resultados, foi utilizado a validação cruzada nos experimentos, dividindo a base em 5 partes, sendo uma utilizada para testes, e as demais para treino, resultando em cinco treinos para cada arquitetura, replicado para modelos pré-treinados e treinados do zero, totalizando 40 experimentos.

Cada experimento teve sua acurácia, precisão e F1-Score medidos com base na partição de teste da validação cruzada. Foi calculado a média e o desvio padrão considerando o modelo obtido na época que teve o melhor resultado. A Tabela 3 que apresenta os resultados obtidos no experimento, com a relação do modelo utilizado, a forma de treinamento, e as métricas de avaliação. Para avaliação da metodologia foram treinados e avaliados as mesmas arquiteturas com inicialização aleatória (treinamento do zero).

Tabela 3 – Resultados obtidos nos experimentos considerando as variações de arquiteturas e forma de treinamento, onde DFT = Deep Fine Tuning, DP = Desvio Padrão.

Modelo	Treinamento	Acurácia		Precisão		F1 Score	
		Média	DP	Média	DP	Média	DP
ConvNeXt Base	Do zero	70,32	2,47	78,27	8,72	59,13	6,91
	DFT	60,35	5,35	49,82	2,11	49,03	3,02
ResNeXt 101	Do zero	84,01	3,81	86,63	1,96	81,53	4,34
	DFT	92,23	3,35	91,82	4,58	90,89	3,78
SE-Resnet 101	Do zero	83,45	4,18	86,41	2,38	79,53	5,47
	DFT	92,53	3,32	92,43	3,23	91,12	4,16
HRNet-W30	Do zero	<i>86,76</i>	3,84	<i>86,97</i>	4,59	<i>83,87</i>	4,84
	DFT	92,68	2,60	92,37	3,30	91,47	2,93

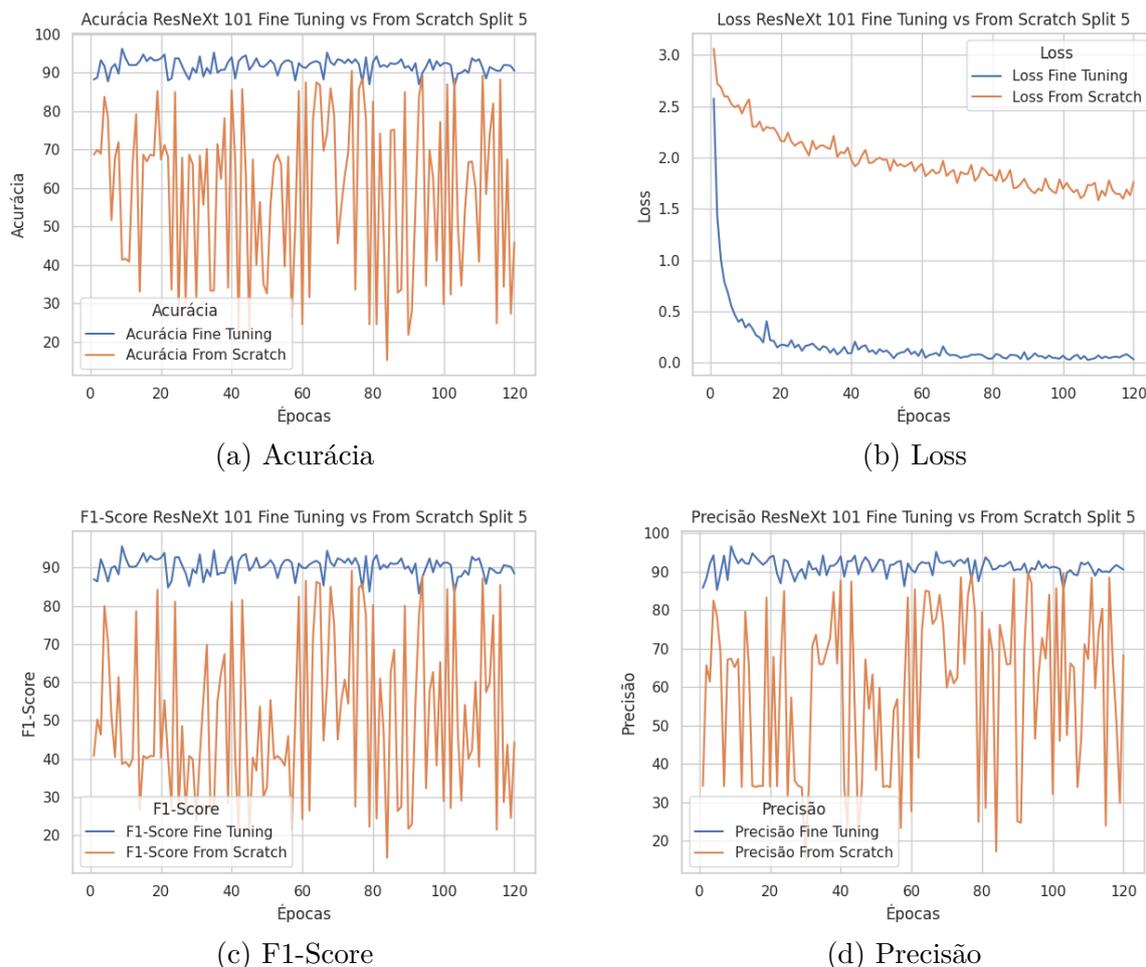
A partir da Tabela 3 é possível observar que a maioria das arquiteturas investigadas possui uma melhor acurácia ao utilizar DFT em seu treinamento, com exceção da ConvNeXt onde é analisada na seção de discussão. Os melhores resultados obtidos em acurácia e f1-score, foram obtidos com a HRNet, utilizando Deep fine tuning. Os resultados da HRNet foram os melhores dentre os experimentos com treinamento do zero.

Para melhor entendermos o comportamento de cada arquitetura, apresentamos abaixo a análise gráfica dos treinamentos das redes.

4.1 Análise do modelo ResNeXt 101

A arquitetura ResNeXt foi fortemente inspirada na arquitetura ResNet, com a adição do conceito de cardinalidade, criando camadas convolucionais em paralelo no bloco.

Figura 15 – Treino ResNeXt-101 Split 5

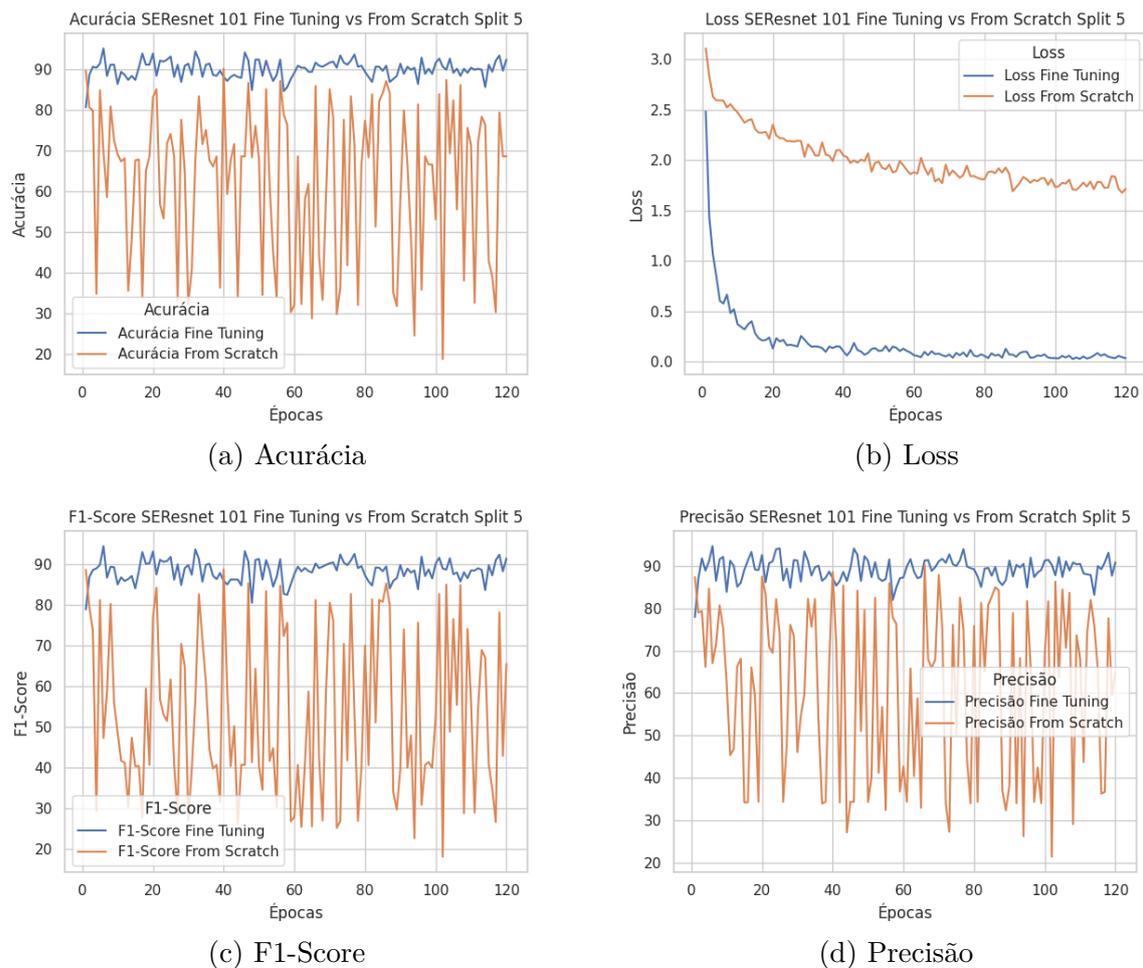


O modelo ResNeXt-101 apresenta uma melhor acurácia nos experimentos com DFT, superior ao dos experimentos com treinamento do zero. Na Figura 15 podemos verificar no gráfico de acurácia o treinamento do zero apresenta grande variância em relação ao treinamento com DFT. A partir do gráfico de Loss é possível observar que o modelo treinado com DFT é mais estável e consegue atingir valores menores de loss ao longo do treinamento. Para a arquitetura ResNeXt os experimentos com DFT além de conseguir uma acurácia absoluta maior, apresenta um desempenho mais estável no decorrer das épocas.

4.2 Análise do modelo SEResNet 101

A arquitetura SEResNet foi baseada na ResNet, com a adição de um recurso chamado Squeeze-and-Excitation (SE), que pondera os mapas de características na saída do bloco convolucional.

Figura 16 – Treino seresnet-101 Split 5

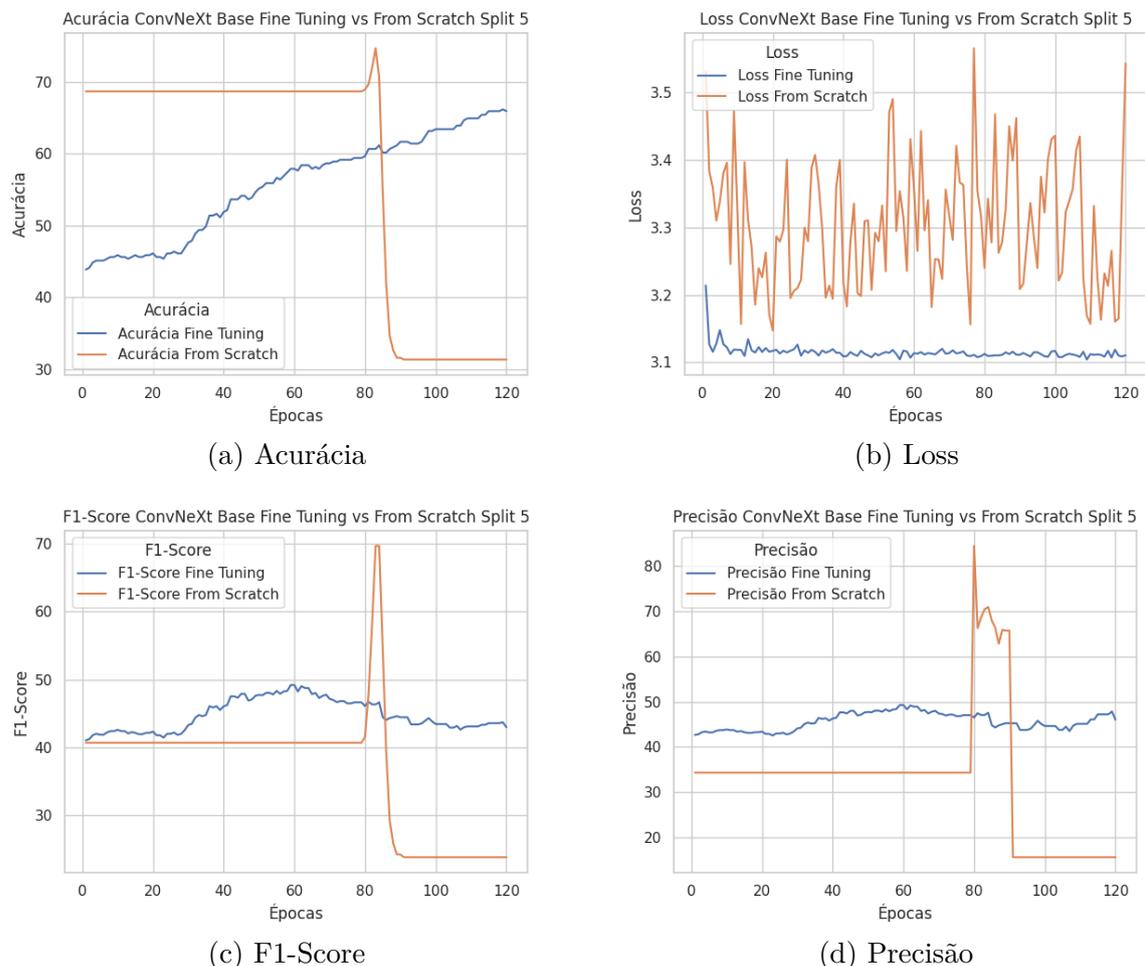


Assim como a ResNeXt-101, a SEResNet-101 apresentou uma acurácia superior no experimento com DFT além de uma variância menor em comparação ao modelo treinado do zero, como fica apresentado na Figura 16. A acurácia máxima do experimento com DFT foi 9 pontos percentuais superiores ao experimento com treinamento do zero. Em ambos os gráficos, o experimento com DFT apresentou um comportamento mais estável e em valores absolutos superiores ao treinamento do zero.

4.3 Análise do modelo ConvNeXt Base

A arquitetura ConvNeXt Base foi desenvolvida como abordagem completamente convolucional com desempenho comparável as redes baseadas em Transformer.

Figura 17 – Treino ConvNeXt Split 5



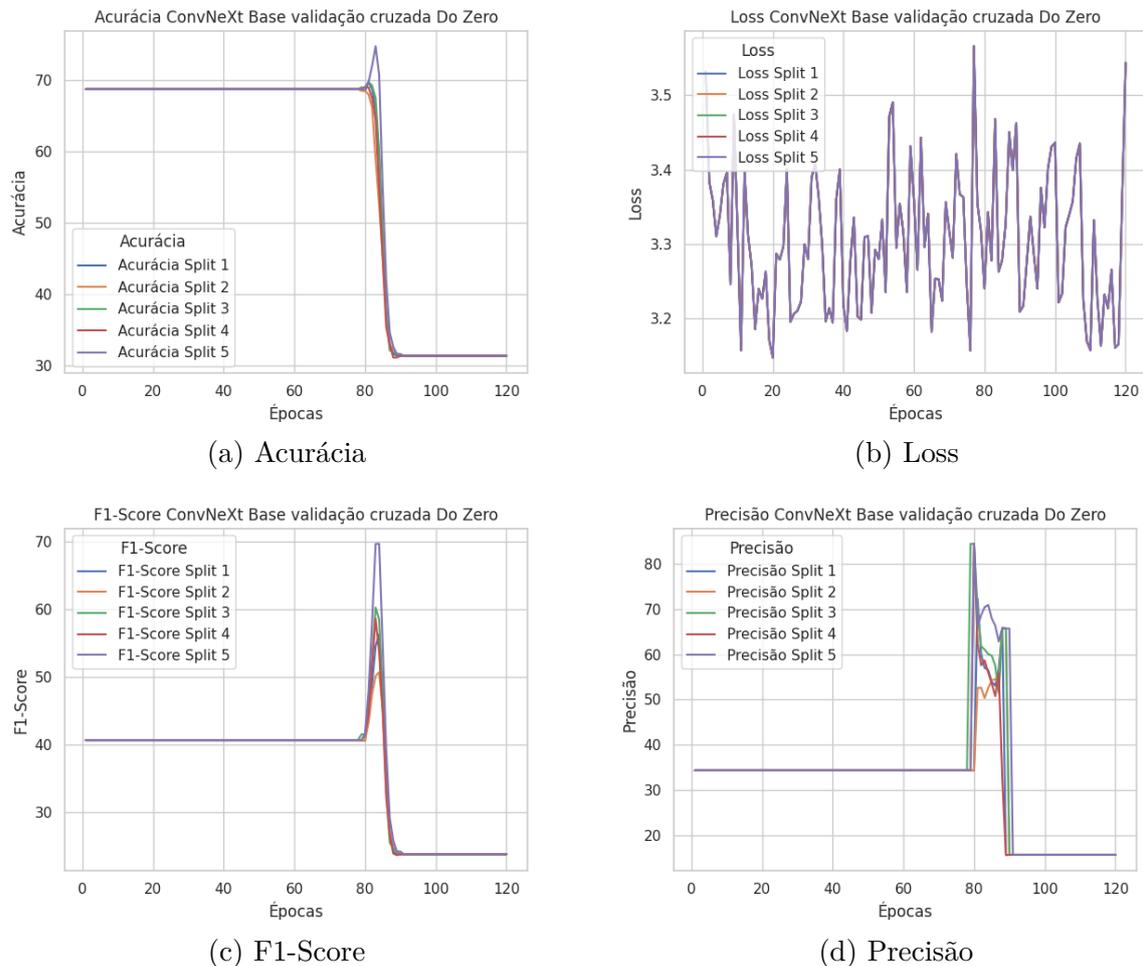
Como podemos ver na Figura 17 o modelo treinado do zero não conseguiu convergir. A curva de acurácia se mantém estável, até a época 80 quando a acurácia cai subitamente. O experimento com DFT apresenta uma tendência de crescimento, mas a velocidade de crescimento é bem menor do que nas arquiteturas anteriormente discutidas. O gráfico de Loss para o treinamento do zero apresenta variância sem uma tendência clara.

Neste caso o gráfico de apenas uma divisão da validação cruzada não nos permite ter uma avaliação aprofundada do experimento. Na Figura 18 podemos ver uma comparação das divisões da validação cruzada com o modelo treinado do zero.

A partir da Figura 18 podemos perceber que o comportamento se repetiu nas outras divisões da validação cruzada, não sendo uma especificidade do split 5. Na Figura 19 podemos ver uma comparação das divisões de treinamento para o experimento com DFT.

Na Figura 19 fica claro que os Splits 1 e 3, entraram em overfit, no gráfico da acurácia podemos analisar a forte queda para as duas divisões. Considerando que este modelo dentre os analisados neste estudo possui a maior quantidade de parâmetros aumentando

Figura 18 – Treino ConvNeXt Splits com treinamento do zero



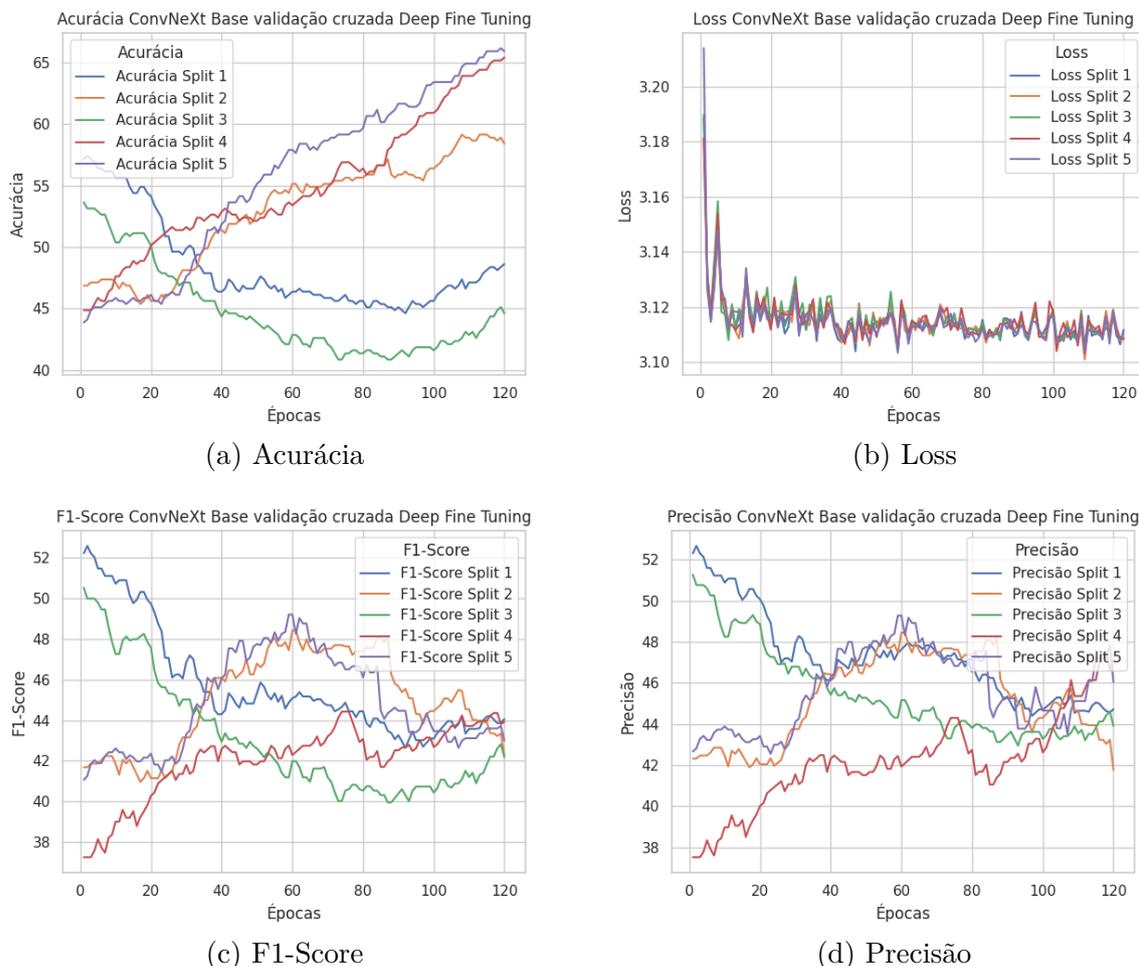
as chances de overfit. Para os modelos treinados do zero, possivelmente enfrentaram um problema de gradiente vanishing, o desaparecimento do sinal de retro propagação durante o treinamento, comum em redes neurais muito profundas.

4.4 Análise do modelo HRNet

A arquitetura HRNet foi elaborada visando manter representações em alta definição ao longo da rede, para algumas tarefas de visão computacional como, estimativa de pose e localização de objetos essas representações permitem uma melhor precisão.

Já a arquitetura HRNet obteve os melhores resultados considerando apenas os experimentos com treinamento do zero. Os experimentos com DFT chegaram no melhor resultado em acurácia e F1-Score, além de um resultado consistente em precisão. Analisando a Figura 20 mesmo com bons resultados, o gráfico de acurácia no experimento com treinamento do zero apresentou alta variância partindo de 90% a 40% em poucas épocas, o que sugere que a taxa de treinamento estava alta para a arquitetura, podendo ter resultados melhores com um ajuste neste hiperparâmetro.

Figura 19 – Treino ConvNeXt Splits com DFT



Os experimentos com DFT além de obterem melhores resultados absolutos mantiveram resultados estáveis durante o treinamento, com variância menor em comparação ao modelo treinado do zero em todas as métricas de avaliação utilizadas neste estudo.

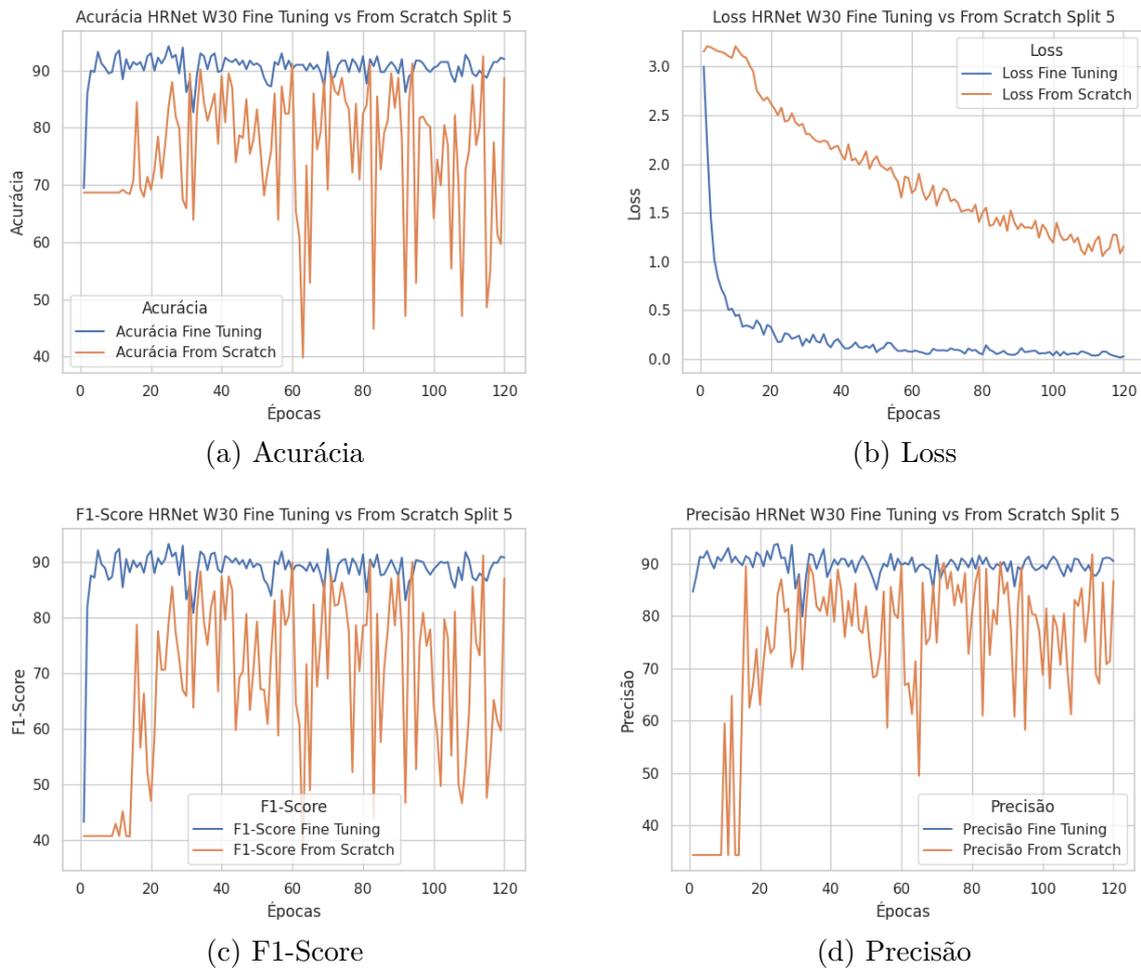
Dentre todos os experimentos realizados neste estudo, a HRNet obteve os resultados mais promissores, com bom desempenho tanto com DFT quanto no treinamento do Zero.

4.5 Análise dos experimentos com DFT

Em suma os experimentos com Deep Fine Tuning apresentaram resultados mais robustos não só em números absolutos das métricas avaliadas, como em estabilidade durante o histórico de treinamento, apresentando gráficos com menor variância, em todas as métricas apuradas. Também foi observado uma maior velocidade de convergência dos experimentos que utilizaram DFT em relação ao treinamento do zero.

Considerando a arquitetura ConvNeXt, os experimentos com treinamento do zero não conseguiram convergir, apresentando um comportamento instável nas métricas

Figura 20 – Treino HRNet-W30 Split 5



analisadas e com o gráfico de loss sem uma tendência clara, isso possivelmente ocorreu pelo problema de gradiente vanishing, enquanto os experimentos com a utilização de DFT, alguns splits da validação cruzada conseguiram convergir, outros apresentaram um comportamento típico de overfit com a redução das métricas na partição de validação. Neste caso a utilização de Deep Fine Tuning permitiu que o modelo fosse treinável com os parâmetros analisados, o que sugere que a utilização da técnica permite a redução dos esforços de otimização dos hiperparâmetros.

5 Conclusão

Métodos computacionais para análise histopatológica tem potencial para acelerar o processo de estadiamento de diversos tipos de câncer, fornecendo ferramentas de diagnóstico auxiliado por computador (CAD), aumentando o rendimento de médicos patologistas, melhorando a confiança dos resultados e os tornando mais reproduzíveis.

Este trabalho se propõe a avaliar o desempenho da utilização de deep fine tuning para classificação binária do dataset breakhis, nas arquiteturas ResNeXt, SEResNet, ConvNeXt e HRNet. Como baseline de comparação, as mesmas arquiteturas foram treinadas do zero com as mesmas divisões de validação cruzada. Todas as arquiteturas foram treinadas com os mesmos hiper-parâmetros.

Neste estudo arquitetura HRNet-W30 apresentou os melhores resultados com e sem fine tuning. Possivelmente as representações em alta resolução auxiliaram na classificação das texturas das imagens histopatológicas e no reconhecimento das estruturas presentes no tecido. O fato de ser a arquitetura com menor quantidade de parâmetros também possa ter contribuído evitando overfit.

A ResNeXt e SEResNet apresentaram resultados semelhantes, ambas arquiteturas fortemente inspiradas na ResNet apresentaram resultados semelhantes a HR-Net com poucos décimos de diferença nos experimentos com fine tuning. A diferença se torna significativa nos experimentos treinados do zero com a HR-Net possuindo resultados mais consistente neste caso.

No estudo onde a ConvNeXt foi proposta, além de uma nova arquitetura, foi utilizada diversos tipos de data-augmentation, metodologias de treinamento e otimizações de hiper-parâmetros para alcançar os resultados demonstrados. Este estudo não reproduz as otimizações propostas naquele estudo, o que possivelmente contribuiu para os resultados inferiores desempenhados pela arquitetura neste trabalho.

Além da reprodução da metodologia de treinamento da ConvNext, este trabalho não otimiza nenhuma das arquiteturas delimitadas, se limitando a avaliar as arquiteturas a partir dos mesmos hiperparâmetros. Essa limitação impede que atinjamos os melhores resultados possíveis para cada arquitetura, mas permite uma comparação equivalente entre elas.

Foi constatado que a utilização de fine-tuning permite a geração de modelos mais robustos, com velocidade de convergência maiores, maior estabilidade das métricas durante o treinamento, resultados absolutos maiores nas métricas avaliadas. A única exceção observada neste estudo foi a ConvNext, como discutido anteriormente a metodologia de

treinamento não foi reproduzida, porém podemos observar que a utilização de fine-tuning permitiu que a rede fosse treinável mesmo sem a otimização de seus hiperparâmetros, sugerindo uma maior facilidade de otimização ao utilizar fine tuning.

Um caminho promissor para trabalhos futuros é a otimização da arquitetura HR-Net, no presente estudo a HR-Net demonstrou os melhores resultados, podendo ser uma abordagem promissora para lidar com esse desafio. Ao otimizar essa arquitetura seria possível explorar sua capacidade de capturar detalhes finos e sutis presentes nas imagens histopatológicas importantes para identificação de padrões e características relevantes para sua classificação.

Referências

- ALBARQOUNI, S.; BAUR, C.; ACHILLES, F.; BELAGIANNIS, V.; DEMIRCI, S.; NAVAB, N. Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1313–1321, 2016. Citado na página 27.
- ARESTA, G.; ARAÚJO, T.; KWOK, S.; CHENNAMSETTY, S. S.; SAFWAN, M.; ALEX, V.; MARAMI, B.; PRASTAWA, M.; CHAN, M.; DONOVAN, M. et al. Bach: Grand challenge on breast cancer histology images. *Medical image analysis*, Elsevier, v. 56, p. 122–139, 2019. Citado na página 13.
- BARDOU, D.; ZHANG, K.; AHMAD, S. M. Classification of breast cancer based on histology images using convolutional neural networks. *Ieee Access*, IEEE, v. 6, p. 24680–24693, 2018. Citado na página 13.
- BAYRAMOGLU, N.; KANNALA, J.; HEIKKILÄ, J. Deep learning for magnification independent breast cancer histopathology image classification. In: IEEE. *2016 23rd International conference on pattern recognition (ICPR)*. Cancun, Mexico, 2016. p. 2440–2445. Citado na página 13.
- BRAY, F.; LAVERSANNE, M.; WEIDERPASS, E.; SOERJOMATARAM, I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*, v. 127, n. 16, p. 3029–3030, 2021. Disponível em: <<https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.1002/cncr.33587>>. Citado na página 12.
- CONTRIBUTORS, M. *OpenMMLab's Image Classification Toolbox and Benchmark*. 2020. <<https://github.com/open-mmlab/mmlclassification>>. Citado 2 vezes nas páginas 28 e 29.
- DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; WEISSENBORN, D.; ZHAI, X.; UNTERTHINER, T.; DEHGHANI, M.; MINDERER, M.; HEIGOLD, G.; GELLY, S. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. Citado na página 28.
- EBENUWA, S. H.; SHARIF, M. S.; AL-NEMRAT, A.; AL-BAYATTI, A. H.; ALALWAN, N.; ALZHRANI, A. I.; ALFARRAJ, O. Variance ranking for multi-classed imbalanced datasets: a case study of one-versus-all. *Symmetry*, MDPI, v. 11, n. 12, p. 1504, 2019. Citado na página 31.
- FAZAL, K.; MUMTAZ, H.; BASIR, A.; BUTT, M. M.; AMAN, M.; KHAN, F.; AHMAD, I.; SHAMS, Y.; BASHIR, A. et al. Color doppler ultrasound for diagnosis of testicular carcinoma: A comparison with gold standard histopathology. *Annals of Medicine and Surgery*, Elsevier, v. 84, p. 104938, 2022. Citado na página 13.
- GHANATE, A.; KOTHIWALE, S.; SINGH, S.; BERTRAND, D.; KRISHNA, C. M. Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario. *Journal of Biomedical Optics*, Society of Photo-Optical Instrumentation Engineers, v. 16, n. 2, p. 025003–025003, 2011. Citado na página 13.

- GOUR, M.; JAIN, S.; KUMAR, T. S. Residual learning based cnn for breast cancer histopathological image classification. *International Journal of Imaging Systems and Technology*, Wiley Online Library, v. 30, n. 3, p. 621–635, 2020. Citado na página 13.
- GUO, Y.; SHI, H.; KUMAR, A.; GRAUMAN, K.; ROSING, T.; FERIS, R. Spottune: transfer learning through adaptive fine-tuning. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Long Beach, California, USA: Institute of Electrical and Electronics Engineers (IEEE), 2019. p. 4805–4814. Citado na página 25.
- GURCAN, M. N.; BOUCHERON, L. E.; CAN, A.; MADABHUSHI, A.; RAJPOOT, N. M.; YENER, B. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, IEEE, v. 2, p. 147–171, 2009. Citado na página 15.
- HE, K.; ZHANG, M.; ZHOU, J.; JIN, Y.; LI, C.-m. Stochastic item descent method for large scale equal circle packing problem. *arXiv preprint arXiv:2001.08540*, 2020. Citado na página 29.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, Nevada, USA.: Institute of Electrical and Electronics Engineers (IEEE), 2016. p. 770–778. Citado na página 19.
- HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Salt Lake City, Utah, USA: Institute of Electrical and Electronics Engineers (IEEE), 2018. p. 7132–7141. Citado 3 vezes nas páginas 20, 21 e 28.
- ITO, H.; APRILPYONE, M.; SHIOTA, S.; KIYA, H. et al. Access control of semantic segmentation models using encrypted feature maps. *APSIPA Transactions on Signal and Information Processing*, Now Publishers, Inc., v. 11, n. 1, 2022. Citado na página 30.
- KAMEL, H. F. M.; AL-AMODI, H. S. B. Cancer biomarkers. In: . Rijeka: InTech Rijeka, 2016. cap. 1. Disponível em: <<https://doi.org/10.5772/62421>>. Citado na página 12.
- KANTHIKAR, S. N.; NIKUMBH, D. B.; DRAVID, N. V. Histopathological overview of central nervous system tumours in north maharashtra, india: a single center study. *Indian Journal of Pathology and Oncology*, v. 4, n. 1, p. 80–84, 2017. Citado na página 13.
- KHAN, A. M.; RAJPOOT, N.; TREANOR, D.; MAGEE, D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE transactions on Biomedical Engineering*, IEEE, v. 61, n. 6, p. 1729–1738, 2014. Citado na página 17.
- KOMURA, D.; ISHIKAWA, S. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, Elsevier, v. 16, p. 34–42, 2018. Citado na página 16.
- KRENKER, A.; BEŠTER, J.; KOS, A. Introduction to the artificial neural networks. *Artificial Neural Networks: Methodological Advances and Biomedical Applications*. InTech, p. 1–18, 2011. Citado na página 18.

- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGESS, C.; BOTTOU, L.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. New York, NY, USA: Curran Associates, Inc., 2012. v. 25. Citado 2 vezes nas páginas 18 e 19.
- KWAK, J. T.; HEWITT, S. M.; SINHA, S.; BHARGAVA, R. Multimodal microscopy for automated histologic analysis of prostate cancer. *BMC cancer*, BioMed Central, v. 11, n. 1, p. 1–16, 2011. Citado na página 12.
- LAAK, J. Van der; LITJENS, G.; CIOMPI, F. Deep learning in histopathology: the path to the clinic. *Nature medicine*, Nature Publishing Group US New York, v. 27, n. 5, p. 775–784, 2021. Citado na página 13.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado na página 18.
- LI, M.; MA, X.; CHEN, C.; YUAN, Y.; ZHANG, S.; YAN, Z.; CHEN, C.; CHEN, F.; BAI, Y.; ZHOU, P.; LV, X.; MA, M. Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images. *IEEE Access*, v. 9, p. 53687–53707, 2021. Citado na página 12.
- LI, T.; XIE, P.; LIU, J.; CHEN, M.; ZHAO, S.; KANG, W.; ZUO, K.; LI, F. Automated diagnosis and localization of melanoma from skin histopathology slides using deep learning: A multicenter study. *Journal of Healthcare Engineering*, Hindawi Limited, v. 2021, p. 1–6, 2021. Citado na página 30.
- LIU, Z.; MAO, H.; WU, C.-Y.; FEICHTENHOFER, C.; DARRELL, T.; XIE, S. A convnet for the 2020s. *arXiv preprint arXiv:2201.03545*, 2022. Citado 3 vezes nas páginas 21, 22 e 28.
- LOSHCHILOV, I.; HUTTER, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. Citado na página 30.
- MACENKO, M.; NIETHAMMER, M.; MARRON, J. S.; BORLAND, D.; WOOSLEY, J. T.; GUAN, X.; SCHMITT, C.; THOMAS, N. E. A method for normalizing histology slides for quantitative analysis. In: IEEE. *2009 IEEE international symposium on biomedical imaging: from nano to macro*. Boston, Massachusetts, 2009. p. 1107–1110. Citado 2 vezes nas páginas 17 e 27.
- MALMSTRØM, M. L.; SĂFTOIU, A.; VILMANN, P.; KLAUSEN, T. W.; GÖGENUR, I. Endoscopic ultrasound for staging of colonic cancer proximal to the rectum: A systematic review and meta-analysis. *Endoscopic ultrasound*, Wolters Kluwer–Medknow Publications, v. 5, n. 5, p. 307, 2016. Citado na página 12.
- MORET, D.; BREITMAN, K.; AMORIM, E.; TALAVERA, J.; MILIDIU, R.; VITERBO, J. Double dip map-reduce for processing cross validation jobs. In: *Proceedings of the 27th Annual ACM Symposium on Applied Computing*. New York, NY, USA: Association for Computing Machinery, 2012. p. 473–477. Citado na página 30.
- POLETTI, V.; RAVAGLIA, C.; BUCCIOLI, M.; TANTALOCCO, P.; PICIUCCHI, S.; DUBINI, A.; CARLONI, A.; CHILOSI, M.; TOMASSETTI, S. Idiopathic pulmonary fibrosis: diagnosis and prognostic evaluation. *Respiration*, Karger Publishers, v. 86, n. 1, p. 5–12, 2013. Citado na página 12.

PRECHELT, L. Automatic early stopping using cross validation: quantifying the criteria. *Neural networks*, Elsevier, v. 11, n. 4, p. 761–767, 1998. Citado na página 30.

REINHARD, E.; ADHIKHMIN, M.; GOOCH, B.; SHIRLEY, P. Color transfer between images. *IEEE Computer graphics and applications*, IEEE, v. 21, n. 5, p. 34–41, 2001. Citado na página 16.

ROY, K.; BANIK, D.; BHATTACHARJEE, D.; NASIPURI, M. Patch-based system for classification of breast histology images using deep learning. *Computerized Medical Imaging and Graphics*, Elsevier, v. 71, p. 90–103, 2019. Citado na página 27.

ROY, S.; kumar Jain, A.; LAL, S.; KINI, J. A study about color normalization methods for histopathology images. *Micron*, v. 114, p. 42–61, 2018. ISSN 0968-4328. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0968432818300982>>. Citado na página 16.

RUBIN, R.; STRAYER, D. S.; RUBIN, E. et al. *Rubin's pathology: clinicopathologic foundations of medicine*. philadelphia pennsylvania: Lippincott Williams & Wilkins, 2008. Citado na página 12.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015. Citado na página 29.

SALEHI, P.; CHALECHALE, A. Pix2pix-based stain-to-stain translation: a solution for robust stain normalization in histopathology images analysis. In: IEEE. *2020 International Conference on Machine Vision and Image Processing (MVIP)*. Qom, Iran, 2020. p. 1–7. Citado na página 17.

SANTOS, M. d. O.; LIMA, F. C. d. S. d.; MARTINS, L. F. L.; OLIVEIRA, J. F. P.; ALMEIDA, L. M. d.; CANCELA, M. d. C. Estimativa de incidência de câncer no brasil, 2023-2025. *Revista Brasileira de Cancerologia*, v. 69, n. 1, p. e–213700, fev. 2023. Disponível em: <<https://rbc.inca.gov.br/index.php/revista/article/view/3700>>. Citado na página 12.

SHAFFIE, A. M. A. Machine learning approaches for lung cancer diagnosis. 2021. Citado na página 12.

SPANHOL, F. A.; OLIVEIRA, L. S.; PETITJEAN, C.; HEUTTE, L. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, IEEE, v. 63, n. 7, p. 1455–1462, 2015. Citado na página 26.

SUNG, H.; FERLAY, J.; SIEGEL, R. L.; LAVERSANNE, M.; SOERJOMATARAM, I.; JEMAL, A.; BRAY, F. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, v. 71, n. 3, p. 209–249, 2021. Disponível em: <<https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21660>>. Citado na página 12.

SWATHI, T.; KRISHNA, S.; RAMESH, M. V. A survey on breast cancer diagnosis methods and modalities. In: IEEE. *2019 international conference on wireless*

communications signal processing and networking (WiSPNET). Chennai, India, 2019. p. 287–292. Citado na página 13.

TAJBAKSHI, N.; SHIN, J. Y.; GURUDU, S. R.; HURST, R. T.; KENDALL, C. B.; GOTWAY, M. B.; LIANG, J. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1299–1312, 2016. Citado 2 vezes nas páginas 25 e 29.

TAN, C.; SUN, F.; KONG, T.; ZHANG, W.; YANG, C.; LIU, C. A survey on deep transfer learning. In: SPRINGER. *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III 27*. Rhodes, Greece, 2018. p. 270–279. Citado na página 24.

VAHADANE, A.; PENG, T.; SETHI, A.; ALBARQOUNI, S.; WANG, L.; BAUST, M.; STEIGER, K.; SCHLITTER, A. M.; ESPOSITO, I.; NAVAB, N. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 8, p. 1962–1971, 2016. Citado na página 17.

WANG, J.; SUN, K.; CHENG, T.; JIANG, B.; DENG, C.; ZHAO, Y.; LIU, D.; MU, Y.; TAN, M.; WANG, X.; LIU, W.; XIAO, B. Deep high-resolution representation learning for visual recognition. *TPAMI*, 2019. Citado na página 28.

WANG, J.; SUN, K.; CHENG, T.; JIANG, B.; DENG, C.; ZHAO, Y.; LIU, D.; MU, Y.; TAN, M.; WANG, X. et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 43, n. 10, p. 3349–3364, 2020. Citado 2 vezes nas páginas 22 e 23.

WANG, J.-K.; LIN, C.-H.; ABERNETHY, J. Escaping saddle points faster with stochastic momentum. *arXiv preprint arXiv:2106.02985*, 2021. Citado na página 30.

WANGHAO. *Stain Normalization: Some stain normalization code*. 2020. https://github.com/wanghao14/Stain_Normalization. Último acesso em 18 de maio 2023. Citado na página 16.

XIE, S.; GIRSHICK, R.; DOLLAR, P.; TU, Z.; HE, K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Hawaii: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. p. 1492–1500. Citado 2 vezes nas páginas 19 e 28.

YAN, R.; REN, F.; WANG, Z.; WANG, L.; ZHANG, T.; LIU, Y.; RAO, X.; ZHENG, C.; ZHANG, F. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, Elsevier, v. 173, p. 52–60, 2020. Citado na página 13.

YAO, Y.; CHEN, T.; XIE, G.-S.; ZHANG, C.; SHEN, F.; WU, Q.; TANG, Z.; ZHANG, J. Non-salient region object mining for weakly supervised semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Virtual Conference: Institute of Electrical and Electronics Engineers (IEEE), 2021. p. 2623–2632. Citado na página 30.

ZEROUAOUI, H.; IDRI, A. Deep hybrid architectures for binary classification of medical breast cancer images. *Biomedical Signal Processing and Control*, Elsevier, v. 71, p. 103226, 2022. Citado na página 14.

ZHANG, X.; ZHANG, X. Global learnable pooling with enhancing distinctive feature for image classification. *IEEE Access*, IEEE, v. 8, p. 98539–98547, 2020. Citado na página 30.

ZHUANG, F.; QI, Z.; DUAN, K.; XI, D.; ZHU, Y.; ZHU, H.; XIONG, H.; HE, Q. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, IEEE, v. 109, n. 1, p. 43–76, 2020. Citado na página 24.