Universidade Estadual de Feira de Santana

Programa de Pós-Graduação em Ciência da Computação

Aprendizado Auto-supervisionado para Classificação e Recuperação de Imagens Histopatológicas

José Solenir Lima Figuerêdo

Feira de Santana

2025

Universidade Estadual de Feira de Santana

Programa de Pós-Graduação em Ciência da Computação

José Solenir Lima Figuerêdo

Aprendizado Auto-supervisionado para Classificação e Recuperação de Imagens Histopatológicas

Dissertação apresentada à Universidade Estadual de Feira de Santana como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Orientador: Rodrigo Tripodi Calumby

Feira de Santana

2025

Ficha Catalográfica - Biblioteca Central Julieta Carteado - UEFS

F495a

Figuerêdo, José Solenir Lima

Aprendizado auto-supervisionado para classificação e recuperação de imagens histopatológicas / José Solenir Lima Figuerêdo. – 2025.

91 f.: il.

Orientador: Rodrigo Tripodi Calumby

Dissertação (mestrado) – Universidade Estadual de Feira de Santana, Programa de Pós-Graduação em Ciência da Computação, Feira de Santana, 2025.

1. Patologia digital. 2. Histopatologia. 3.Fine-tuning. 4. Processamento de imagens. 5. Lesões . I. Calumby, Rodrigo Tripodi, orient. II. Universidade Estadual de Feira de Santana. III. Título.

CDU 004.383.8:616-001

Daniela Machado Sampaio Costa - Bibliotecária - CRB-5/2077

UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
Autorizada pelo Decreto Federal nº 77.496 de 27/04/1976
Reconhecida pela Portaria Ministerial nº 874 de 19/12/1986
Recredenciada pelo Decreto nº 9.271 de 14/12/2004
Recredenciada pelo Decreto nº 17.228 de 25/11/2016
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE MESTRADO EM CIÊNCIA DA COMPUTAÇÃO Nº 36

No dia 26 de Março de 2025, às 09:00, na Sala S8 do Labotec III - Campus da UEFS, realizou-se a Sessão Pública de Defesa de Dissertação de Mestrado em Ciência da Computação número 36, do(a) mestrando(a) **José Solenir Lima Figuerêdo**, matrícula nº 14111182, intitulada *Aprendizado Auto-supervisionado para Classificação e Recuperação de Imagens Histopatológicas* do Programa de Pós-Graduação em Ciência da Computação (PGCC). Inicialmente, a Banca Examinadora foi instalada, sendo composta pelos seguintes membros: o(a) Orientador(a) do(a) mestrando(a) e Presidente da Banca Examinadora, Dr(a). Rodrigo Tripodi Calumby (UEFS), Dr(a). Angelo Conrado Loula (UEFS) e Dr(a). Tiago Amador Coelho (UEFS). Em seguida, foram esclarecidos os procedimentos e a palavra foi passada ao(à) mestrando(a), que apresentou o seu trabalho. Ao final da apresentação, a Banca Examinadora passou à arguição do(a) candidato(a). Ato contínuo, a Banca Examinadora reuniu-se para elaborar seu parecer final. Concluída a reunião, foi lido o parecer final sobre a dissertação apresentada, tendo a Banca Examinadora atribuído o conceito **APROVADO** à referida dissertação, sendo esta aprovação um requisito parcial para a obtenção do título de Mestre em Ciência da Computação. O(A) mestrando(a) terá 90 (noventa) dias para realizar modificações consideradas essenciais para a aprovação da dissertação, conforme o parecer exarado pela Banca Examinadora, anexo a este documento. Nada mais havendo a tratar, foi encerrada a sessão e lavrada a presente ata, abaixo assinada pelo Presidente e demais membros da Banca Examinadora.

Feira de Santana, 26 de Março de 2025.

Dr(a). Rodrigo Tripodi Calumby (Presidente)

Dr(a). Angelo Conrado Loula

Dr(a). Tiago Amador Coelho

Abstract

Histopathology analyzes tissues and lesions for disease diagnosis, but manual analysis is often laborious and error-prone. To improve accuracy and reduce the workload for pathologists, computer-aided diagnosis systems can be used, including disease classifiers and content-based image retrieval tools. However, to be used, these systems need to extract features from images, but labeling these images also requires a great deal of effort, and is often time-consuming and expensive. To mitigate these problems, approaches centered on self-supervised learning can be explored, taking advantage of the inherent structure of the data for learning. In this context, considering the outlined challenges, this study developed and experimentally evaluated self-supervised models applied to histopathological image classification and retrieval tasks. A comparative analysis was conducted using both pre-trained feature extractors trained on natural images—ResNet50, DINOv2 (Small and Base)—and state-of-the-art extractors specifically designed for histopathology, namely RetCCL and Phikon ViT-B. Furthermore, given the promising results obtained by the DINOv2 Small and Base models, fine-tuning was performed, leading to the development of DINOHist-S and DINOHist-B models to assess their effectiveness on histopathological images. To evaluate the potential of the models, extensive experiments were carried out with seven databases. The results in the lesion classification task demonstrated that, by fine-tuning the DINOv2 models with few iterations, it is possible to match or surpass the state-of-the-art extractors in the histopathological domain. For some scenarios, the evaluated models achieved efficiency above 99% for multiple measurements. Furthermore, it was observed that the state-of-the-art models were significantly inferior in kidney lesion databases, while fine-tuning allowed better results, even for small datasets. In the content-based image retrieval task, the results demonstrated that Phikon ViT-B outperformed the other models at multiple ranking levels, achieving MAP values above 90% in most databases and in all rankings evaluated. For most datasets, Phikon ViT-B was statistically superior, with the exception of the kidney lesion databases. For these, DINOHist-S and DINOHist-B presented the best results. In general, the results demonstrated that the adjustment of base models in specific domains can outperform specialized models trained on a large scale, requiring only a fraction of the resources and training time compared to state-of-the-art models. To extend this work, future research can be conducted, including the evaluation of new architectures, the use of other datasets and the analysis of different fine-tuning strategies, among others.

Keywords: Histopathology, Lesions, Self-Supervised Learning, Classification, Content-Based Image Retrieval.

Resumo

A histopatologia analisa tecidos e lesões para diagnóstico de doenças, mas essa análise manual costuma ser trabalhosa e sujeita a erros. Para melhorar a acurácia e reduzir a carga dos patologistas, sistemas de diagnóstico assistido por computador podem ser usados, incluindo classificadores de doenças e ferramentas de recuperação de imagens baseada em conteúdo. No entanto, para serem utilizados, esses sistemas necessitam extrair *features* das imagens, mas a rotulação dessas imagens também exige um grande esforço, sendo demorada e onerosa. Para atenuar estes problemas, abordagens centradas em aprendizado auto-supervisionado podem ser exploradas, aproveitando a estrutura dos próprios dados para aprendizado. Neste contexto, levando em consideração os desafios apresentados, este trabalho desenvolveu e avaliou experimentalmente modelos auto-supervisionados, aplicados em tarefas de classificação e recuperação de imagens histopatológicas. Para isso, foi conduzida uma análise comparativa, incluindo extratores pré-treinados com imagens naturais: ResNet50, DINOv2 (Small e Base), e extratores do estado da arte específicos de histopatologia, RetCCL e Phikon ViT-B. Além disso, considerando os resultados promissores que os modelos DINOv2 Small e Base têm alcançado, foi realizado o *fine-tuning*, dando origem aos modelos DINOHist-S e DINOHist-B, para avaliar seu comportamento com imagens histopatológicas. Para avaliar o potencial dos modelos foram realizados amplos experimentos com sete bases de dados. Os resultados na tarefa de classificação de lesões demonstraram que, ao ajustar os modelos DINOv2 com poucas iterações, pode-se igualar ou superar os extratores de última geração do domínio histopatológico. Para alguns cenários, os modelos avaliados alcançaram eficácia acima de 99% para múltiplas medidas. Além disso, observou-se que os modelos de estado da arte foram significativamente inferiores em bases de dados de lesões renais, enquanto o *fine-tuning* permitiu melhores resultados, mesmo para conjuntos de dados pequenos. Na tarefa de recuperação de imagens por conteúdo, os resultados demonstraram que o Phikon ViT-B superou os demais modelos em múltiplos níveis de *ranking*, alcançando valores de MAP superiores a 90% na maioria das bases de dados e em todos os rankings avaliados. Para a maioria dos *datasets*, o Phikon ViT-B foi estatisticamente superior, com exceção das bases de lesões renais. Para estas, o DINOHist-S e o DINOHist-B apresentaram os melhores resultados. Em geral, os resultados demonstraram que o ajuste de modelos base em domínios específicos podem superar modelos especializados treinados em larga escala, demandando apenas uma fração dos recursos e do tempo de treinamento em comparação a modelos de estado da arte.

Para estender este trabalho, pesquisas futuras podem ser conduzidas, incluindo a avaliação de novas arquiteturas, a utilização de outros conjuntos de dados, a análise de diferentes estratégias de *fine-tuning*, entre outras.

Palavras-chave: Histopatologia, Lesões, Aprendizado Auto-Supervisionado, Classificação, Recuperação de Imagens por conteúdo.

Prefácio

Esta dissertação de mestrado foi submetida à Universidade Estadual de Feira de Santana (UEFS) como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

A dissertação foi desenvolvida no Programa de Pós-Graduação em Ciência da Computação (PGCC), tendo como orientador o Prof. Dr. **Rodrigo Tripodi Calumby**.

Esta pesquisa foi financiada pela CAPES.

Agradecimentos

Primeiramente, agradeço a Deus por me conceder força, sabedoria e perseverança ao longo desta jornada acadêmica. Sua presença constante em minha vida me guiou nos momentos de incerteza e me deu coragem para superar os desafios e tribulações. ELE me deu coragem nos momentos mais difíceis, quando a ansiedade me paralisava e não me deixava continuar.

Agradeço, também, à Virgem Maria, por sua intercessão e proteção, sempre presente em meus pensamentos e orações. Seu exemplo de fé e humildade foi uma fonte de inspiração ao longo desta caminhada. A certeza de sua intercessão me deu força e coragem para trilhar essa jornada.

Gostaria de expressar minha profunda gratidão ao meu orientador e amigo amado, Rodrigo Tripodi Calumby, por sua orientação, paciência e por acreditar em meu potencial. Suas contribuições foram essenciais para o desenvolvimento deste trabalho.

Aos meus colegas e amigos do laboratório ADAM, agradeço pelo apoio, pelas trocas de conhecimento e pela amizade ao longo dos anos. O ambiente colaborativo foi essencial para o meu crescimento pessoal e profissional.

Aos meus pais, José Mota Figuerêdo e Lecy Lima Figuerêdo, que me apoiaram em todos os momentos, oferecendo amor, compreensão e incentivo. Vocês são a base de tudo o que conquistei até aqui.

Aos meus irmãos, Leciane Lima Figuerêdo e Jusceny Lima Figuerêdo, sou profundamente grato pelo apoio e pelas palavras de encorajamento. Mesmo à distância, vocês nunca deixaram de estar ao meu lado, oferecendo-me força e auxílio inestimáveis ao longo desta jornada.

A Jorge Luis Cazumbá da Silva, que esteve ao meu lado nesta jornada. Seu amor, paciência e apoio incondicional foram fundamentais para que eu pudesse alcançar este objetivo. Nos momentos de dificuldade, suas palavras de encorajamento e seu carinho me deram a força necessária para seguir em frente.

Aos meus amigos, que estiveram ao meu lado, oferecendo palavras de encorajamento e momentos de alegria, meu sincero obrigado.

Agradeço também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Ministério da Educação (CAPES) pelo suporte financeiro e pelas oportunidades oferecidas, que viabilizaram a realização deste estudo.

Agradeço também ao projeto Pathospotter, pelo fornecimento das bases de dados de lesões renais, essenciais para a condução deste estudo.

Por fim, dedico esta dissertação a Deus, à Virgem Maria e a todos aqueles que, de alguma forma, contribuíram para a concretização deste sonho.

"A verdadeira viagem da descoberta consiste não em procurar novas paisagens, mas em ter novos olhos."
– Marcel Proust

Sumário

Alinhamento com a Linha de Pesquisa

Linha de Pesquisa: Computação Inteligente

A presente dissertação está alinhada com a linha de pesquisa ao desenvolver e avaliar metodologias computacionais inteligentes voltadas para a solução de problemas complexos na área de patologia digital. Ao aplicar técnicas de aprendizado auto-supervisionado, a pesquisa busca aprimorar modelos existentes e propor novas abordagens para a classificação e recuperação de imagens histopatológicas. Assim, o trabalho contribui diretamente para os avanços em áreas como Inteligência Artificial, Aprendizado de Máquina, Ciência de Dados, Recuperação da Informação, e Processamento de Imagens, abordando desafios relacionados à extração de conhecimento e reconhecimento de padrões em bases de dados complexas.

Produções Bibliográficas, Produções Técnicas e Premiações

Nesta capítulo são apresentadas produções bibliográficas, produções técnicas e premiações alcançadas ao longo do Mestrado, relacionadas à linha de pesquisa desta dissertação. Ressalta-se que foi considerando o período enquanto estudante especial e regular.

0.1 Produções Bibliográficas

Figuerêdo, J. S. L., Maia, A. L. L. M., & Calumby, R. T. (2023, September). A Novel Graph-based Diversity-aware Rank Fusion Method Applied to Image Metasearch. In Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados (pp. 324-329). SBC.

Figuerêdo, J. S. L., Araujo-Calumby, R. F., & Calumby, R. T. (2023, September). Towards Effective and Reliable Data-driven Prognostication: An Application to COVID-19. In Anais do XI Symposium on Knowledge Discovery, Mining and Learning (pp. 81-88). SBC.

Figuerêdo, J. S. L., Ferreira, M. E. D. C., & Calumby, R. T. (2023, November). Machine Learning for Soil Attribute Prediction: An Effectiveness and Dimensionality Reduction Analysis. In Anais do XIV Congresso Brasileiro de Agroinformática (pp. 302-309). SBC.

Figuerêdo, J. S. L., Sarinho, V. T., & Calumby, R. T. (2021, October). Low-Cost Machine Learning for Effective and Efficient Bad Smells Detection. In Anais do IX Symposium on Knowledge Discovery, Mining and Learning (pp. 113-120). SBC.

Figuerêdo, J. S. L., Maia, A. L. L. M., & Calumby, R. T. (2024, August). GDRF: An Innovative Graph-Based Rank Fusion Method for Enhancing Diversity in Image Metasearch. Journal of Information and Data Management (JIDM).

0.2 Artigos Aceitos para Publicação

Figuerêdo, J. S. L., Araujo-Calumby, R. F., & Calumby, R. T. (2024, November). Enhancing COVID-19 Prognosis Prediction with Machine Learning and LIME Ex-

planation. Journal of Information and Data Management (JIDM).

0.3 Premiações

Menção honrosa com o trabalho A Novel Graph-based Diversity-aware Rank Fusion Method Applied to Image Metasearch, 38º Simpósio Brasileiro de Bancos de Dados (SBBD).

Lista de Tabelas

Lista de Figuras

Lista de Abreviações

Abreviação	**Descrição**
CBIR	Recuperação de Imagens Baseada em Conteúdo (Content-Based Image Retrieval)
WSI	Imagens de Slides Inteiros (Whole Slide Imaging)
GPU	Unidade de Processamento Gráfico (Graphics Processing Unit)
IA	Inteligência Artificial
FDA	Administração de Alimentos e Medicamentos (Food and Drug Administration)
TCGA	Atlas do Genoma do Câncer (The Cancer Genome Atlas)
BoB	Um monte de códigos de barras (Bunch of Barcodes)
MIM	Modelagem de Imagem Mascarada (Masked Image Modeling)
PAMS	Picroanilina Metenamina Prata (Picroaniline Methenamine Silver)
PS	Picro-Sirius Vermelho (Picro-Sirius Red)
AZAN	Azul de Alcian e Naranja (Azul de Alcian e Naranja)
PICRO	Picro-Crômico de Weigert (Picro-Crômico de Weigert)
KMC	Faculdade de Medicina de Kasturba (Kasturba Medical College)
LORA	Adaptação de Baixa Ordem (Low-Rank Adaptation)
VIT	Transformador de Visão (Vision Transformer)
MAP	Média da Precision Média (Mean Average Precision)
FIOCRUZ	Fundação Oswaldo Cruz
ML	Machine Learning
PAIP	Pathology AI Platform

Lista de Símbolos

Símbolo	Descrição
$\sum$	Somatório
θ	Letra grega theta

Capítulo 1

Introdução

"O progresso é impossível sem mudança; e aqueles que não conseguem mudar suas mentes não podem mudar nada."

– George Bernard Shaw

A histopatologia dedica-se à investigação e análise de células biológicas e estruturas de tecidos, apoiada por um microscópio (Orchard e Nation, 2012). Este procedimento permite que o patologista realize o diagnóstico examinando um pequeno pedaço de tecido da pele, fígado, rim ou outro órgão. A análise microscópica é amplamente utilizada para diagnosticar doenças como o câncer, doenças hepáticas, doenças renais, entre outras (Anwar et al., 2018; Erfankhah et al., 2019; Abels et al., 2019; Chagas et al., 2020; L'Imperio et al., 2021). De modo geral, a histopatologia desempenha um papel relevante no diagnóstico, planejamento de tratamento e pesquisa médica. Em um dia de trabalho comum, os patologistas analisam lâminas histológicas visualmente para identificar anormalidades celulares, padrões teciduais e marcadores de doenças (Filiot et al., 2023). Embora seja uma prática comum, essa avaliação manual é trabalhosa, demorada, subjetiva e propensa a erros. Diante disso, para reduzir a carga de trabalho desses profissionais, bem como melhorar a objetividade da análise dessas imagens, sistemas de diagnóstico auxiliado por computador(do inglês, Computer-Aided Diagnosis (CAD)) oferecem uma promissora aplicação (Shi et al., 2018).

Em muitas aplicações médicas baseadas em computação inteligente, imagens coletadas são utilizadas para gerar modelos de detecção de doenças, deste modo, quando novas imagens são adquiridas, o patologista utiliza esse sistema para ajudá-lo no diagnóstico. Outra situação recorrente em que o patologista se aproveita de ferramentas computacionais, é na busca de imagens similares à que está sob análise. O uso de sistemas de recuperação de imagens baseada em conteúdo (do inglês, *Content-*

Based Image Retrieval (CBIR)) auxilia os patologistas a encontrar em grandes bases de dados casos similares ao que está sob avaliação. Com isso, é possível consultar, e.g., o diagnóstico realizado para casos semelhantes. Ambas as abordagens têm o potencial de modernizar e melhorar o processo de tomada de decisão clínica desses profissionais.

A crescente disponibilidade de tecnologias que permitem a criação rotineira de imagens de slides inteiros (do inglês, Whole Slide Imaging (WSIs[1])) de alta resolução revolucionou a patologia computacional/digital[2] (Navid Farahani et al., 2015). A Figura 1.1 apresenta um *workflow* comumente usado em estudos relacionados à patologia computacional. Em geral, o fluxo de trabalho compreende seis etapas: (a) Seleção de lâminas; (b) Digitalização de lâminas; (c) Definição do problema, rotulagem/anotação do *ground-truth*; (d) Desenvolvimento de modelos de Inteligência Artificial(IA)/*Machine Learning* (ML); (e) Implantação e, por fim, (f) Avaliação.

Com o avanço da patologia digital, surgem desafios significativos, especialmente devido à geração de grandes volumes de dados provenientes da digitalização de amostras patológicas. A busca manual nessas bases de dados se torna inviável, comprometendo a eficiência e a precisão dos diagnósticos. Para enfrentar esse problema, é necessário adotar soluções tecnológicas que otimizem o processo de análise. Nesse contexto, a combinação de sistemas de CBIR com IA surge como uma alternativa promissora.

CBIR é uma tecnologia utilizada para buscar imagens semelhantes com base em seu conteúdo visual. As imagens são representadas por *features*, em geral de alta dimensionalidade, que são utilizadas para calcular a similaridade entre uma imagem de referência (também chamada de imagem de consulta) e as imagens armazenadas em um banco de dados (Gudivada e Raghavan, 1995). Essa abordagem tem sido desenvolvida para diferentes domínios de aplicação e para imagens médicas vem sendo amplamente explorada com diversas fontes de imagem, como radiografia, dermatologia, mamografia, ressonância magnética, tomografia computadorizada e histopatologia (Barata e Santiago, 2021; Mohammad Alizadeh et al., 2023; Wickstrøm et al., 2023b). Os sistemas de CBIR exploram o conhecimento baseado em evidências de casos anteriores e disponibilizam os resultados (geralmente um *ranking* com as imagens mais similares à imagem de consulta) aos patologistas para uma tomada de decisão mais eficiente e fundamentada (Komura e Ishikawa, 2018). Ou seja, esses sistemas apoiam os médicos na recuperação de imagens relevantes de grandes bases de dados a partir da comparação com uma imagem de referência, contornando a ineficiente busca manual.

Nos últimos anos, vários sistemas CBIR baseados em *deep learning* foram desenvolvidos, alcançando resultados notáveis (Kalra et al., 2020; Haq et al., 2021; Souid

[1]Refere-se à digitalização de lâminas de vidro convencionais para produzir lâminas digitais.

[2]Ressalta-se que patologia computacional envolve a aplicação de métodos computacionais e estatísticos para a análise de dados biológicos e médicos, enquanto a patologia digital está mais focada na digitalização e armazenamento dessas amostras para facilitar o diagnóstico, a comunicação e a colaboração

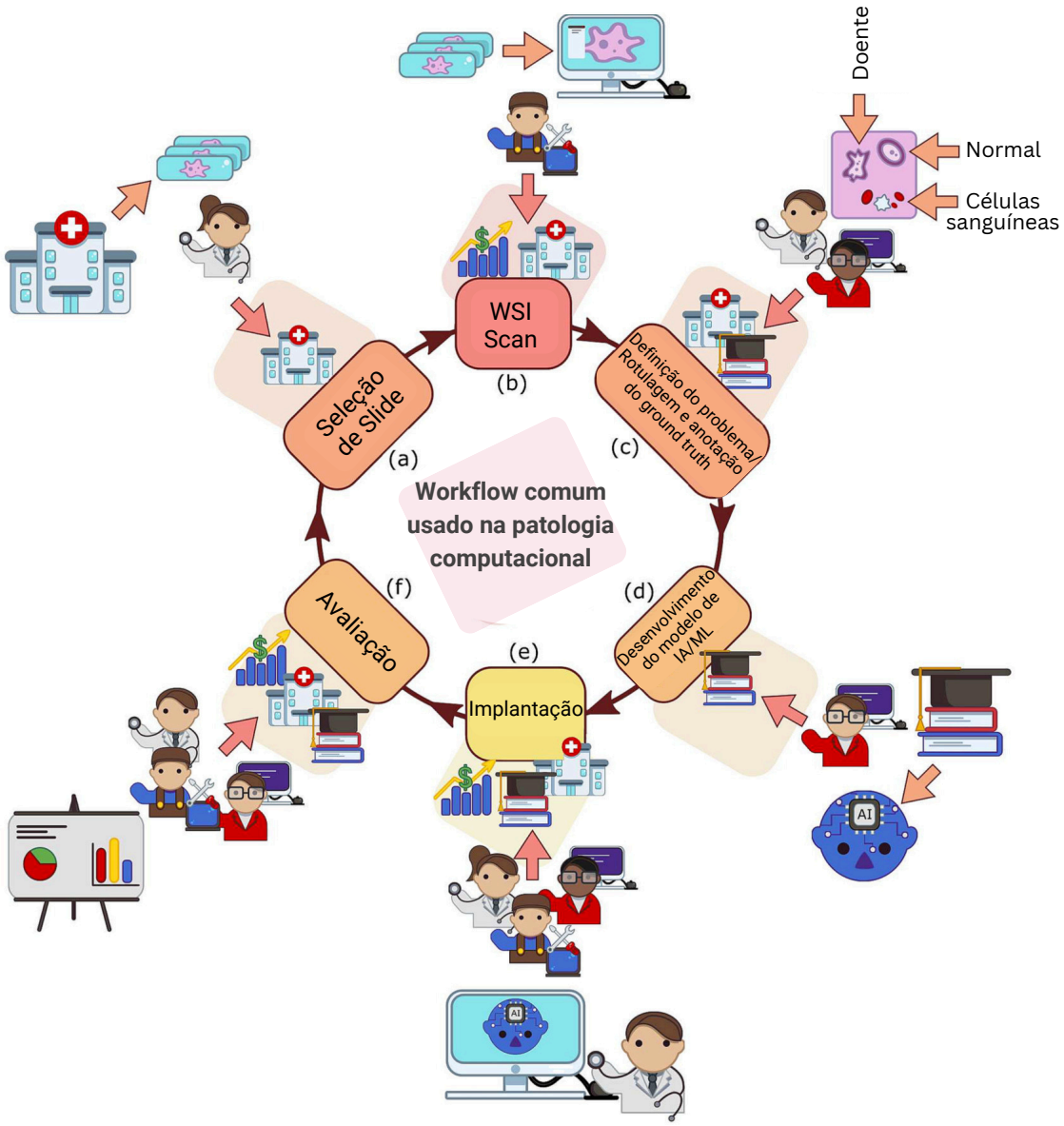


Figura 1.1: Workflow comumente usado em estudos da patologia computacional/digital. Adaptado de Hosseini et al. (2024)

et al., 2023). Para gerar os modelos de extração de *features* usado por esses sistemas, os modelos baseados em *deep learning* dependem, comumente, de dados em larga escala rotulados para o treinamento (Yoshinobu et al., 2020; Jing e Tian, 2021). Contudo, rotular grandes quantidades de dados pode ser desafiador, especialmente tratando-se de dados médicos. Anotar a WSI, por exemplo, no nível do slide ou do pixel, pode ser demorado e complexo, mesmo para patologistas treinados, o que limita o desenvolvimento de CBIR baseados em *deep learning* (Wang et al., 2020). Além disso, dependendo do tipo de patologia, o processo de rotulação torna-se ainda mais desafiador, estando sujeito, inclusive, à variabilidade entre observadores (Kang

et al., 2023a). Por exemplo, na patologia renal, todos os glomérulos devem ser cuidadosamente identificados, tarefa geralmente demorada e sujeita a erros (Yamaguchi et al., 2021). Ou ainda, quando considera-se a análise de câncer, a anotação torna-se ainda mais difícil, dada a diversidade de tipo e protocolos de preparação de tecidos, sobretudo devido à variabilidade de cor, textura, coloração e morfologia celulares.

Para minimizar o problema da rotulação dos dados, vários estudos na área de patologia computacional têm empregado estratégias de *transfer-learning* a partir do ImageNet (Kalra et al., 2020; Ozen et al., 2021; Majumdar et al., 2023). Geralmente, são utilizadas redes neurais convolucionais (CNN), visando o aprendizado de representações visuais robustas por meio de imagens naturais, que servem de extratores de *features* para imagens histopatológicas, e que podem ser posteriormente utilizadas pelos sistemas de CBIR. Outra estratégia mais recente que tem sido explorada para enfrentar essa limitação é o aprendizado auto-supervisionado.

O aprendizado auto-supervisionado é um paradigma que permite a aprendizagem de *features* semânticas através da geração de sinais de supervisão a partir de um conjunto de dados não rotulados (Chen et al., 2019). Em seguida, o modelo pré-treinado, pode ser usado em tarefas subsequentes onde a quantidade de dados anotados é limitada, sendo ajustado ao novo domínio via *fine-tuning*. Tipicamente, o *fine-tuning* é conduzido a partir de uma base de dados rotulada. No entanto, esse processo também pode ser realizado de forma auto-supervisionada. O uso dessa metodologia torna-se ainda mais relevante, pois, em cenários reais, os dados nem sempre estarão rotulados (Gui et al., 2024). Embora a sua rotulação seja uma solução, este processo exige um esforço significativo de especialistas, que nem sempre estão disponíveis ou cuja atuação pode ser inviável economicamente para determinados projetos.

Considerando a importância do aprendizado auto-supervisionado, várias técnicas/modelos foram desenvolvidos nos últimos anos, destacando-se o SwAV (Caron et al., 2020), MoCo v3 (Chen et al., 2021), iBOT (Zhou et al., 2022), EsVIT (Li et al., 2022). Estes métodos são utilizados em diversas aplicações, como classificação de doenças, detecção de objetos, segmentação semântica, reconhecimento da ação humana, estimativa de profundidade, entre outras (Jing e Tian, 2021; Chen et al., 2022; Filiot et al., 2023).

Geralmente, as abordagens auto-supervisionadas são utilizadas no processo de pré-treino de modelos em um conjunto de dados em larga escala, frequentemente na *ImageNet*. Em seguida, são treinados de modo supervisionado considerando uma tarefa alvo, frequentemente em um conjunto reduzido de dados. Embora essa seja uma prática comum, confiar apenas no pré-treinamento fora do domínio pode acarretar várias limitações (Kataria et al., 2023), dada a particularidade intrínseca de cada domínio, especialmente no contexto médico. Diferentemente de imagens naturais, imagens histológicas exibem características complexas e específicas, incluindo estruturas celulares, morfologia dos tecidos e padrões de coloração, que podem não ser capturadas adequadamente por modelos pré-treinados a partir de imagens de domínios mais generalistas, baseados em categorias de objetos ou imagens da natureza.

Considerando estes desafios, foram propostos modelos de base treinados em dados de imagens médicas, mas não extensivamente em dados histopatológicos (Azizi et al., 2023; Moor et al., 2023; Wickstrøm et al., 2023b).

Especificamente, alguns estudos foram conduzidos com foco na exploração de imagens histopatológicas, destacando-se os trabalhos realizados por Wang et al. (2023) e Filiot et al. (2023). Em Wang et al. (2023) foi proposto o retCCL (do inglês, Retrieval with Clustering-guided Contrastive Learning). Trata-se de *framework* de recuperação de imagens histopatológicas, aplicável tanto para recuperação em nível de WSI quanto em nível de *patch*[3]. O método desenvolvido utilizou dados em larga escala de imagens não rotuladas (15 milhões), oriundas do TCGA[4] (do inglês, The Cancer Genome Atlas) e PAIP[5] (do inglês, pathology AI platform), o que contribuiu para aprender características universais. Isto possibilitou, conforme indicado pelos autores, utilizar o modelo diretamente para tarefas subsequentes sem a aplicação de *fine-tuning* extra.

Ainda no contexto de imagens histopatológicas, Filiot et al. (2023) desenvolveram modelos pré-treinados de modo auto-supervisionado em um conjunto com cerca de 43 milhões de imagens, originárias do TCGA, a partir da arquitetura iBOT. Estes modelos foram avaliados em várias tarefas de classificação, incluindo predição do subtipo histológico, predição de subtipo molecular, identificação do tipo de câncer, predição de alterações genômicas e detecção de metástases, alcançando resultados superiores a outros métodos. No entanto, apesar dos resultados promissores, os modelos desenvolvidos por Filiot et al. (2023) não foram avaliados em tarefas de CBIR, sendo desconhecida a sua eficácia nesse contexto. Além disso, tanto para o retCCL quanto para os modelos de Filiot et al. (2023), ainda é necessário avaliá-los em alguns domínios histológicos, como o nefropatológico. Considerando as características peculiares das imagens renais, onde uma avaliação cuidadosa dos glomérulos deve ser executada para identificar alguma lesão ou patologia associada, esses modelos podem ser úteis para prover uma análise mais acurada, aproveitando-se da sua natureza auto-supervisionada.

Apesar dos avanços recentes do aprendizado auto-supervisionado, o treinamento de redes inicializadas com pesos aleatórios nem sempre é viável, uma vez que o processo de pré-treino geralmente demanda recursos computacionais significativos. O treinamento de métodos auto-supervisionados pode exigir muitos recursos computacionais, pois, para aprender representações sem a rotulação explícita, é essencial processar grandes volumes de dados, além disso, é importante que se faça a otimização dos hiperparâmetros. Consequentemente, isso pode exigir infraestruturas mais robustas, como unidades de processamento gráfico (do inglês, *Graphics Processing Unit*(GPU)) de alto desempenho, aumentando o custo geral do desenvolvimento. Por exemplo, considerando o *retCCL* e o principal modelo desenvolvido em Filiot

[3]Refere-se a uma subseção menor, retangular ou quadrada da WSI.

[4]`https://portal.gdc.cancer.gov/`

[5]`http://www.wisepaip.org/paip`

et al. (2023), o processo de pré-treino utilizou 32 GPUs NVIDIA V100 de 32Gb por 300 horas e 16 a 64 GPUs NVIDIA V100 de 32Gb por 1216 horas, respectivamente. Para contornar essa limitação, os modelos de base (do inglês, *Foundation Models*), surgem como uma alternativa. Recentemente, um desses modelos, denominado como DINOv2, despertou o interesse da comunidade científica (Oquab et al., 2023). O DINOv2 é um modelo base de código aberto pré-treinado com aprendizagem auto-supervisionada e têm alcançado resultados de estado da arte em várias tarefas. O DINOv2 tem superado vários métodos alternativos em uma ampla gama de *benchmarks* (Oquab et al., 2023). No entanto, apesar de alcançar resultados de estado da arte, ainda persistem questões relativas à adaptabilidade do DINOv2 para imagens histopatológicas, especialmente imagens renais. Além disso, sua aplicação em tarefas de CBIR requer investigações mais aprofundadas. Adicionalmente, é essencial investigar a sua adaptação como extrator de características para bases de dados pequenas não rotuladas e aplicado em patologias ou lesões especificas, especialmente em cenários cujos recursos computacionais são limitados.

A partir do contexto e dos desafios acima mencionados, foram definidas as seguintes questões de pesquisa:

- **Q1:** Considerando modelos auto-supervisionados do estado da arte pré-treinados com imagens histopatológicas e com imagens naturais, quais seus níveis e diferença em termos de eficácia no contexto de classificação de doenças histopatológicas?
- **Q2:** Considerando modelos auto-supervisionados do estado da arte pré-treinados com imagens histopatológicas e com imagens naturais, qual seu nível de eficácia no contexto de recuperação de imagens histopatológicas e quais demonstram melhor adaptação para cada um dos sub-domínios da histopatologia considerados neste estudo?
- **Q3:** Em que medida modelos pré-treinados em imagens naturais, ajustados auto-supervisionadamente com poucos dados histopatológicos, impactam a eficácia na classificação e recuperação em comparação com extratores de estado da arte?
- **Q4:** Como se comportam os extratores de características pré-treinados com imagens naturais quando aplicados ao domínio histopatológico?
- **Q5:** Qual o impacto do *fine-tuning* em modelos auto-supervisionados no aprimoramento da representação de dados e na melhoria da eficácia em tarefas específicas, como classificação e recuperação de imagens?

1.1 Objetivos

1.1.1 Objetivo geral

Diante dos desafios apresentados, este estudo tem como objetivo geral projetar, desenvolver e avaliar experimentalmente modelos auto-supervisionados, para classificação de lesões e a recuperação de imagens baseada em conteúdo aplicada à histopatologia.

1.1.2 Objetivos específicos

Especificamente, este projeto visa:

- Comparar experimentalmente a eficácia de modelos auto-supervisionados, considerando os cenários com e sem *fine-tuning*, no contexto da classificação e recuperação de imagens histopatológicas;
- Comparar experimentalmente a eficácia de modelos auto-supervisionados ajustados em bases de dados reduzidas com a de extratores de *features* especializados e reconhecidos como estado da arte no domínio;
- Comparar experimentalmente modelos auto-supervisionados pré-treinados com imagens naturais e modelos pré-treinados de modo supervisionado no mesmo domínio;
- Avaliar experimentalmente a eficácia de modelos auto-supervisionados ajustados em bases de dados com poucas imagens no domínio histopatológico, abrangendo diferentes tipos de patologias e lesões;

1.2 Contribuições

As principais contribuições deste trabalho são:

- Avaliação do impacto do *fine-tuning* de modelos auto-supervisionados ajustados em base de dados com poucas imagens.
- Avaliação dos modelos ajustados por *fine-tuning* em relação aos modelos base não ajustados, destacando impactos na recuperação de imagens histopatológicas e fornecendo informações sobre a necessidade de *fine-tuning*;
- Avaliação da eficácia e qualidade das representações geradas pelos modelos, oferecendo *insights* sobre a eficácia das representações vetoriais na recuperação e análise de imagens histopatológicas.
- Disponibilização online de todos os modelos desenvolvidos neste estudo, de modo a incentivar que outros pesquisadores expandam a avaliação conduzida neste trabalho.

- Análise dos achados e discussões sobre possibilidades de avanços em termos de avaliação de novas arquiteturas, utilização de outros conjuntos de dados, análise da influência dos hiperparâmetros, avaliação de diferentes estratégias *fine-tuning*, avaliação do impacto da redução da dimensionalidade e experimentação com múltiplos classificadores.
- Verificação se o *fine-tuning* de modelos de base em dados específicos da tarefa pode superar modelos especializados de domínio treinados em larga escala, demandando apenas uma fração dos recursos e do tempo de treinamento para produção de resultados similares ou até mesmo superiores.

1.3 Organização do Trabalho

O restante deste trabalho segue estruturado da seguinte forma. No Capítulo 2, apresentamos uma revisão bibliográfica, incluindo também os principais trabalhos relacionados. Capítulo 3 descreve os procedimentos metodológicos utilizados na condução deste estudo, incluindo a descrição dos experimentos realizados. Em seguida, o Capítulo 4 apresenta e discute os resultados alcançados. Por fim, no Capítulo 5 discutimos as conclusões obtidas a partir deste trabalho e possíveis direções para pesquisas futuras.

Capítulo 2

Revisão Bibliográfica

"O maior inimigo do conhecimento não é a ignorância, é a ilusão do conhecimento."
– Stephen Hawking

Este capítulo descreve os principais conceitos e trabalhos relacionados a este estudo. Na seção 2.1 apresentamos uma visão geral sobre patologia computacional. A seção 2.2 descreve o mecanismo de recuperação de imagens baseada em conteúdo. Em seguida, a seção 2.3 apresenta os princípios relacionados a aprendizagem auto-supervisionada. A seção 2.4 aborda algumas medidas que são utilizadas para avaliar a eficácia de sistemas de classificação e recuperação de imagens, considerando um cenário de múltiplas classes. Por fim, a seção 2.5 descreve trabalhos relacionados à recuperação de imagens baseada em conteúdo no contexto de imagens histopatológicas.

2.1 Patologia Computacional

A patologia computacional refere-se a um campo interdisciplinar, que combina a ciência da computação e IA com a patologia tradicional, objetivando aprimorar o diagnóstico e a análise de doenças, especialmente no contexto da patologia digital (Abels et al., 2019; Hosseini et al., 2024). Em outra perspectiva, a patologia computacional desenvolve infraestrutura e fluxos de trabalho de diagnósticos digitais como um sistema CAD assistivo para patologia clínica, facilitando mudanças transformacionais no diagnóstico e tratamento de doenças, principalmente do câncer (Hosseini et al., 2024).

Nas últimas décadas, a análise de imagens médicas foi facilitada, sobretudo em razão de avanços em tecnologias de captura e armazenamento de dados, impulsionando

o crescimento da patologia digital. Recentemente, assistiu-se a uma explosão de técnicas de imagem digital, no qual enormes quantidades de imagens médicas foram produzidas com qualidade e diversidade cada vez maiores (Hosseini et al., 2024). Junto a este avanço surgiram vários desafios, a exemplo de como analisar essas bases em larga escala, tendo em vista que métodos convencionais tem sucesso limitado neste tipo de análise, pois não são capazes de lidar com a enorme quantidade de dados de imagem.

Considerando as limitações de métodos convencionais, ao longo dos anos foram desenvolvidas abordagens para análise de imagens médicas em larga escala, que são baseadas principalmente em avanços recentes na visão computacional, aprendizado de máquina e recuperação de informação, culminando com avanços na patologia computacional. A patologia computacional possibilita muitos benefícios para a área médica, ao automatizar e otimizar a análise de lâminas histológicas digitalizadas, permitindo a extração de características quantitativas e a identificação de padrões que auxiliam no diagnóstico médico. Isso é particularmente útil em áreas como oncologia, onde a detecção e classificação de tumores podem ser melhoradas com técnicas computacionais (Hosseini et al., 2024). Apesar dos muitos benefícios oferecidos para melhorar a eficiência e a precisão na patologia, seu uso prático ainda é limitado (Abels et al., 2019; Kumar et al., 2020; Koohbanani et al., 2021). Essa falta de adoção e integração na prática clínica ressalta a necessidade de intensificar estudos nesta área, de modo a evidenciar cada vez mais a sua utilidade em tais cenários.

A patologia computacional é aplicada em várias tarefas da área médica, como por exemplo, detecção de doenças, classificação de tecidos, diagnóstico de doença, segmentação e tarefas de prognóstico (Hosseini et al., 2024). A Figura 2.1 exemplifica essas tarefas visualmente. Além dessas aplicações, a patologia computacional também tem explorado aplicações relacionadas a recuperação de imagens baseada em conteúdo. Esta recuperação é conduzida ao indexar e minerar imagens que contêm um conteúdo visual semelhante (por exemplo, forma, morfologia, estrutura, etc.). Para que uma nova imagem médica seja analisada, um sistema CBIR pode inicialmente recuperar imagens visualmente semelhantes em um conjunto de dados existente. Em seguida, suas descrições e interpretações de alto nível podem ser exploradas com base nas imagens recuperadas. Mais detalhes desse tipo de sistema é abordado na seção 2.2.

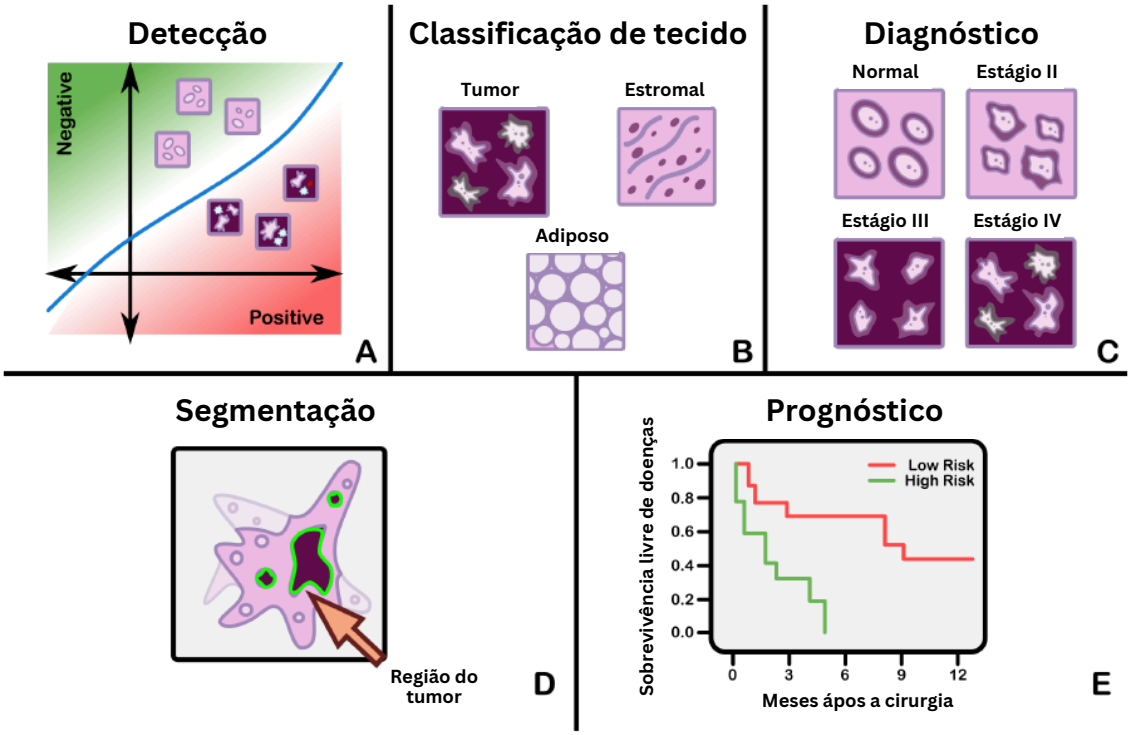


Figura 2.1: Exemplos de tarefas englobadas pela patologia computacional. A) Detecção: tarefa de detecção comum, como diferenciar classes positivas de negativas, como malignas de benignas, B) Classificação de subtipo de tecido: tarefa de classificação para tecido tumoral, estroma e tecido adiposo, C) Diagnóstico de doença: tarefa comum de diagnóstico de doença, como estágio do câncer, D) Segmentação: segmentação de tumor em imagens digitais e E) Tarefas de prognóstico: mostra um gráfico comparando a taxa de sobrevivência e os meses após a cirurgia. Adaptado de (Hosseini et al., 2024)

Em geral, antes de ser analisado, seja na análise microscópica ou utilizando técnicas computacionais, o tecido extraído passa por um processo de preparação. Um desses processos corresponde na utilização de contraste ou corante para melhorar a visualização de estruturas específicas. Existem vários tipos de corantes, cada um com uma aplicação particular. Por exemplo, tem-se o Ácido Periódico de Schiff (do inglês, *periodic acid–Schiff* (PAS)), a Hematoxilina e Eosina (H&E), PAMS (Picroaniline Methenamine Silver), PS (Picro-Sirius Red), AZAN (Azul de Alcian e Naranja) e PICRO (Picro-Crômico de Weigert). Ressalta-se que o tipo de corante utilizado pode interferir no processo de análise dessas imagens, uma vez que determinadas estruturas celulares podem ser suprimidas ou realçadas.

2.2 Recuperação de Imagens Baseada em Conteúdo

A recuperação de imagens representa uma área central de estudo, com inúmeras aplicações práticas, tais como: diagnóstico por imagem, identificação de objetos ou pessoas, análise forense, personalização de recomendação, monitoramento ambiente, entre outras (Li et al., 2021; Hameed et al., 2021). A recuperação das imagens é realizada a partir de uma consulta de entrada. A consulta pode ser realizada a partir de diferentes formatos de entrada. Por exemplo, em uma abordagem tradicional de recuperação de imagens, explora-se passagens textuais associadas a elas. Os fragmentos textuais podem ser oriundos de várias fontes, por exemplo, páginas web, palavras-chave, *tags*, legendas, artigos e descrições feitas pelos usuários. Assim, técnicas tradicionais de recuperação textual podem ser aplicadas, mediante modelagem vetorial, probabilística ou até mesmo modelagem de linguagem.

Outra abordagem que pode ser utilizada é a busca baseada em conteúdo. Essa estratégia é centrada na noção de similaridade entre as imagens, isto é, dado um banco de dados com um grande número de imagens, o usuário deseja recuperar as imagens mais similares a um padrão de consulta (normalmente uma imagem). Para conduzir esta busca, as abordagens de CBIR descrevem o conteúdo de uma imagem geralmente como um vetor de números reais, conhecido como vetor de características. Deste modo, se um vetor de características abrange as propriedades visuais descritivas de uma imagem, então a busca por imagens semelhantes torna-se um problema de correspondência de vizinhos mais próximos (Kalra et al., 2020). Assim, imagens com conteúdo semelhante poderiam ser recuperadas com base na comparação de seus vetores de características e não com base nos metadados textuais associados. Diversos domínios podem se beneficiar desses sistemas para melhorar a tomada de decisões. Por exemplo, um médico pode querer investigar como pacientes com uma doença semelhante à de um novo paciente, como metástase hepática, foram diagnosticados em um grande banco de dados. As informações dos diagnósticos anteriores podem então ser usadas para determinar o tratamento adequado para o novo paciente.

A Figura 2.2 apresenta a arquitetura típica de um sistema de recuperação de imagens baseada em conteúdo, que é dividida em três módulos: Interface, Processamento de Consultas e Base de dados. Na interface o usuário especifica uma consulta a partir de uma imagem de interesse, e visualiza as imagens recuperadas. O módulo de processamento de consultas extrai o vetor de características da imagem submetida, e em seguida aplica uma medida de distância para verificar a similaridade entre a imagem de consulta e as imagens da base. Ainda neste módulo, as imagens são ordenadas de acordo com a similaridade, retornando as mais similares para a interface. Por fim, o módulo base de dados permite armazenar as imagens a serem recuperadas e seus vetores de características. Sistemas com tais funcionalidades são utilizados em várias aplicações, por exemplo, para apoiar médicos na recuperação de imagens relevantes de um grande banco de dados, o que reduz a trabalhosa pesquisa

manual e auxilia no diagnóstico - devido a evidências passadas de outros pacientes. A Figura 2.3 apresenta um resultado gerado por um sistema de recuperação baseado em conteúdo, considerando imagens do *dataset* PatchCamelyon (Veeling et al., 2018). O exemplo simula a recuperação das dez imagens mais relevantes, dada uma imagem de entrada. Imagens não relvantes à imagem de consulta são destacadas em vermelho.

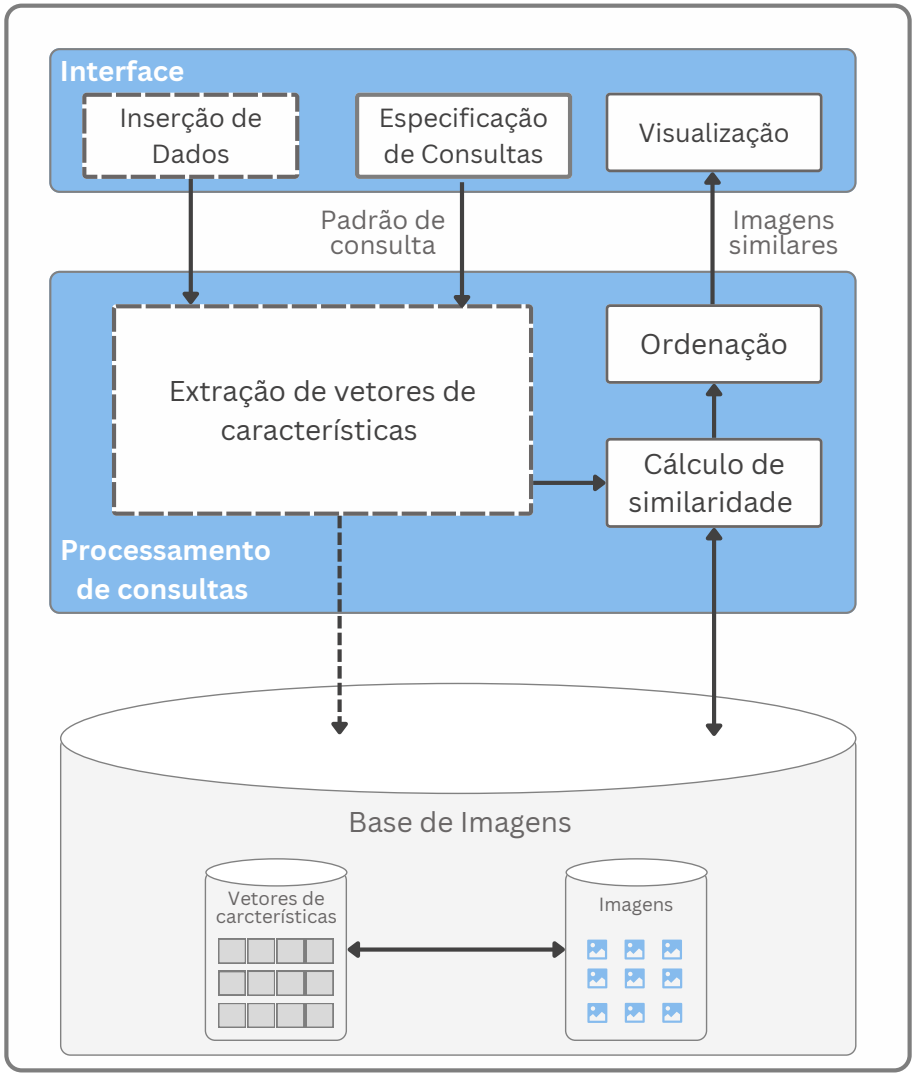


Figura 2.2: Arquitetura típica de um sistema de recuperação de imagens por conteúdo. Adaptado de Torres e Falcão (2006).

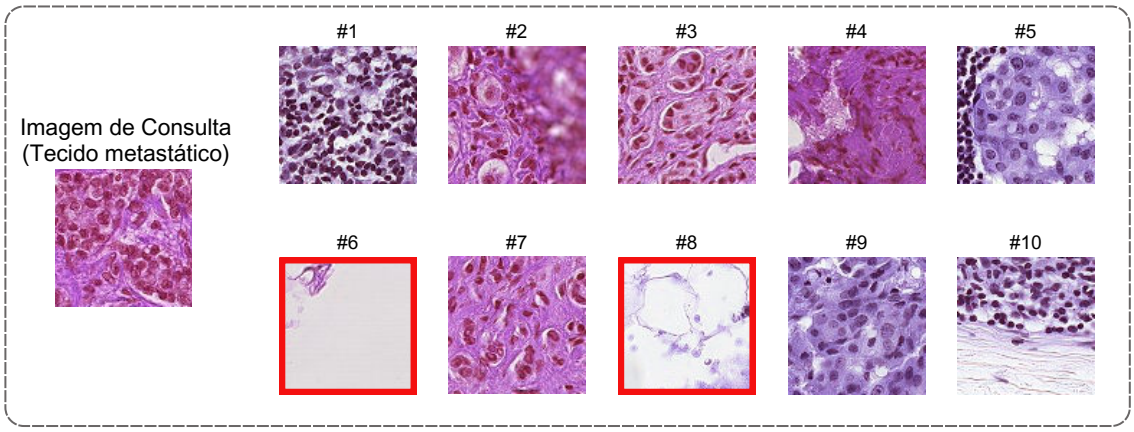


Figura 2.3: Top-10 resultados para a imagem de consulta especificada do *dataset* PatchCamelyon. Imagens não relevantes (que não apresentam tecido metastático) são destacadas em vermelho.

2.3 Aprendizado Auto-Supervisionado

Nos últimos anos, abordagens baseadas em *deep learning* foram desenvolvidas para resolver problemas em várias áreas (Pouyanfar et al., 2018). Diferentes arquiteturas foram projetadas para dar suporte a este paradigma, tais como: AlexNet, VGG, ResNet, MobileNet, entre outras (Jing e Tian, 2021). Nestas redes, o desempenho depende muito de sua capacidade de aprendizagem e da quantidade de dados de treinamento. Deste modo, quanto mais dados disponíveis maior o potencial de aprendizado. Contudo, a coleta e a anotação de conjuntos de dados em grande escala são processos demorados e dispendiosos. A dificuldade em coletar e anotar estes dados pode limitar o desenvolvimento de estudos em algumas áreas. Uma solução que vem sendo aplicada para atenuar esse problema é o uso de estratégias auto-supervisionadas.

A aprendizagem auto-supervisionada é um paradigma de aprendizagem recente que permite a aprendizagem de *features* semânticas através da geração de sinais de supervisão a partir de um conjunto de dados não rotulados, sem a necessidade de anotação humana (Chen et al., 2019). Por conseguinte, as *features* aprendidas durante o aprendizado auto-supervisionado são usadas para tarefas subsequentes onde a quantidade de dados, com ou sem anotação, é limitada. Geralmente, o aprendizado auto-supervisionado envolve duas tarefas, denominadas como *pretext task* e *downstream task*. Na *pretext task* é onde ocorre de fato a aprendizagem auto-supervisionada. Nesta etapa, um modelo é aprendido de forma supervisionada a partir de rótulos criados artificialmente, usando os dados não rotulados, permitindo o modelo aprender uma representação útil dos dados (Shurrab e Duwairi, 2022). Por outro lado, a *downstream task* é usada para avaliar a qualidade das *features* aprendidas na fase auto-supervisionada. Assim, nesta etapa é realizado o *fine-tuning* para atingir o objetivo pretendido, aprimorando muitas aplicações quando os da-

dos de treinamento são escassos. Tipicamente, é necessário que os dados usados nesta etapa estejam rotulados, mas existem algumas aplicações na qual isso não é necessário (Jing e Tian, 2021).

Na Figura 2.4 é apresentado o pipeline geral da aprendizagem auto-supervisionada. No passo 1 é realizado o treinamento do modelo a partir de uma *pretex task* predefinida. A forma como os pseudo-rótulos são gerados dependem da estratégia utilizada. Finalizado o treinamento, o passo 2 é iniciado. Nesta etapa, as *features* aprendidas anteriormente podem ser transferidas para serem utilizadas em uma *downstream task*. Na *downstream task* normalmente é realizado o *fine-tuning* dos modelos pré-treinados, considerando a nova tarefa alvo. Ressalta-se que diferentes arquiteturas de redes são utilizadas no aprendizado auto-supervisionado, tais como: VGG, AlexNet e ResNet. No entanto, o *Vision Transformer (ViT)*, que tem alcançado o estado da arte em muitas tarefas de visão computacional (Dosovitskiy et al., 2021), tem progressivamente substituído essas arquiteturas baseadas em redes convolucionais. Entre as principais abordagens auto-supervisionadas estão: SwAV (Caron et al., 2020), SimSiam Chen e He (2021), MoCo v3 (Chen et al., 2021), iBOT (Zhou et al., 2022), EsVIT (Li et al., 2022), e DINOv2 (Oquab et al., 2023).

Normalmente, o aprendizado auto-supervisionado é dividido em três abordagens: aprendizagem contrastiva, aprendizagem não contrastiva e baseadas em agrupamento. A aprendizagem contrastiva consiste em amostrar pares positivos e negativos, com o objetivo de aprender uma representação em que os pares positivos sejam mapeados para regiões próximas no espaço de características, enquanto os negativos sejam projetados para regiões distantes (Wickstrøm et al., 2023b). Em contrapartida, no método não contrastivo, ao invés de utilizar pares negativos e positivos, são utilizados apenas pares positivos. Assim, o método aprende como produzir uma representação útil maximizando a concordância entre pares positivos de amostras (Wickstrøm et al., 2023b; Kang et al., 2023b). Já nas abordagens baseadas em agrupamento, o método discrimina entre *clusters* de representações de imagens em vez de pares explícitos de imagens. São utilizados algoritmos de agrupamento para produzir pseudo-rótulos que, em sequência, são usados para aprender uma representação útil dos dados.

A condução da *pretex task* pode ser executada mediante diferentes estratégias. Na Figura 2.2 apresentamos um exemplo de tarefa de predição de posição relativa, comumente utilizada como *pretext task* no paradigma de aprendizado auto-supervisionado (Doersch et al., 2015). No exemplo, uma imagem é dividida em nove *patches*, no qual o *patch* central (aquele sem número) representa a âncora e os demais representam os de consulta. Neste tipo de tarefa, o treinamento auto-supervisionado consiste em utilizar um *patch* de âncora e um *patch* de consulta. Em resumo, dada uma coleção de imagens não rotuladas, são extraídos pares aleatórios de *patches* de cada imagem, seguido por um treinamento utilizando uma rede neural convolucional para prever a posição do segundo *patch* em relação ao primeiro. Para isso, é utilizada uma arquitetura de fusão tardia, a partir de duas redes convolucionais (AlexNet), que processam cada *patch* separadamente. Nesta arquitetura, a saída de cada uma

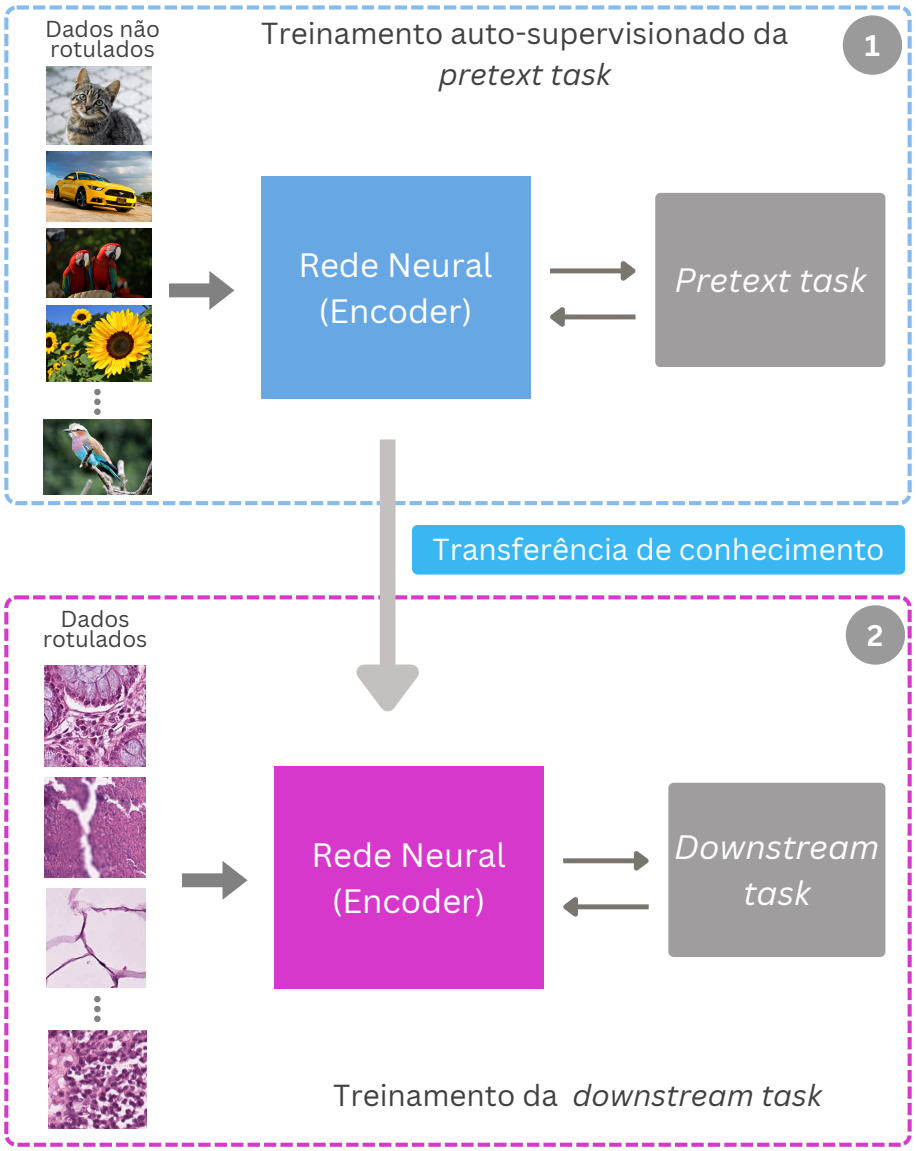


Figura 2.4: Pipeline geral da aprendizagem auto-supervisionada.

das redes (um *embedding* de 4096), como na AlexNet, são fundidas. Ressalta-se que os pesos destas redes são vinculados entre ambos os lados, de modo que a mesma função de *embedding* da camada totalmente conectada seja computada para ambos os *patches* (Doersch et al., 2015). Conforme salienta Doersch et al., espera-se que ao longo do processo de treinamento, o contexto espacial dos objetos na imagem de entrada seja compreendido.

A Figura 2.6 apresenta outro exemplo de *pretext task*, denominado como codificador de contexto (Pathak et al., 2016). Trata-se de uma tarefa generativa, que visa aprender as representações por meio do preenchimento de lacunas. Nesta abordagem, parte da imagem de entrada é recortada ou mascarada, e o objetivo da rede é aprender a completar/restaurar a parte oculta. O aprendizado das representações

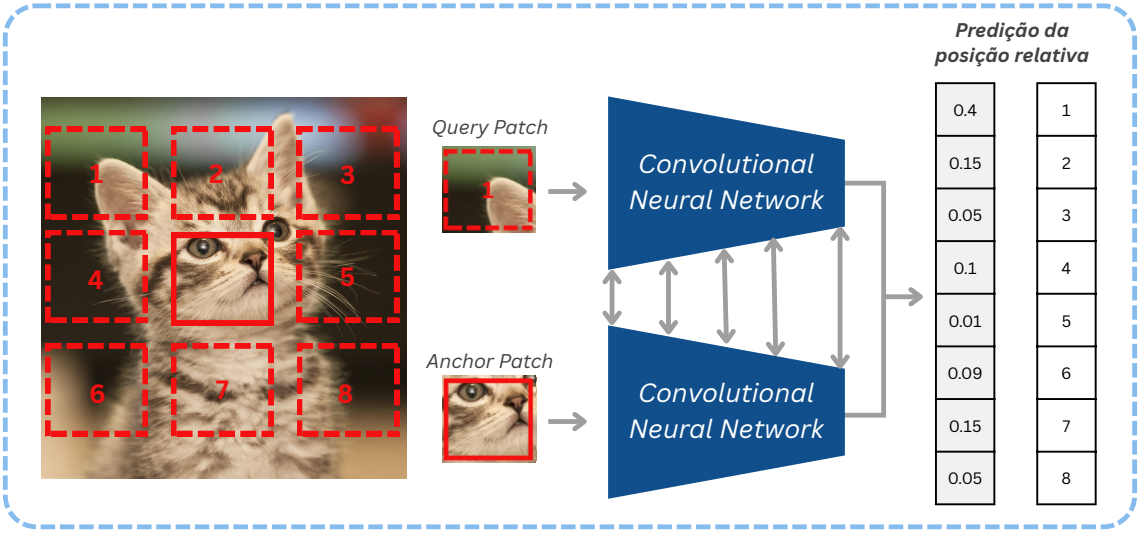


Figura 2.5: Exemplo de aprendizagem auto-supervisionada utilizando a tarefa de predição da posição relativa. Adaptado de (Shurrab e Duwairi, 2022)

utiliza uma rede de autocodificadores e um espaço latente totalmente conectados. Além disso, conforme indicado pelos autores, utiliza uma função de perda combinada que integra a perda da reconstrução e a perda adversarial, de modo a possibilitar um treinamento mais eficaz.

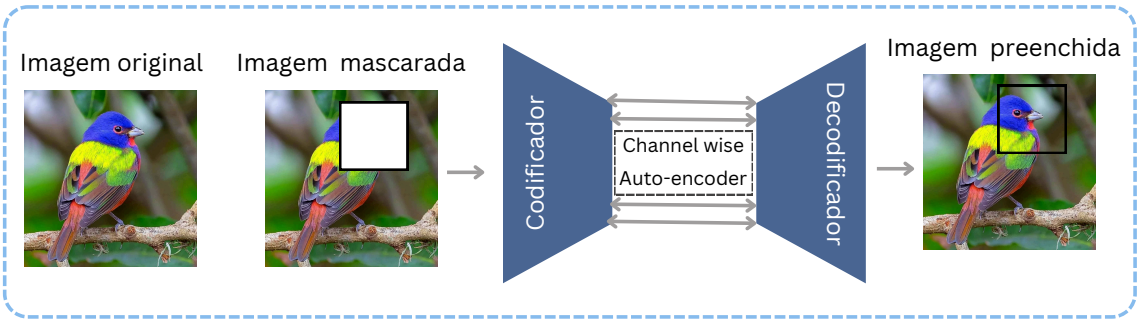


Figura 2.6: Exemplo de aprendizagem auto-supervisionada utilizando codificador de contexto

Esses dois casos exemplificam estratégias que podem ser utilizadas no processo de aprendizado de representações a partir do paradigma auto-supervisionado. No entanto, salienta-se que, na literatura, existem muitas outras estratégias, tais como: colorir imagens em tons de cinza, previsão de rotação, *quebra-cabeça de imagem*, modelagem de imagem mascarada, entre outros (Jing e Tian, 2021; Bao et al., 2021).

2.4 Balanceamento de Dados e Medidas de Avaliação

O número de classes que compõem um *dataset* é uma das características que devem ser cuidadosamente analisadas, dado o impacto direto no processo de geração dos modelos. Quando há grande variação no número de classes presentes, os modelos tendem a identificar algumas classes com maior eficácia que outras, prejudicando a qualidade geral dos modelos desenvolvidos (Han et al., 2011). Assim, verificar a proporção das classes contribui para análises mais precisas do desempenho destes sistemas. Além disso, compreender o nível de desbalanceamento da base de dados também auxilia a determinar as medidas mais adequadas para conduzir a avaliação. Afinal, algumas medidas, como a acurácia, podem induzir a interpretações equivocadas devido ao grau de desbalanceamento.

Durante o processo de desenvolvimento de um sistema, seja ele voltado para classificação de imagens ou aplicado ao contexto de CBIR, uma etapa tão importante quanto as demais é o processo de avaliação de sua eficácia. Essa etapa é ainda mais significativa quando o sistema sob análise será utilizado em aplicações na área média, tendo em vista o impacto direto na vida das pessoas. Para esta categoria de aplicação, um erro pode resultar em consequências irreversíveis.

A seleção de quais medidas são aplicadas para avaliar a eficácia desses sistemas variam conforme o objetivo da tarefa e as características da base de dados. Por exemplo, medidas que são utilizadas em problemas binárias não podem ser aplicadas diretamente em tarefas multiclasse (Han et al., 2011). Do mesmo modo, determinadas medidas, se aplicadas em bases de dados desbalanceadas, podem levar a interpretações equivocadas. A avaliação da eficácia é tão importante quanto a etapa de treinamento dos modelos, pois, através do *feedback* que essas medidas fornecem, ajustes podem ser feitos, melhorando o desempenho geral do sistema.

Considerando a tarefa de classificação multiclasse, comumente são utilizadas as medidas: macro precision, macro recall, macro F_1, micro precision, micro recall e micro F_1 (Bishop, 2006). A seguir, apresenta-se detalhadamente cada uma delas.

- **Macro Precision:** Corresponde a média da *precision* de todas as classes. Inicialmente calcula-se a *precision* para cada classe individualmente e depois é feita a média. Ressalta-se que a *precision* é a proporção de verdadeiros positivos em relação ao total de positivos previstos. A macro precision é calculada pela expressão:

$$Macro\ Precision = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FP_i}$$

- **Macro Recall:** Corresponde a média do *recall* de todas as classes. Inicialmente calcula-se o *recall* para cada classe individualmente e depois é feita a

média. O *recall* é a proporção de verdadeiros positivos em relação ao total de positivos reais. Assim, esta medida é capaz de avaliar a cobertura do modelo. A macro recall é calculado pela expressão:

$$Macro\ Recall = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FN_i}$$

- **Macro F_1:** Corresponde a média do F_1 de todas as classes. Inicialmente calcula-se o F_1 para cada classe individualmente e depois é feita a média. Ressalta-se que o F_1 representa a média harmônica entre *precision* e *recall*, fornecendo uma única métrica que balanceia ambas as medidas. A macro F_1 é calculado pela expressão:

$$\text{Macro } F_1 = \frac{1}{C}\sum_{i=1}^{C}\frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

  Onde Precision_i e Recall_i são a *precision* e o *recall* da classe i, respectivamente.

- **Micro Precision:** Corresponde a proporção de verdadeiros positivos (TP) em relação a todos os exemplos de itens preditos como positivos, considerando todas as classes. Ela mede o quão preciso o modelo é ao prever o rótulo positivo. A micro precision é calculada pela expressão:

$$\text{Micro Precision} = \frac{\sum_{i=1}^{C}\text{TP}_i}{\sum_{i=1}^{C}(\text{TP}_i + \text{FP}_i)}$$

- **Micro Recall:** Corresponde a proporção de verdadeiros positivos (TP) em relação a todos os exemplos reais positivos, considerando todas as classes. Assim, essa medida avalia o quão bem o modelo identifica os exemplos positivos. A micro recall é calculada pela expressão:

$$\text{Micro Recall} = \frac{\sum_{i=1}^{C}\text{TP}_i}{\sum_{i=1}^{C}(\text{TP}_i + \text{FN}_i)}$$

- **Micro F_1:** Corresponde a média harmônica entre a micro precision e a micro recall. Isto é, fornece uma visão equilibrada do desempenho do modelo, considerando precisão e sensibilidade. A micro F_1 é calculada pela expressão:

$$\text{Micro } F_1 = \frac{2 \cdot \text{Micro Precision} \cdot \text{Micro Recall}}{\text{Micro Precision} + \text{Micro Recall}}$$

Notações:

- C: Número de classes.

- TP_i: Número de verdadeiros positivos da classe i.
- FP_i: Número de falsos positivos da classe i.
- FN_i: Número de falsos negativos da classe i.

As medidas macro e micro apresentadas são usadas em problemas de classificação multiclasse, mas diferem na forma como tratam as classes e agregam os resultados. Para ilustrar as suas principais diferenças, a tabela 2.1 apresenta um comparativo, considerando os seguintes aspectos: Forma de calcular, peso das classes, melhor cenário de uso e sensibilidade ao desbalanceamento.

Tabela 2.1: Principais diferenças entre medidas macro e micro

Aspecto	**Medidas Macro**	**Medidas Micro**
Forma de calcular	Média das medidas calculadas por classe	Agregação dos valores de *TP*, *FP* e *FN* de todas as classes.
Peso das Classes	Todas as classes têm o mesmo peso	Classes maiores têm maior impacto no resultado.
Melhor cenário de uso	Quando todas as classes têm igual importância.	Em datasets desbalanceados.
Sensibilidade ao desbalanceamento	Sensível ao desbalanceamento de classes	Menos sensível ao desbalanceamento

Considerando a avaliação de modelos em relação à tarefa de recuperação de imagens baseada em conteúdo, várias medidas podem ser utilizadas, destacando-se a *Mean Average Precision (MAP) at K* ou simplesmente $MAP@K$. Esta medida permite avaliar tanto a relevância das imagens ranqueadas quanto a eficácia do sistema em posicionar as imagens relevantes nas primeiras posições (Manning et al., 2008). Em geral, cada imagem ranqueada possui um grau de relevância previamente atribuído por um especialista. Contudo, existem outras maneiras de avaliar a relevância das imagens sem necessariamente utilizar um escore. Por exemplo, pode-se considerar como relevantes as imagens que pertencem à classe da imagem de consulta.

Entre as principais características da $MAP@K$, destacam-se:

- **Avaliação de Relevância:** A $MAP@K$ permite avaliar a qualidade, ao verificar a eficácia de um sistema de recuperação de imagens em posicionar imagens relevantes nas primeiras posições do *ranking*. Em um sistema de CBIR, especialmente no contexto médico, essa característica é essencial, pois permite que os patologistas tenham acesso imediato às imagens mais semelhantes à imagem de consulta, já nas primeiras posições do resultado.
- **Comparação de Sistemas:** A $MAP@K$ pode ser utilizada para comparar a eficácia de diferentes sistemas de recuperação de imagens. Deste modo, sistemas que têm uma $MAP@K$ mais alta são melhores em posicionar imagens relevantes no topo do *ranking*, proporcionando uma melhor experiência ao patologista.
- **Análise de Desempenho:** Ao utilizar a $MAP@K$ é possível identificar como o desempenho de um sistema varia conforme o número K muda. Esse tipo de

avaliação permite fornecer *insights* sobre como o sistema lida com a recuperação de imagens relevantes em diferentes profundidades do *ranking* resultante.

A $MAP@K$ pode ser calculada a partir da expressão:

$$\text{MAP@K} = \frac{1}{Q} \sum_{q=1}^{Q} \text{AP@K}(q)$$

onde:

$$\text{AP@K}(q) = \frac{1}{\text{N@K}(q)} \sum_{i=1}^{K} \text{Precision@i}(q) \cdot \text{rel}(i, q)$$

Q é o número total de consultas.
AP@K(q) é a *precision* média na posição K para a consulta q.
N@K(q) é o número de imagens relevantes retornadas nos primeiros K resultados para a consulta q.
Precision@i(q) é a *precision* na posição i para a consulta q.
rel(i, q) é uma função que é 1 se o item na posição i for relevante para a consulta q e 0 caso contrário.

Além das medidas de avaliação de eficácia, existem algumas aplicações onde é necessário calcular a similaridade entre pares de dados, para gerar o resultado esperado, como os sistemas de CBIR. Este cálculo pode ser conduzido a partir de vários métodos, sendo a similaridade cosseno uma das mais difundidas. A similaridade cosseno é uma métrica usada para medir o grau de semelhança entre dois vetores em um espaço multidimensional. A similaridade cosseno entre dois vetores **A** e **B** é definida como:

$$\text{similaridade cosseno} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

onde:

- $\mathbf{A} \cdot \mathbf{B}$ é o produto interno (ou escalar) dos vetores **A** e **B**.
- $\|\mathbf{A}\|$ e $\|\mathbf{B}\|$ são as normas (ou magnitudes) dos vetores **A** e **B**, respectivamente.
- θ é o ângulo entre os vetores **A** e **B**.

A similaridade cosseno varia de -1 a 1, onde:

- **1** indica que os vetores estão perfeitamente alinhados na mesma direção.
- **0** indica que os vetores são ortogonais.
- **-1** indica que os vetores estão em direções opostas.

2.5 Trabalhos Relacionados

Sistemas baseados em computação inteligente para manipulação de imagens médicas têm sido cada vez mais desenvolvidos por empresas e pesquisadores em todo mundo (Hosseini et al., 2024). Esse crescimento se deve, especialmente, ao desenvolvimento de métodos robustos de computação inteligente, ganhos contínuos no poder de processamento, velocidade de transferência de dados, avanços em soluções de armazenamento em nuvem, entre outros. Isso tem permitido o uso de imagens digitais para uma ampla variedade de finalidades na medicina, em especial patologia (Navid Farahani et al., 2015). De fato, avanços recentes nas técnicas de visão computacional, aprendizado de máquina e recuperação de informação têm facilitado o processo de análise de imagens médicas, superando os métodos de aprendizado de máquina tradicionais (Chan et al., 2020) e abrindo portas para que sistemas cada vez mais eficazes sejam desenvolvidos e implantados na prática diária dos serviços de saúde.

As primeiras tentativas de usar computadores para analisar automaticamente imagens médicas remontam a década de 1960. No entanto, apesar de alguns estudos demonstrarem a viabilidade do uso prático, a ideia não atraiu muito apoiadores, provavelmente devido ao acesso limitado a dados de imagens digitalizadas de alta qualidade e recursos computacionais escassos da época (Chan et al., 2020). Nos anos seguintes, novas pesquisas foram desenvolvidas, incluindo estudos usando aprendizado de máquina (Doi, 2007). Contudo, somente em 1998 o primeiro sistema comercial dos Estados Unidos foi aprovado pela *Food and Drug Administration (FDA)* para uso como segunda opinião no rastreamento de mamografia (Dromain et al., 2012). Atualmente, sistemas dessa categoria integram muitas rotinas clínicas de radiologia diagnóstica e mais recentemente tornou-se também uma aposta para a patologia digital (Hosseini et al., 2024). Isso tem contribuído para o crescimento contínuo de pesquisas nessa área, que abrangem vários domínios de imagens médicas, tais como: radiografia, mamografia, ressonância magnética, tomografia computadorizada e histopatologia.

Com o aumento da carga de trabalho dos patologistas, existe uma necessidade de integrar sistemas de apoio à tomada de decisão nas suas rotinas de trabalho, como forma de auxiliar esses profissionais em suas práticas e manejo clínico. De modo geral, estes sistemas podem ser divididos em duas categorias: aqueles baseados em classificação e os relacionados à recuperação de imagens baseada em conteúdo. Os avanços gerados por esses sistemas tem estimulado um debate sobre a plausibilidade do diagnóstico assistido, usando tais recursos computacionais. A confiabilidade dos sistemas de IA continua sendo uma preocupação, e exige um debate amplo com delineamento de protocolos para fundamentar projetos experimentais e dar suporte a conclusões confiáveis (Calumby et al., 2023). Apesar dessa preocupação constante, a integração dessas soluções de software oferece uma solução inovadora, não apenas para a busca em grandes bases de dados, mas também para melhorar a qualidade e a agilidade nos diagnósticos patológicos, representando um avanço significativo na

área da saúde.

Em um sistema de CBIR, as imagens que compartilham atributos visuais, como cor, textura e forma, com uma imagem de consulta, são indexadas e recuperadas de grandes bases de dados. Neste contexto, nos últimos anos, vários trabalhos foram desenvolvidos, utilizando-se frequentemente abordagens mais avançadas de Inteligência Artificial, como *deep learning* e aprendizado auto-supervisionado. Entre trabalhos recentes relevantes, destacamos: Hegde et al. (2019), Yang et al. (2020), Kalra et al. (2020), Zheng et al. (2022), Mohammad Alizadeh et al. (2023), Wickstrøm et al. (2023b), Wang et al. (2023) e Filiot et al. (2023), que são descritos a seguir.

Hegde et al. (2019) desenvolveram uma ferramenta de recuperação de imagens histopatológicas baseada em uma arquitetura de rede neural convolucional chamada rede de classificação profunda Wang et al. (2014). Esta rede é baseada em um módulo de *embeddings* que comprime partes da imagem de entrada (de dimensões largura x altura x canais) em um vetor de comprimento fixo. Os autores denominaram a ferramenta de SMILY, do inglês *Similar Medical Images Like Yours*. A abordagem foi avaliada usando dados oriundos do TCGA, considerando três tipos de órgãos: próstata, cólon e pulmão. O método foi avaliado usando anotações fornecidas por patologistas e por meio de estudos prospectivos onde os patologistas avaliaram a qualidade dos resultados gerados. Em ambos os tipos de avaliações, o SMILY conseguiu recuperar imagens similares à imagem de consulta submetida.

Yang et al. (2020) projetaram um mecanismo de busca de imagens para patologia digital focado na representação de WSIs de maneira compacta. Os autores desenvolveram um algoritmo de indexação para representar WSIs como um mosaico de *patches* que são então convertidos em códigos de barras, denominado de "Bunch of Barcodes" (BoB). Similar ao trabalho de Hegde et al. (2019), os autores avaliaram a abordagem utilizando um *ground-truth*, e também uma avaliação subjetiva, com usuários especialistas e não especialista do domínio. Os resultados encontrados foram promissores, indicado que o mecanismo pode recuperar com acurácia, órgãos e malignidades, e sua ordenação semântica mostra uma concordância efetiva com a avaliação subjetiva de observadores humanos e o mecanismo de busca. Em Zheng et al. (2022), os autores também propuseram uma abordagem para representação da WSI. Foi projetado um *framework* baseado em grafo para prover busca por regiões de interesse. Além disso, foi proposto um processo de codificação da *features* por meio de valores binários. Os resultados encontrados apontaram a superioridade do método desenvolvido frente a outras abordagens similares.

Em Mohammad Alizadeh et al. (2023), os autores propuseram um novo método de rede neural convolucional siamesa baseado em *hashing* para recuperação de imagens histopatológicas. Os autores também utilizaram uma codificação binária das *features*. Para isso, foram utilizados dois modelos de *hashing* profundos com pesos e estruturas compartilhadas. Também foi projetada uma nova função de custo para aprimorar o processo de treinamento e recuperação das imagens. O método foi avaliado em dois *datasets* públicos, e de acordo com os resultados experimentais, o

modelo desenvolvido superou outros métodos baseados em *hashing*.

Em um outro estudo, diferentemente das abordagens anteriores, os autores desenvolveram uma *framework* auto-supervisionado para recuperação de imagens de tomografia computadorizada do fígado (Wickstrøm et al., 2023b). Além disso, considerando a importância da explicabilidade, especialmente para aplicações médicas, foi realizada a primeira análise de explicabilidade do aprendizado de representação no contexto de CBIR de imagens de tomografia computadorizada do fígado. A explicabilidade foi conduzida por um framework denominado RELAX (do inglês, Representation Learning Explainability) (Wickstrøm et al., 2023a). Os resultados encontrados apontam que a abordagem auto-supervisionada permitiu extrair *features* clinicamente relevantes. Além disso, a usabilidade prática do *framework* proposto também foi validado por um especialista, comparando o resultado recuperado pelo sistema com o *ground-truth*.

Em outro estudo, foi desenvolvido o RetCCL, um *framework* para recuperação de imagens histopatológicas, tanto no nível de WSI quanto no nível de Patch (Wang et al., 2023). Este framework integra um novo método de aprendizado de *features* auto-supervisionado e um algoritmo global de classificação e agregação. O treinamento do modelo utilizado pelo *framework* foi conduzido a partir de uma estratégia auto-supervisionada, utilizando-se aproximadamente 15 milhões de imagens histopatológicas não rotuladas. Dada a utilização dessa base de dados em larga escala, os autores argumentam que o modelo desenvolvido pode ser aplicado diretamente em tarefas subsequentes, sem a necessidade de *fine-tuning*, tendo em vista que, durante o processo de treinamento, *features* consideradas universais são aprendidas. O RetCCL foi avaliado utilizando imagens de patologias associadas a diferentes tipos e subtipos de câncer. Os resultados demonstraram que o *framework* superou significativamente os métodos de estado da arte disponíveis até então, com uma média de aproximadamente 10% em termos de $mMV@10$ (voto majoritário no topo dos k resultados da busca).

Filiot et al. (2023) exploraram a aplicação da modelagem de imagem mascarada (do inglês, Masked Image Modeling (MIM)) em imagens histológicas usando a estrutura iBOT baseada em ViT. A MIM é projetada para treinar modelos de visão computacional ao aprender a prever partes ocultas ou mascaradas de uma imagem. No trabalho de Filiot et al. (2023), o modelo pré-treinado foi avaliado em várias tarefas *downstream*, todas elas relacionadas com o domínio de câncer. Apesar dos autores não utilizarem o modelo pré-treinado no contexto de sistemas CBIR, produziram grandes contribuições, por exemplo, a disponibilização dos modelos pré-treinados para servirem como modelos base para estudos futuros, avaliação utilizando diferentes tarefas de classificação, vários *datasets*, análise da escalabilidade da MIM, entre outros.

Diferentemente dos trabalhos anteriores, em nosso estudo, são utilizados métodos auto-supervisionados ajustados e avaliados em *datasets* reduzidos não rotulados, como forma de verificar a sua aplicabilidade em tais cenários. Também avalia-se

o poder de generalização de métodos de estado da arte, considerando vários tipos de lesões e patologias. Para todos os cenários investigados, considera-se a tarefa de classificação e recuperação de imagens baseada em conteúdo. Considerando que, no mundo real, as bases de dados geralmente não são rotuladas, este estudo oferece uma investigação relevante para aplicações futuras na área médica ou outras áreas que enfrentam desafios semelhantes.

Capítulo 3

Processo Experimental

"Em algum lugar, alguma coisa incrível está esperando para ser descoberta."

– Carl Sagan

O processo experimental conduzido neste estudo está ilustrado na Figura 3.1. Foi composto por 6 etapas: (1) Coleta de dados; (2) Particionamento dos *datasets*; (3) Pré-processamento; (4) Adaptação dos modelos pré-treinados (*fine-tuning*); (5) Extração dos vetores de características; e, por fim, (6) Avaliação dos modelos em tarefas de classificação e de recuperação de imagens histopatológicas.

3.1 Seleção de dados

No processo experimental foram utilizados sete *datasets*, abrangendo diversos tipos de tecidos histológicos e diferentes patologias ou lesões associadas. Neste estudo, os *datasets* foram denominados como: AIDPATH (Pública), PathoSpotter-HE (Privada), PathoSpotter-PAS (Privada), PathoSpotter-MultiContraste (Privada), Kather (Pública), RCC (Pública) e NCT-CRC-HE(Pública). A seguir, apresentamos uma descrição detalhada de cada uma dessas bases de dados.

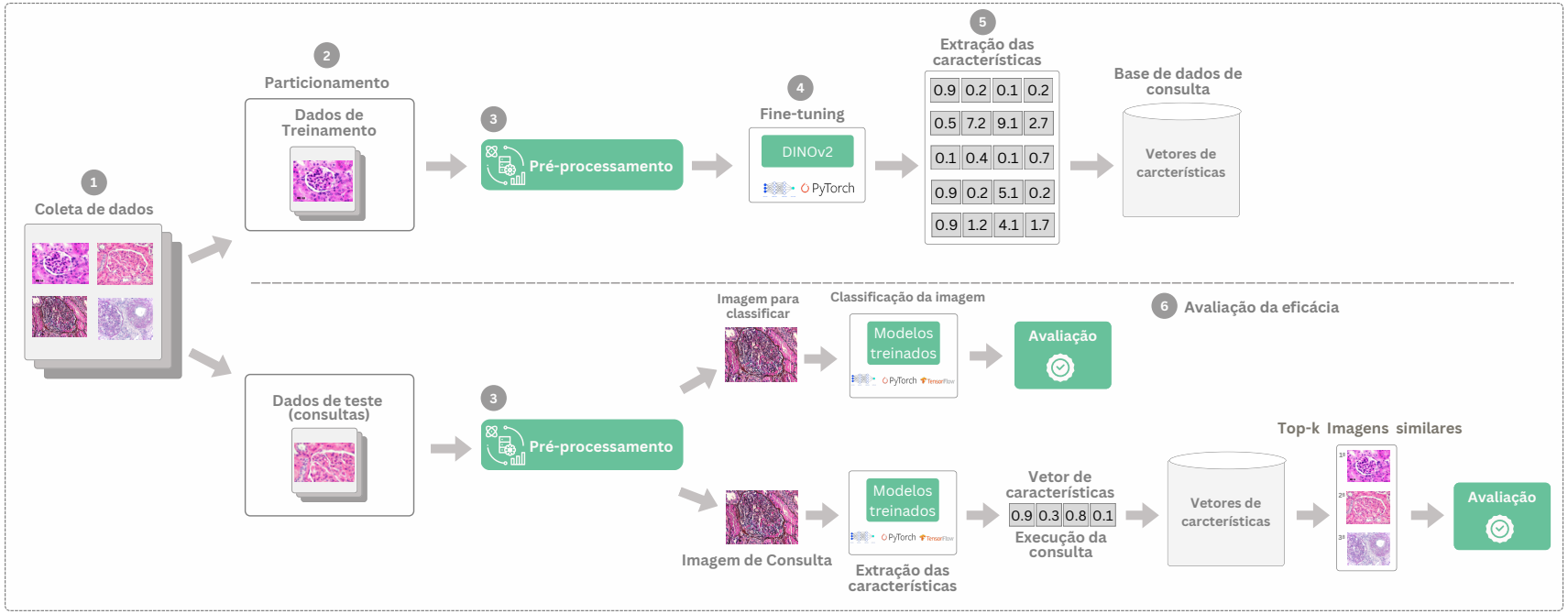
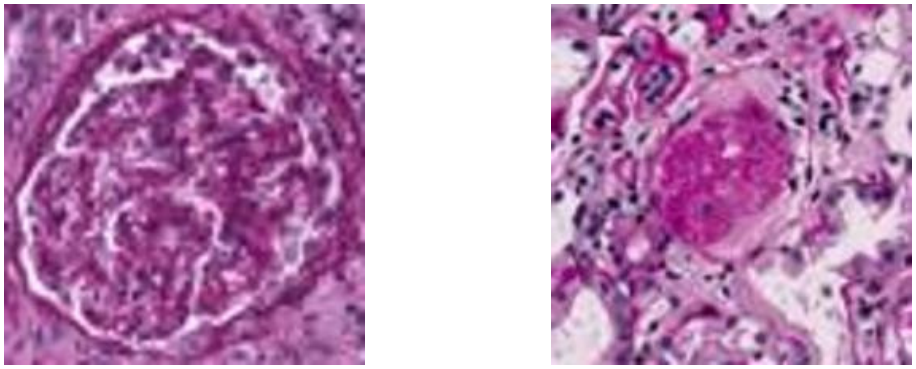


Figura 3.1: Workflow experimental utilizado no desenvolvimento deste estudo

3.1.1 AIDPATH

Esta base faz parte dos conjuntos de dados de WSI gerados no projeto europeu AIDPATH[1]. A base de dados contém 2340 imagens, cada uma apresentando um único glomérulo, que são utilizadas para o desenvolvimento de estudos relacionados à identificação de Glomeruloesclerose (Bueno et al., 2020). Das 2340 imagens, 1170 apresentam glomérulos normais, enquanto as demais 1170 apresentam glomérulos esclerosados. A Figura 3.2 ilustra um exemplo de cada uma das classes de interesse. Para este *dataset* foi utilizado o corante PAS. A coloração PAS é comumente usada devido à sua eficiência na coloração de polissacarídeos, que estão presentes no tecido renal e no realce das membranas basais glomerulares (Bueno et al., 2020).



(a) Glomérulo normal (b) Glomérulo com esclerose

Figura 3.2: Exemplo de imagens do dataset AIDPATH.

3.1.2 PathoSpotter-HE

Este dataset faz parte do projeto *PathoSpotter*[2], vinculado a Fundação Oswaldo Cruz (Fiocruz) da Bahia. De modo geral, o projeto *PathoSpotter* objetiva auxiliar os patologistas no cálculo das estatísticas de atividade e cronicidade das lesões renais. A ideia central do projeto é promover correlações em larga escala das lesões com dados clínicos e históricos dos pacientes e, ao mesmo tempo, criar um conjunto de ferramentas para auxiliar a prática diária dos patologistas. Este *dataset* possui 4947 imagens, divididas em seis classes de interesse: Normal (869), Hipercelularidade (1237), Membranosa primária (712), Membranosa secundária (1354), Esclerose com membranosa (264) e Esclerose sem membranosa (511). Para este *dataset*, o HE presente no nome PathoSpotter-HE, indica que foi utilizado o corante Hematoxilina e Eosina (H&E). Na Figura 3.3 apresentamos amostras de imagens para cada uma das classes rotuladas no PathoSpotter-HE.

[1] `https://aidpath.eu/`

[2] `https://pathospotter.bahia.fiocruz.br/`

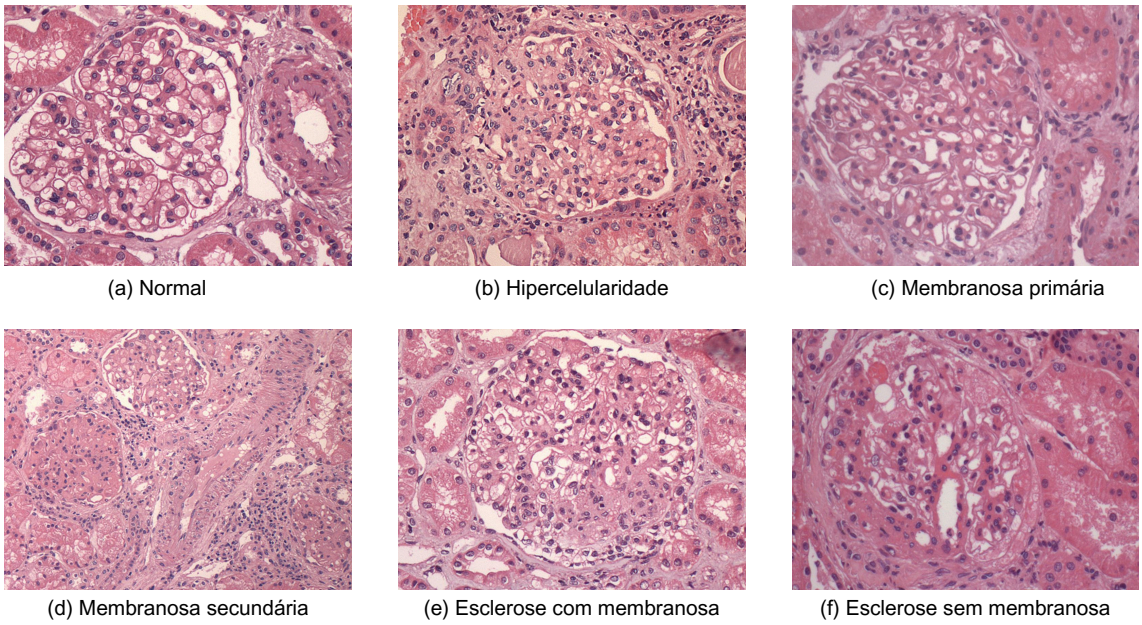


Figura 3.3: Exemplo de imagens do *dataset* PathoSpotter-HE.

3.1.3 PathoSpotter-PAS

Este *dataset* também faz parte do projeto PathoSpotter, mas diferentemente do PathoSpotter-HE, utiliza o PAS como corante. Esta base de dados possui 2390 imagens, e também está dividida em seis classes de interesse: Normal (293), Hipercelularidade (637), Membranosa primária (367), Membranosa secundária (609), Esclerose com membranosa(156) e Esclerose sem membranosa (328). A Figura 3.4 apresenta imagens das classes de interesse.

3.1.4 PathoSpotter-MultiContraste

Este *dataset* também integra o projeto PathoSpotter, mas ao contrário do PathoSpotter-HE e PathoSpotter-PAS, ele contém imagens obtidas com diferentes corantes. Além de incluir imagens processadas com H&E e PAS, também apresenta imagens obtidas com PAMS, PS, AZAN e PICRO. A base é composta por 10184 imagens, também abrangendo seis classes: Normal (1570), Hipercelularidade (2043), Membranosa primária (1608), Membranosa secundária (3170), Esclerose com membranosa(513) e Esclerose sem membranosa (1280). A Figura 3.5 apresenta exemplos de imagens pertencentes a este conjunto de dados.

3.1.5 Kather

Este *dataset* é composto por imagens de amostras histológicas de câncer colorretal humano, coradas com H&E (Kather et al., 2019). É um subconjunto da base de

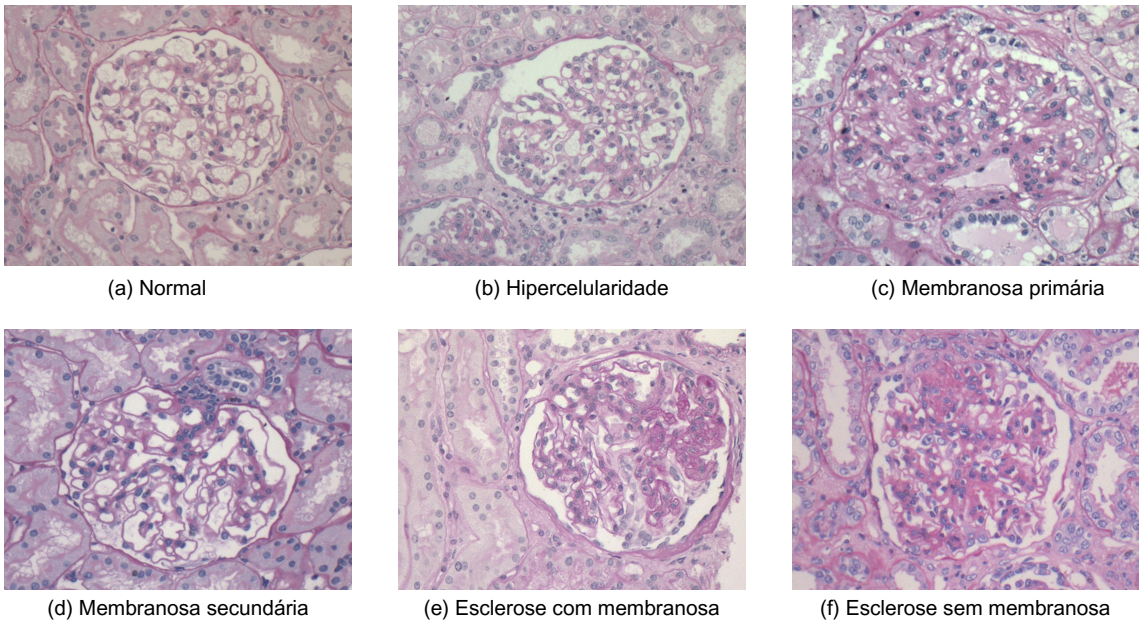


Figura 3.4: Exemplo de imagens do *dataset* PathoSpotter-PAS.

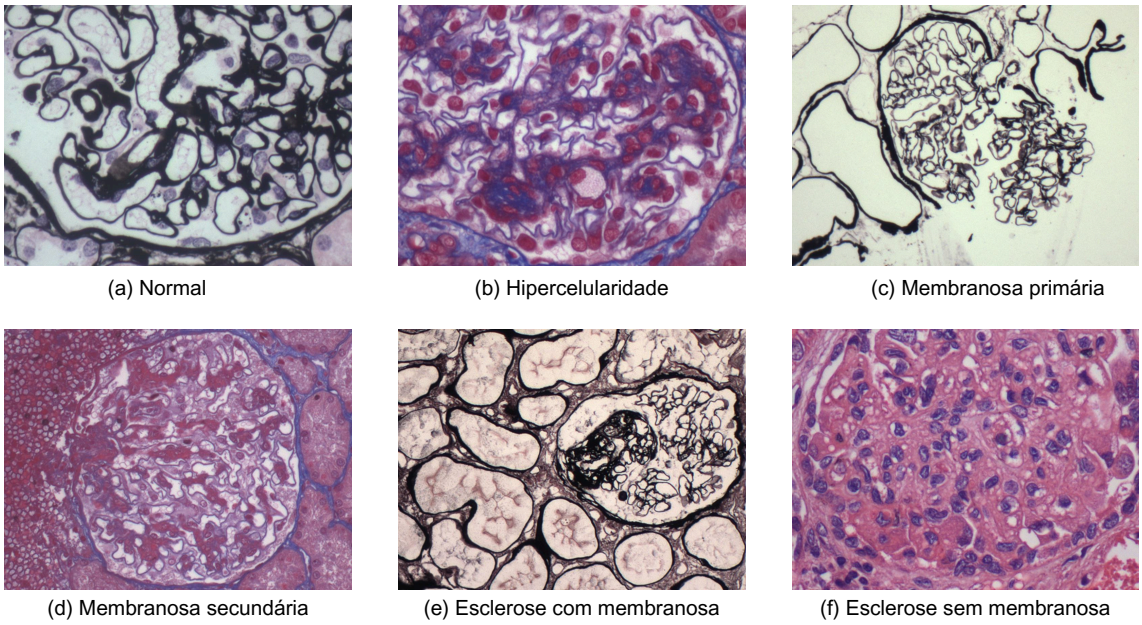
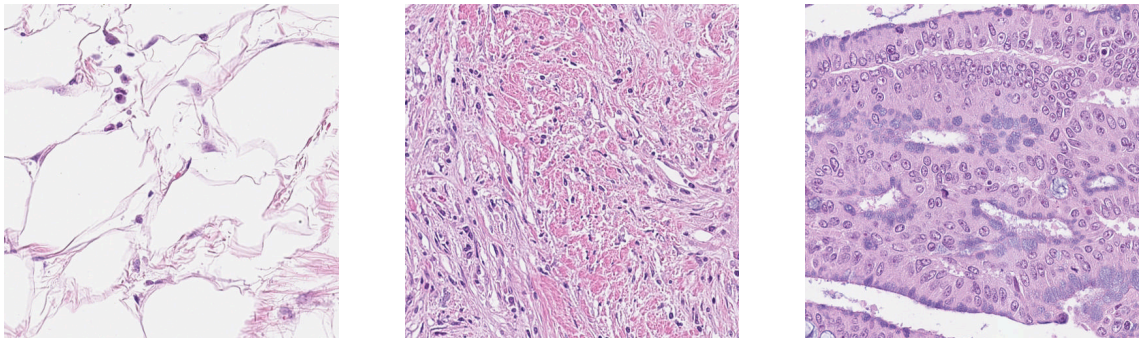


Figura 3.5: Exemplo de imagens do *dataset* PathoSpotter-MultiContraste.

dados denominada NCT-CRC-HE-100K[3], que contém 100.000 imagens histológicas de câncer colorretal humano e tecido saudável. O *dataset* possui 11977 imagens, categorizadas em três classes: tecido adiposo e muco (ADIMUC) com 3977 imagens, estroma e músculo (STRMUS) com 4000 imagens e tecido epitelial de câncer

[3] `https://zenodo.org/records/1214456`

colorretal e tecido epitelial de câncer de estômago (TUMSTU), também com 4000 imagens. A Figura 3.6 exemplifica amostras das classes de interesse.



(a) ADIMUC (b) STRMUS (c)TUMSTU

Figura 3.6: Exemplo de imagens do *dataset* Kather.

3.1.6 RCC

Este *dataset* possui imagens de carcinoma de células renais, obtidas de diferentes pacientes e graus de severidade variados, como parte de um estudo clínico do departamento de Patologia do Kasturba Medical College (KMC), Manipal, Índia (K. et al., 2023). O conjunto de dados inclui imagens não cancerosas (Grau 0) e cancerosas (Grau 1 a Grau 4) do carcinoma. As amostras foram coletadas por biópsia cirúrgica (aberta) de tecido renal, coradas com H&E. O *dataset* é composto por 4077 imagens, subdividido em 5 classes ou graus do câncer: Grau 0 (836), Grau 1 (839), Grau 2 (769), Grau 3 (861) e Grau 4 (772). A Figura 3.7 apresenta amostras representativas para cada um dos graus do câncer.

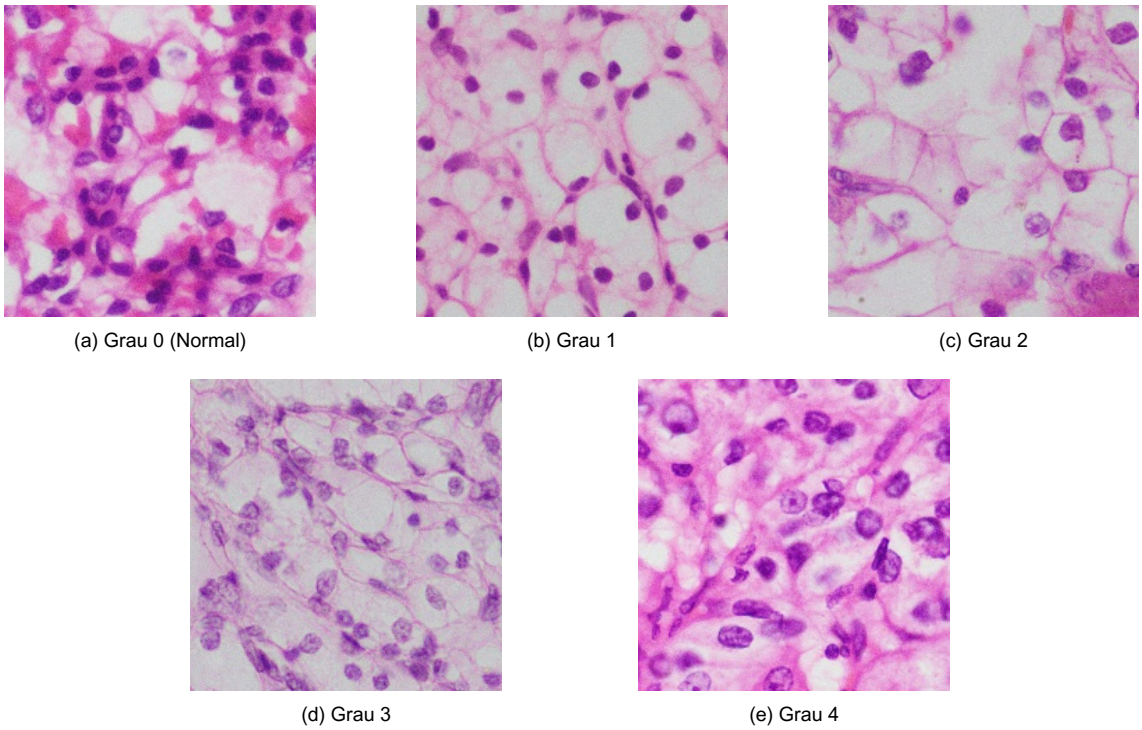


Figura 3.7: Exemplo de imagens do *dataset* RCC.

3.1.7 NCT-CRC-HE

Este *dataset* é composto pelas bases de dados NCT-CRC-HE-100K e CRC-VAL-HE-7K, denominado a partir deste ponto como NCT-CRC-HE. O NCT-CRC-HE-100K é um conjunto de dados de imagens de patologia, projetado para classificação de imagens, compreendendo 100.000 imagens histológicas coradas com H&E de câncer colorretal humano e tecidos saudáveis extraídos de 86 pacientes (Kather et al., 2018). É uma base amplamente utilizada em vários estudos envolvendo análise de imagens histopatológicas. Por outro lado, o conjunto CRC-VAL-HE-7K consiste em 7180 imagens extraídas de 50 pacientes com adenocarcinoma colorretal e foi usado para representar um conjunto de dados que não se sobrepõe aos pacientes do conjunto de dados NCT-CRC-HE-100K. Somadas, as bases incluem 107.180 imagens, categorizadas em nove classes: ADI (11.745), BACK(11.413), DEB(11.851), LYM(12.191), MUC(9.931), MUS(14.128), NORM(9.504), STR(10.867) e TUM(15.550). A Figura 3.8 apresenta exemplos representativos de cada tipo de tecido que compõe esta base de dados, que são descritos a seguir:

- ADI: Tecido adiposo, geralmente encontrado em áreas com gordura, fora da região do tumor.
- BACK: Áreas sem tecido significativo ou estruturas relevantes, geralmente regiões de fundo da lâmina.

- DEB: Resíduos celulares, necrose ou material fragmentado resultante de processos tumorais ou inflamatórios.
- LYM: Áreas ricas em linfócitos, indicando infiltrados imunológicos que podem estar associados a respostas imunológicas locais.
- MUC: Secreção de muco, comumente associada ao epitélio glandular e condições específicas do tecido colorretal.
- MUS: Tecido muscular, normalmente do tipo liso, encontrado em regiões do trato gastrointestinal.
- NORM: Tecido epitelial colorretal saudável, sem sinais de anomalias ou câncer.
- STR: Tecido conjuntivo de suporte, frequentemente associado ao crescimento do câncer.
- TUM: Células epiteliais de adenocarcinoma colorretal, um tipo de câncer de cólon.

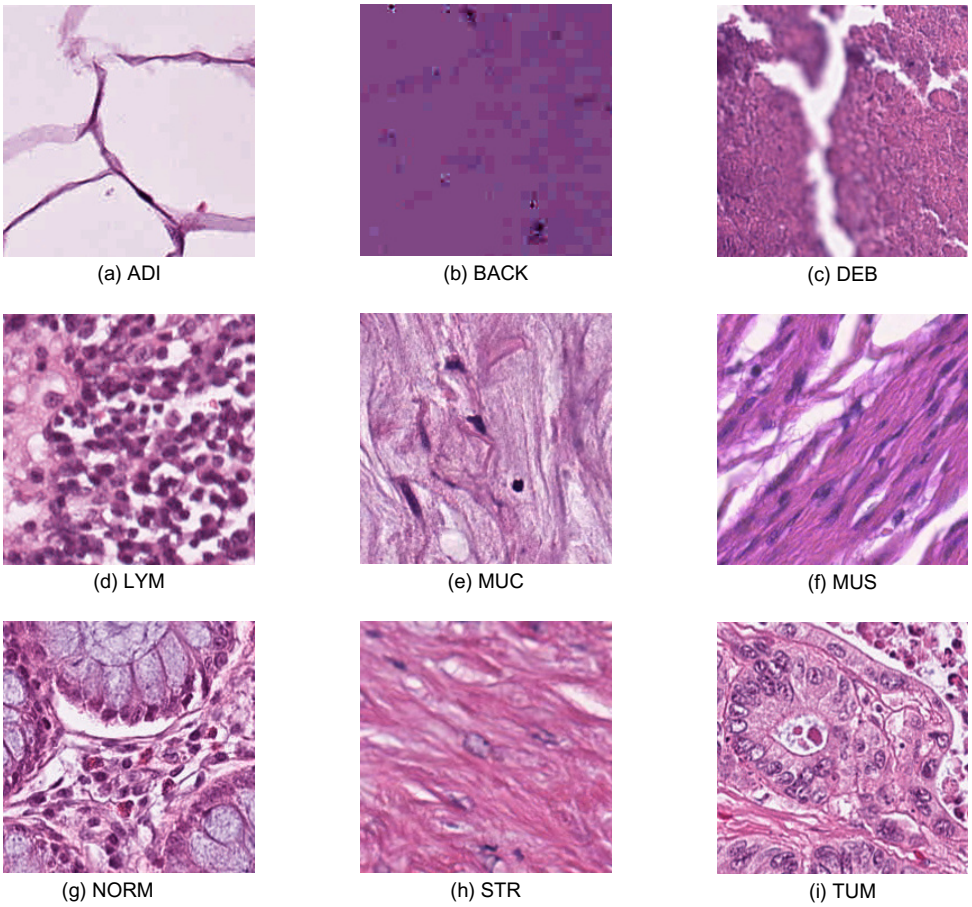


Figura 3.8: Exemplo de imagens do *dataset* NCT-CRC-HE

3.2 Sobre a organização e distribuição dos dados

Esta parte do estudo consistiu no particionamento dos conjuntos de dados. Os dados foram particionados em treinamento e teste, de modo estratificado, seguindo a proporção de 80:20, respectivamente. As únicas exceções foram os *datasets* RCC e NCT-CRC-HE. No caso do conjunto de dados RCC, o particionamento original era composto por treino, validação e teste. Desse modo, para manter a equivalência de organização das demais bases, o conjunto de validação foi combinado com o de treinamento. Para o NCT-CRC-HE, o particionamento em treino e teste já havia sido originalmente realizado pelos fornecedores, assim, a divisão original foi preservada, mesmo que esta não seguisse a proporção de 80:20. A Tabela 3.1 apresenta uma descrição detalhada da quantidade de amostras dos conjuntos após efetuar o particionamento.

Tabela 3.1: Estatísticas dos *datasets*

Dataset	Número de Classes	Treino	Teste	Total
AIDAPATH	2	1872	468	2340
PathoSpotter-HE	6	3957	990	4947
PathoSpotter-PAS	6	1912	478	2390
PathoSpotter-MultiContraste	6	8147	2037	10184
Kather	3	9581	2396	11977
RCC	5	3935	142	4077
NCT-CRC-HE-100K	9	100.000	7180	107.180

Considerando a importância de conhecer o nível de desbalanceamento dos *datasets*, realizou-se um levantamento das proporções das classes para cada um deles. O resultado é apresentado na Tabela 3.2. Observa-se que, entre as bases utilizadas, as bases PathoSpotter-HE, PathoSpotter-PAS e PathoSpotter-MultiContraste são as mais desbalanceadas. As demais, exceto a AIDPATH e a Kather, que são balanceadas, apresentam um pequeno grau de desbalanceamento.

3.3 Pré-processamento

Após a etapa de particionamento, um pipeline de pré-processamento foi conduzido. O objetivo central desta etapa foi padronizar as imagens dos conjuntos de dados, preparando-os para processamento pelos modelos auto-supervisionados avaliados.

Diante disso, foi aplicado o seguinte pipeline:

- **Redimensionamento:** Todas as imagens foram ajustadas para as dimensões de 224 x 224 *pixels*, conforme especificado pela arquitetura da rede utilizada

Tabela 3.2: Proporção das classes de cada *dataset*

Dataset	Classe	Proporção (%)
AIDAPATH	Normal	50.00
	Esclerosada	50.00
PathoSpotter-HE	Normal	17.56
	Hipercelularidade	24.99
	Membranosa primária	14.40
	Membranosa secundária	27.37
	Esclerose com membranosa	5.33
	Esclerose sem membranosa	10.34
PathoSpotter-PAS	Normal	12.24
	Hipercelularidade	26.67
	Membranosa primária	15.38
	Membranosa secundária	25.47
	Esclerose com membranosa	6.54
	Esclerose sem membranosa	13.70
PathoSpotter-MultiContraste	Normal	15.42
	Hipercelularidade	20.06
	Membranosa primária	15.78
	Membranosa secundária	31.13
	Esclerose com membranosa	5.04
	Esclerose sem membranosa	12.57
Kather	ADIMUC	33.20
	STRMUS	33.40
	TUMSTU	33.40
RCC	Grade 0	20.20
	Grade 1	20.64
	Grade 2	18.91
	Grade 3	21.35
	Grade 4	18.91
NCT-CRC-HE	ADI	10.41
	BACK	10.57
	DEB	11.51
	LYM	11.56
	MUC	8.90
	MUS	13.54
	NORM	8.76
	STR	10.45
	TUM	14.32

(descrita na seção 3.4), que foi projetada para operar com entradas desse tamanho. Este é um procedimento padrão utilizado pelas principais redes modernas, sobretudo pelas redes baseadas na arquitetura *vision transformer* (Dosovitskiy et al., 2021). Ressalta-se que o método de reamostragem LANCZOS foi utilizado para suavizar as imagens ao reduzir sua resolução, objetivando minimizar artefatos visuais como *aliasing*[4].

- **Conversão para o padrão RGB:** Converte a imagem para o modo de cor RGB, caso não esteja nesse formato, garantindo que a imagem seja processada com três canais de cor.
- **Normalização dos valores dos pixels:** Uma normalização é aplicada aos valores dos *pixels* da imagem usando médias e desvios padrão específicos para cada canal de cor (R, G, B).

3.4 Configuração Experimental

Um dos objetivos deste estudo é avaliar a eficácia de modelos baseados em aprendizado auto-supervisionado, aprimorados por meio de *fine-tuning*, aplicados ao contexto de classificação e recuperação de imagens histopatológicas. A adaptação ou *fine-tuning* refere-se ao ajuste adicional de um modelo previamente treinado em uma grande base de dados, com o objetivo de adaptá-lo a um novo conjunto de dados ou tarefa específica. Em geral, este processo é conduzido com bases de dados rotuladas. No entanto, neste estudo investigamos o potencial desse ajuste considerando bases não rotuladas. Além disso, este processo de avaliação é realizado em *datasets* pequenos de diferentes patologias ou lesões.

Para avaliar a eficácia, foi utilizado o DINOv2 nas suas versões Small (ViT/14) e Base (ViT/14). A Tabela 3.4 apresenta alguns detalhes destes modelos, como número de parâmetros e tamanho dos vetores de características. Originalmente, o DINOv2 foi pré-treinado com imagens naturais, utilizando redes do tipo ViT. Embora existam outros modelos de base treinados por meio de aprendizagem auto-supervisionada, o DINOv2 foi selecionado devido à robustez das suas representações, que tem alcançado desempenho competitivo em múltiplas tarefas, para diferentes tipos de mídia (vídeo e imagem). Além disso, é considerando como um dos modelos mais promissores do paradigma auto-supervisionado (Oquab et al., 2023).

O *fine-tuning* foi conduzido usando a implementação original do DINOv2[5]. Os pesos pré-treinados disponíveis publicamente para os *backbones* foram carregados na rede *teacher* e *student*. A adaptação é conduzida para cada um dos *datasets* utilizando o respectivo conjunto de treinamento. O conjunto de teste é empregado para avaliar a qualidade das representações extraídas dos modelos. Ressalta-se que todas as imagens utilizadas estão em nível de *patch*.

[4] Aliasing em imagens é o efeito visual de distorções, como serrilhados ou padrões indesejados, que surgem quando a resolução é insuficiente para representar detalhes finos.

[5] https://github.com/facebookresearch/dinov2

Tabela 3.3: Número de parâmetros e dimensão dos vetores de características dos modelos avaliados

Modelo	#Parâmetros	#Tamanho do Vetor
ResNet50	25.6 M	1024
RetCCL	23.5 M	2048
Phikon ViT-B	86 M	768
DINOv2 (ViT-S)	21 M	384
DINOv2 (ViT-B)	86 M	768

Em nossos experimentos, utilizou-se as mesmas configurações do pipeline de treinamento original utilizado para pré-treinamento da rede, com uma taxa de aprendizado base de 2×10^{-4}. Também avaliamos uma taxa de aprendizado de 5×10^{-4}, conforme recomendação de trabalhos anteriores (Filiot et al., 2023; Oquab et al., 2023). O número de épocas de treinamento foi definido especificamente para cada *dataset*. Para o AIDPATH, Kather, RCC e NCT-CRC-HE foram utilizadas 300 épocas, enquanto que para os conjuntos vinculados ao PathoSpotter foram utilizadas 2000 épocas. Utilizou-se um *batch* de 32 para as arquiteturas Small (ViT/14) e Base (ViT/14) e todos os *datasets*, exceto o NCT-CRC-HE para o qual aplicou-se um *batch* de 256 e 128 para as arquiteturas DINOv2 Small e Base, respectivamente. Para os conjuntos de dados associados ao PathoSpotter também foi necessário, ao invés de utilizar 8 *crops*, modificou-se para 1 *crop*. Essa alteração foi realizada devido a experimentos preliminares indicarem que o uso de um *crop* global beneficiava essas bases, considerando as especificidades das estruturas glomerulares. De fato, os glomérulos apresentam uma ampla diversidade estrutural, mesmo em tecidos saudáveis, e essa variação se intensifica em condições patológicas. Pequenas alterações na estrutura podem indicar lesões distintas, tornando a distinção entre padrões normais e anômalos uma tarefa difícil. Para o contexto deste estudo, considerar o aspecto global, mostrou-se mais benéfico que às partes individuais. Os hiperparâmetros restantes foram mantidos iguais aos valores especificados no arquivo de configuração fornecido na implementação oficial.

Os experimentos foram realizados para cada conjunto de dados, arquitetura e variação dos hiperparâmetros. Após ajustar os modelos, a etapa de extração de *features* foi então realizada. Os vetores extraídos são armazenados em uma base de dados para avaliação subsequente em tarefas de classificação e CBIR. Para a classificação, a avaliação foi conduzida usando a implementação scikit-learn[6] do *K-nearest neighbors* (KNN) com $K = 20$, utilizando o cosseno como métrica de cálculo da distância. Este valor de K foi determinado após experimentos preliminares indicarem este como o melhor hiperparâmetro, considerando os valores [1, 5, 10, 20]. O kNN foi utilizado devido seu papel valioso em contextos de aprendizado auto-supervisionado, onde o

[6]https://scikit-learn.org/1.5/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

objetivo principal comumente é verificar a qualidade do espaço latente, obtido pelos modelos. Dentre as vantagens em se utilizar o kNN, estão:

- **Avalia diretamente a qualidade do espaço latente:** O desempenho do kNN reflete a organização intrínseca das representações aprendidas, sem influências de treinamento adicional.
- **Não supervisionado para ajuste:** Não requer otimização de pesos ou treinamento adicional, tornando-o mais agnóstico em relação à tarefa.
- **Sem viés de regularização ou treinamento:** Evita enviesar os resultados devido à escolha de regularizadores, taxas de aprendizado, ou outros hiperparâmetros de treinamento.

Em relação a tarefa de CBIR, realizou-se uma busca com $K = 50$, utilizando todas as imagens do respectivo conjunto de teste dos *datasets* considerados. Todos os experimentos foram conduzidos em uma única GPU NVIDIA A100-SXM (80 GB). Os modelos adaptados neste estudo foram disponibilizados no endereço eletrônico: `https://drive.google.com/drive/folders/1wuglpSPZLSpXAZLHP08SDwiZeF4Jh4HC?usp=sharing`.

3.5 Avaliação

A análise comparativa conduzida neste estudo inclui os seguintes extratores de características: (i) O ResNet50 (He et al., 2016), pré-treinado com a coleção ImageNet; (ii) O DINOv2, nas suas versões Small e Base, que foi pré-treinado em um conjunto de imagens naturais em larga escala (Oquab et al., 2023). (iii) Extratores do estado da arte específicos de histopatologia: RetCCL (Wang et al., 2023) e Phikon ViT-B (Filiot et al., 2023); iv) Modelos propostos DINOHist-S e DINOHist-B. Apesar do ResNet50 ser um modelo mais antigo, seu uso ainda se justifica por sua ampla aplicação como extrator de características (Wang et al., 2023).

Os modelos gerados são submetidos à avaliação em duas categorias de tarefas: Classificação e CBIR. Para a tarefa de classificação foram utilizadas seis medidas de avaliação, sendo elas: *macro precision*, *macro recall*, *macro* F_1, *micro precision*, *macro recall* e *micro* F_1. Foram aplicadas tanto as medidas macro quanto as micro, dadas suas diferentes perspectivas de quantificação da eficácia. Isso é especialmente importante devido às variações nos conjuntos de dados, como tamanho, tipo de lesão ou patologia, além do desbalanceamento das classes.

Em relação à tarefa de recuperação de imagens baseada em conteúdo, a avaliação da eficácia foi conduzida a partir da $MAP@K$. Especificamente, o cálculo foi efetuado para diferentes níveis de profundidade do ranking: $MAP@5, MAP@10, MAP@20, MAP@30, MAP@40, MAP@50$. Para este cenário, o conjunto de treino foi utilizado como base de busca das imagens. Por outro lado, o conjunto de teste foi empregado para verificar a qualidade das representações

extraídas dos modelos, ou seja, corresponde ao conjunto das imagens de consulta. Cada imagem de teste foi usada individualmente como consulta. Adicionalmente, para a comparação estrita dos resultados de eficácia, os modelos desenvolvidos foram comparados usando o teste de Wilcoxon com 95% de confiança, a fim de avaliar a significância estatística. Ressalta-se que foram executadas várias consultas, portanto, o valor reportado pela $MAP@K$ corresponde a média dessas execuções. Ressalta-se que o cálculo de similaridade entre as imagens foi realizado com base na similaridade cosseno.

Capítulo 4

Resultados e Discussão

"Quanto mais aumenta nosso conhecimento, mais evidente fica nossa ignorância."

– John F. Kennedy

Este capítulo apresenta os resultados e discussões deste estudo. A Seção 4.1 aborda a eficácia dos modelos na tarefa de classificação. Por sua vez, a Seção 4.2 analisa os resultados da recuperação de imagens baseada em conteúdo.

4.1 Eficácia na Classificação de Lesões

A Tabela 4.1 apresenta a eficácia dos modelos avaliados neste estudo, considerando cada *dataset*, para as medidas *Macro Precision*, *Macro Recall* e *Macro* F_1. Esta análise inicial é conduzida do ponto de vista das medidas macro, que capturam a eficácia considerando que as classes tem pesos iguais independentemente do número de amostras de teste em cada classe. Os resultados numericamente superiores, entre todos os modelos avaliados, são destacados em negrito. A coluna "Histo" identifica se o modelo foi treinado com dados histopatológicos. De modo geral, observa-se que os modelos avaliados alcançaram resultados promissores em todas as medidas avaliadas, sobretudo os modelos auto-supervisionados, em grande parte dos *datasets*. Para a maioria dos *datasets*, os modelos alcançaram eficácia superior a 93%. Em particular, para as bases AIDPATH e Kather, os modelos DINOHist-S e Phikon ViT-B alcançaram quase 100% de eficácia. Ressalta-se que, para garantir o rigor experimental, a avaliação foi realizada em dados diferentes daqueles utilizados para o treinamento dos modelos.

Em trabalhos anteriores, considerando outros domínios, como imagens de ressonância magnética (Müller-Franzes et al., 2024), imagens da radiologia (Baharoon et al.,

2024), imagens geológicas (Brondolo e Beaussant, 2024) e detecção de anomalias na indústria (Damm et al., 2025), o DINOv2 superou as demais abordagens em diferentes cenários. Os nossos resultados também reforçam essa adaptabilidade do DINOv2 para dados histopatológicos. Embora os resultados obtidos para as bases de dados vinculadas ao PathoSpotter não tenham sido tão expressivos, o fato do *fine-tuning* ter proporcionado um melhoria frente aos demais modelos, abre caminhos para trabalhos futuros explorarem outras estratégias.

Foi verificado que os modelos pré-treinados com imagens do domínio de interesse foram mais eficazes que modelos pré-treinados com imagens naturais. Essa diferença é ainda maior para *datasets* no contexto da nefropatologia, no qual os modelos ajustados via *fine-tuning* apresentaram ganhos superiores a 30% para algumas medidas.

Também foi conduzida uma análise comparativa entre os modelos não específicos para histopatologia, ou seja, o ResNet50, o DINOv2 (ViT-S) e DINOv2 (ViT-B). Observou-se que o DINOv2 (ViT-S) e o DINOv2 (ViT-B) foram superiores numericamente ao ResNet50 para todas os *datasets*, exceto o PathoSpotter-HE. Esse resultado é indicativo do potencial das abordagens auto-supervisionadas baseadas em *Transformers* para tarefas de classificação em imagens histopatológicas, mesmo quando não foram explicitamente treinadas neste domínio. Apesar da eficácia obtida pelo DINOv2, é importante destacar a exceção observada no conjunto PathoSpotter-HE, no qual o ResNet50 apresentou eficácia competitiva. Esse resultado pode indicar limitações dos modelos *Transformers* em contextos que envolvem padrões morfológicos complexos ou estruturas teciduais homogêneas, onde a variação estrutural é limitada e menos pronunciada, como nos casos investigados.

Tomando a medida Macro F_1, o DINOv2 (ViT-B) apresentou um ganho 12.76% considerando a base RCC. Em relação à NCT-CRC-HE, este ganho é ainda maior, alcançando $15,81\%$. Tais resultados apontam para uma maior capacidade do DINOv2 (ViT-B) em capturar representações discriminativas relevantes, mesmo na presença de alta variabilidade morfológica, como variação de tamanho, forma, distribuição celular e etc. Com base neste resultado, em um cenário restrito aos modelos mencionados, os modelos DINOv2 seriam os mais adequados como extratores de características.

Considerando os conjuntos de dados utilizados, verificou-se que aqueles associados às lesões renais, foram os mais desafiadores. Para estes *datasets*, o modelo DINOHist-B foi considerado o mais eficaz. Tomando a medida Macro F_1, por exemplo, o DINOHist-B obteve 0.6664 e 0.6137 para o PathoSpotter-HE e PathoSpotter MultiContraste, respectivamente. Em relação ao PathoSpotter-PAS, a maior Macro F_1 foi alcançada pelo Phikon ViT-B. Estes resultados, apesar de estarem em patamar abaixo daqueles obtidos para as demais bases de dados, pode ser considerado promissor, dada a dificuldade de classificação de glomérulos renais e valores expressivamente maiores do que os do *baseline*. Particularmente, classificar glomérulos renais utilizando um algoritmo auto-supervisionado pode apresentar diversas dificuldades, por exemplo:

Tabela 4.1: Eficácia dos modelos avaliados neste estudo considerando as medidas *Macro Precision*, *Macro Recall* e *Macro* F_1

Dataset	**Modelo**	**Histo**	**Macro Precision**	**Macro Recall**	**Macro** F_1
AIDPATH	ResNet50		0.9561	0.9551	0.9551
	DINOv2 (ViT-S)		0.9604	0.9594	0.9594
	DINOv2 (ViT-B)		0.9604	0.9594	0.9594
	RetCCL	✓	0.9808	0.9808	0.9808
	Phikon ViT-B	✓	0.9830	0.9829	0.9829
	DINOHist-S	✓	**0.9979**	**0.9979**	**0.9979**
	DINOHist-B	✓	0.9835	0.9829	0.9829
PathoSpotter-HE	ResNet50		0.5629	0.4857	0.5019
	DINOv2 (ViT-S)		0.5489	0.4848	0.4950
	DINOv2 (ViT-B)		0.5308	0.4903	0.4955
	RetCCL	✓	0.6537	0.5911	0.6087
	Phikon ViT-B	✓	0.6672	0.6432	0.6474
	DINOHist-S	✓	0.6850	0.6272	0.6459
	DINOHist-B	✓	**0.7037**	**0.6503**	**0.6664**
PathoSpotter-PAS	ResNet50		0.4754	0.4432	0.4471
	DINOv2 (ViT-S)		0.5186	0.4874	0.4914
	DINOv2 (ViT-B)		0.5199	0.4663	0.4742
	RetCCL	✓	0.6242	0.5729	0.5895
	Phikon ViT-B	✓	**0.6569**	0.5927	**0.6111**
	DINOHist-S	✓	0.6483	**0.5962**	0.6034
	DINOHist-B	✓	0.6043	0.5602	0.5671
PathoSpotter MultiContraste	ResNet50		0.5284	0.4753	0.4876
	DINOv2 (ViT-S)		0.5460	0.4990	0.5111
	DINOv2 (ViT-B)		0.5219	0.4780	0.4863
	RetCCL	✓	0.6273	0.5822	0.5962
	Phikon ViT-B	✓	0.6390	0.5948	0.6059
	DINOHist-S	✓	0.6298	0.5868	0.5917
	DINOHist-B	✓	**0.6546**	**0.5992**	**0.6137**
Kather	ResNet50		0.9841	0.9841	0.9841
	DINOv2 (ViT-S)		0.9888	0.9887	0.9887
	DINOv2 (ViT-B)		0.9925	0.9925	0.9925
	RetCCL	✓	0.9954	0.9954	0.9954
	Phikon ViT-B	✓	**0.9987**	**0.9988**	**0.9987**
	DINOHist-S	✓	0.9938	0.9937	0.9938
	DINOHist-B	✓	0.9946	0.9946	0.9946
RCC	ResNet50		0.8090	0.7949	0.7821
	DINOv2 (ViT-S)		0.8904	0.8841	0.8800
	DINOv2 (ViT-B)		0.8893	0.8854	0.8819
	RetCCL	✓	0.9357	0.9332	0.9291
	Phikon ViT-B	✓	0.9429	**0.9409**	0.9397
	DINOHist-S	✓	**0.9452**	**0.9409**	**0.9403**
	DINOHist-B	✓	0.9360	0.9335	0.9325
NCT-CRC-HE	ResNet50		0.7834	0.7841	0.7729
	DINOv2 (ViT-S)		0.8756	0.8873	0.8777
	DINOv2 (ViT-B)		0.8911	0.9008	0.8951
	RetCCL	✓	0.9035	0.9048	0.9025
	Phikon ViT-B	✓	**0.9322**	**0.9332**	**0.9326**
	DINOHist-S	✓	0.9147	0.9211	0.9176
	DINOHist-B	✓	0.9204	0.9233	0.9216

- **Variabilidade Morfológica:** Os glomérulos renais possuem formas e tamanhos variáveis dependendo do estado de saúde do tecido (ex. normal, inflamação ou esclerose). Essa heterogeneidade pode dificultar a generalização do modelo.
- **Sutileza nas Alterações Patológicas:** Diferenças visuais entre glomérulos normais e alterados (como fibrose ou presença de células inflamatórias) podem ser extremamente sutis e não facilmente capturadas por representações aprendidas automaticamente (Yao et al., 2022; Hemmatirad et al., 2023).
- **Complexidade do Contexto Tissular:** A classificação pode ser influenciada pelo microambiente ao redor dos glomérulos, incluindo células vizinhas e estruturas que podem ser confundidas, dificultando a extração precisa de características relevantes.

Além das dificuldades mencionadas, existem outros fatores que podem ter exercido influência. Um desses fatores está associado a similaridade entre imagens de algumas classes diferentes, por exemplo, a classe membranosa primária e membranosa secundária. A Figura 4.1 apresenta um exemplo em que é possível perceber, visualmente, a similaridade entre ambas as classes. Dada a semelhança entre ambas, o modelo gera muitos falsos positivos, ou seja, indica que é membranosa primária em vez de secundária, e vice-versa. A Figura 4.2 apresenta a matriz de confusão do *dataset* PathoSpotter-HE, considerando o modelo DINOHist-B. De fato, ao analisar a matriz de confusão gerada na fase experimental, verificou-se que esse comportamento era comum. Além disso, outras classes seguem esse mesmo comportamento. Por exemplo, muitas imagens da classe esclerose com membranosa foram incorretamente previstas como da classe hipercelularidade.

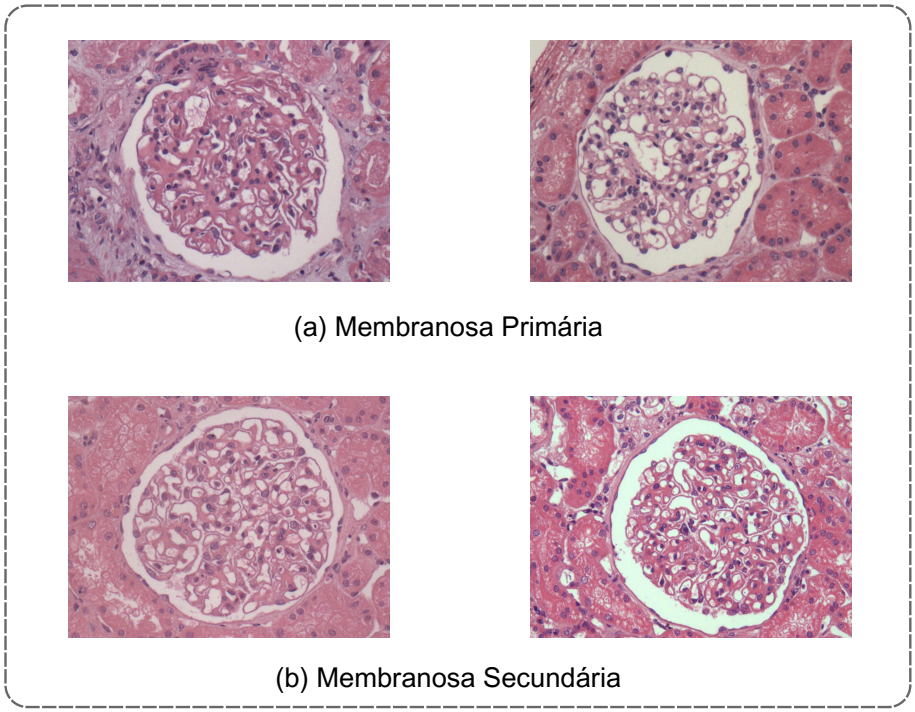


Figura 4.1: Amostras de imagens da classe membranosa primária e membranosa secundária do *dataset* PathoSpotter-HE

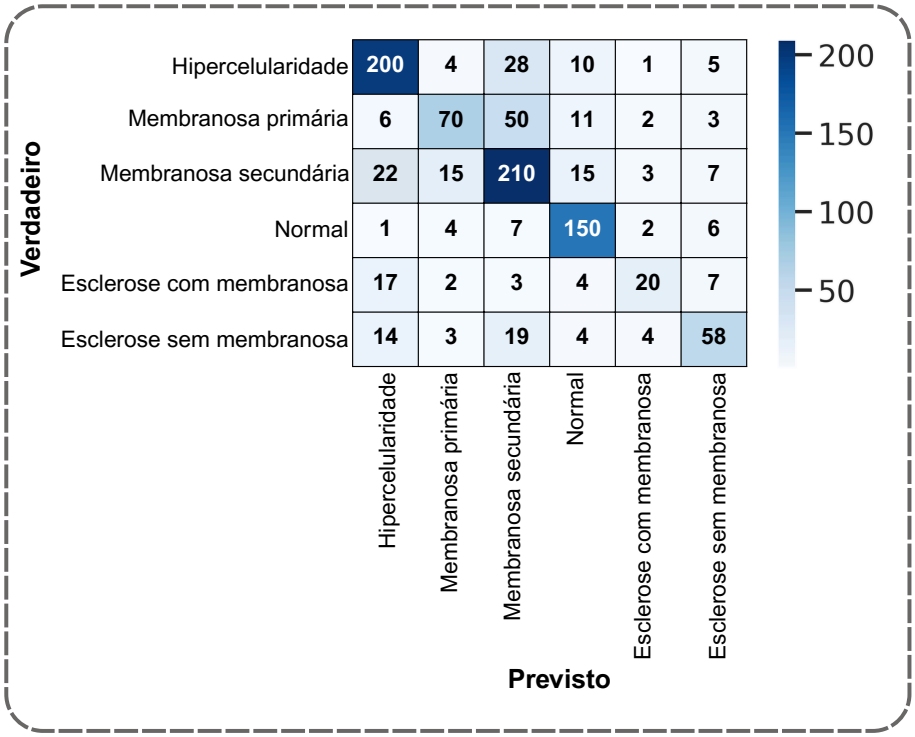


Figura 4.2: Matriz de confusão do *dataset* PathoSpotter-HE

Outra característica que pode ter influenciado refere-se ao redimensionamento. Diferentemente das imagens pertencentes aos outros conjuntos de dados, as imagens das lesões renais possuíam uma maior resolução. Assim, ao reduzir a resolução é possível que isso tenha afetado a capacidade de manutenção de detalhes importantes dos glomérulos, mesmo tendo sido aplicada uma técnica de amostragem considerada eficaz e comum na literatura. Por fim, os hiperparâmetros originalmente recomendados do DINOv2 podem não ter sido adequados para este domínio. Isto foi parcialmente verificado para um dos hiperparâmetros. Observou-se que o número de *crops* exercia uma influência direta para as bases do PathoSpotter. Ao manter o valor originalmente recomendado no arquivo de configuração, a eficácia do modelo não evoluía. Ao alterar o número de *crops*, o efeito na eficácia do modelo foi imediato. Deste modo, é preciso estender essas análises para investigar outras possíveis causas e, com isso, promover ações para aperfeiçoar os modelos. Neste sentido, direcionamentos de pesquisa podem ser encontrados na Seção 5.1.

De modo geral, o DINOHist-S e DINOHist-B alcançaram o melhor desempenho dentre os modelos avaliados. Este resultado é de grande relevância, tendo em vista que esses modelos foram ajustados usando uma única GPU a partir de *datasets* pequenos. Além disso, indica que a adaptação dos modelos de base em dados de tarefas específicas pode superar os modelos específicos de domínio treinados em larga escala, necessitando apenas de uma fração dos recursos e do tempo de treinamento. Por exemplo, considerando o *dataset* RCC, o modelo DINOHist-S levou aproximadamente 40 minutos de treinamento. Comparado ao Phikon ViT-B, foi utilizado apenas 0.05% (40 minutos vs. 1.216 horas) de horas de GPU para atingir os resultados apresentados na Tabela 4.1. Este resultado é de grande relevância, pois destaca o potencial da aplicação dos modelos auto-supervisionados em cenários em que há escassez de dados rotulados e o acesso aos recursos computacionais são limitados.

A Tabela 4.2 apresenta a eficácia dos modelos avaliados levando em consideração as medidas *Micro Precision*, *Micro Recall* e *Micro* F_1. Com estas medidas, espera-se avaliar o desempenho global do modelo, uma vez que no cálculo são consideradas todas as instâncias das classes conjuntamente. Os resultados numericamente superiores são destacados em negrito. Semelhante ao verificado para as medidas Macro, observa-se que os modelos avaliados alcançaram resultados promissores para todas medidas aplicadas, especialmente para os modelos auto-supervisionados. Para a maioria dos *datasets*, os modelos alcançaram eficácia acima de 94%. No entanto, a aplicação direta dos modelos, sem a etapa de *fine-tuning*, resultou em desempenho limitado nos conjuntos de dados do PathoSpotter. Por outro lado, ao aplicar o *fine-tuning*, observou-se uma melhoria significativa na eficácia desses modelos. Para estas bases, os melhores resultados foram obtidos pelo DINOHist-S e pelo DINOHist-B. Por exemplo, considerando o PathoSpotter-HE, o DINOHist-B Micro F_1 de 0.7182, superando os demais modelos bases. Este resultado reforça a importância do *fine-tuning* em cenários desafiadores, como a classificação de lesões glomerulares. Além disso, indica o potencial do aprendizado auto-supervisionado em cenários com dados reduzidos e de domínio distinto do modelo base.

Tabela 4.2: Eficácia dos modelos avaliados neste estudo considerando as medidas *Micro Precision*, *Micro Recall* e *Micro* F_1

Dataset	**Modelo**	**Histo**	**Micro Precision**	**Micro Recall**	**Micro** F_1
AIDPATH	ResNet50		0.9551	0.9551	0.9551
	DINOv2 (ViT-S)		0.9594	0.9594	0.9594
	DINOv2 (ViT-B)		0.9594	0.9594	0.9594
	RetCCL	✓	0.9808	0.9808	0.9808
	Phikon ViT-B	✓	0.9829	0.9829	0.9829
	DINOHist-S	✓	**0.9979**	**0.9979**	**0.9979**
	DINOHist-B	✓	0.9829	0.9829	0.9829
PathoSpotter-HE	ResNet50		0.5515	0.5515	0.5515
	DINOv2 (ViT-S)		0.5535	0.5535	0.5535
	DINOv2 (ViT-B)		0.5485	0.5485	0.5485
	RetCCL	✓	0.6636	0.6636	0.6636
	Phikon ViT-B	✓	0.6798	0.6798	0.6798
	DINOHist-S	✓	0.7030	0.7030	0.7030
	DINOHist-B	✓	**0.7182**	**0.7182**	**0.7182**
PathoSpotter-PAS	ResNet50		0.5084	0.5084	0.5084
	DINOv2 (ViT-S)		0.5460	0.5460	0.5460
	DINOv2 (ViT-B)		0.5314	0.5314	0.5314
	RetCCL	✓	0.6213	0.6213	0.6213
	Phikon ViT-B	✓	0.6234	0.6234	0.6234
	DINOHist-S	✓	**0.6548**	**0.6548**	**0.6548**
	DINOHist-B	✓	0.6192	0.6192	0.6192
PathoSpotter MultiContraste	ResNet50		0.5376	0.5376	0.5376
	DINOv2 (ViT-S)		0.5616	0.5616	0.5616
	DINOv2 (ViT-B)		0.5415	0.5415	0.5415
	RetCCL	✓	0.6446	0.6446	0.6446
	Phikon ViT-B	✓	0.6456	0.6456	0.6456
	DINOHist-S	✓	0.6559	0.6559	0.6559
	DINOHist-B	✓	**0.6583**	**0.6583**	**0.6583**
Kather	ResNet50		0.9841	0.9841	0.9841
	DINOv2 (ViT-S)		0.9887	0.9887	0.9887
	DINOv2 (ViT-B)		0.9925	0.9925	0.9925
	RetCCL	✓	0.9954	0.9954	0.9954
	Phikon ViT-B	✓	**0.9987**	**0.9987**	**0.9987**
	DINOHist-S	✓	0.9937	0.9937	0.9937
	DINOHist-B	✓	0.9946	0.9946	0.9946
RCC	ResNet50		0.7986	0.7986	0.7986
	DINOv2 (ViT-S)		0.8921	0.8921	0.8921
	DINOv2 (ViT-B)		0.8921	0.8921	0.8921
	RetCCL	✓	0.9353	0.9353	0.9353
	Phikon ViT-B	✓	**0.9496**	**0.9496**	**0.9496**
	DINOHist-S	✓	**0.9496**	**0.9496**	**0.9496**
	DINOHist-B	✓	0.9424	0.9424	0.9424
NCT-CRC-HE	ResNet50		0.8121	0.8121	0.8121
	DINOv2 (ViT-S)		0.9015	0.9015	0.9015
	DINOv2 (ViT-B)		0.9174	0.9174	0.9174
	RetCCL	✓	0.9262	0.9262	0.9262
	Phikon ViT-B	✓	**0.9565**	**0.9565**	**0.9565**
	DINOHist-S	✓	0.9412	0.9412	0.9412
	DINOHist-B	✓	0.9464	0.9464	0.9464

De modo geral, os modelos DINOv2 submetidos ao *fine-tuning* foram os que apresentaram os melhores resultados. Em alguns cenários, como no caso da base NCT-CRC-HE, as variações do DINOv2 apresentaram resultados numericamente inferiores em comparação ao Phikon ViT-B. Contudo, é importante destacar que seu treinamento foi mais eficiente, alcançando resultados promissores, mesmo com uma base de dados significativamente menor (433 vezes menor) do que a utilizada pelo modelo mencionado.

Considerando as medidas macro e micro, observou-se que os melhores resultados foram consistentes para ambas. Tanto para as medidas macro quanto para as medidas micro, os melhores modelos foram os mesmos, exceto para o *dataset* PathoSpotter-PAS. Para este conjunto, há uma diferença entre o melhor modelo apontado. Nas medidas macro, o Phikon ViT-B é superior em *Macro Precision* e *Macro* F_1, enquanto o DINOHist-S é superior em *Macro Recall*. Por outro lado, para as medidas micro, o DINOHist-S obteve os melhores resultados para todas as medidas. Esse resultado reforça a importância do *fine-tuning* na adaptação de modelos a domínios distintos daquele utilizado no pré-treinamento. Embora os autores do DINOv2 (Oquab et al., 2023) afirmem que o modelo pode ser aplicado diretamente, sem necessidade de *fine-tuning*, nossos experimentos demonstraram que, em alguns conjuntos de dados, o desempenho é limitado. No entanto, ao aplicar o *fine-tuning*, observou-se uma melhoria consistente, evidenciada por ambas as medidas.

4.2 Eficácia Considerando a Tarefa de CBIR

A Tabela 4.3 apresenta a eficácia dos modelos avaliados neste estudo, considerando diferentes níveis de profundidade do *ranking* (MAP@5, MAP@10, MAP@20, MAP@30, MAP@40, MAP@50). Os resultados apresentados na Tabela 4.3 indicam que o modelo Phikon ViT-B supera os demais em vários níveis de *ranking*, para a maioria dos *datasets*. Verifica-se que, mesmo aqueles modelos pré-treinados com imagens naturais, alcançam resultados acima de 90% na maioria dos *datasets*. Em relação as bases de dados vinculadas ao PathoSpotter, os modelos DINOHist-S e DINOHist-B foram os mais eficazes. Esse resultado indica que, para determinadas lesões ou patologias em alguns órgãos, aplicar o *fine-tuning* para adaptar o modelo especificamente ao domínio de aplicação foi a melhor opção para melhoria de eficácia. A superioridade proporcionada pelo *fine-tuning* torna-se ainda mais evidente nos conjuntos de dados do PathoSpotter. Nesses casos, os modelos DINOHist-S e DINOHist-B apresentaram ganhos superiores a 20% em alguns níveis de ranqueamento, quando comparados às suas respectivas variantes sem *fine-tuning*.

Tabela 4.3: Eficácia dos modelos considerando a tarefa de CBIR.

Dataset	Modelo	Histo	MAP@5	MAP@10	MAP@20	MAP@30	MAP@40	MAP@50
AIDPATH	ResNet50		0.9420	0.9001	0.8519	0.8110	0.7842	0.7645
	DINOv2 (ViT-S)		0.9769	0.9531	0.9192	0.8924	0.8805	0.8644
	DINOv2 (ViT-B)		0.9810	0.9572	0.9158	0.8900	0.8664	0.8448
	RetCCL	✓	0.9737	0.9657	0.9368	0.9245	0.9145	0.8983
	Phikon ViT-B	✓	0.9867	**0.9771**	**0.9478**	**0.9330**	**0.9218**	**0.9088**
	DINOHist-S	✓	**0.9899**	0.9742	0.9437	0.9254	0.9050	0.8866
	DINOHist-B	✓	0.9889	0.9701	0.9414	0.9249	0.9058	0.8926
PathoSpotter-HE	ResNet50		0.5571	0.4933	0.4341	0.3938	0.3711	0.3606
	DINOv2 (ViT-S)		0.5608	0.4845	0.4292	0.3919	0.3792	0.3602
	DINOv2 (ViT-B)		0.5460	0.4982	0.4427	0.4070	0.3846	0.3696
	RetCCL	✓	0.6804	0.5865	0.5149	0.4734	0.4428	0.4235
	Phikon ViT-B	✓	**0.7257**	0.6206	0.5148	0.4654	0.4269	0.4044
	DINOHist-S	✓	0.7135	0.6320	**0.5685**	0.5136	0.4762	0.4469
	DINOHist-B	✓	0.7131	**0.6369**	0.5623	**0.5231**	**0.4856**	**0.4577**
PathoSpotter-PAS	ResNet50		0.5366	0.4873	0.4319	0.3954	0.3767	0.3574
	DINOv2 (ViT-S)		0.5527	0.4933	0.4203	0.3868	0.3582	0.3369
	DINOv2 (ViT-B)		0.5455	0.5009	0.4231	0.3935	0.3688	0.3488
	RetCCL	✓	0.6224	0.5378	0.4656	0.4219	**0.3911**	**0.3815**
	Phikon ViT-B	✓	0.6532	0.5528	0.4608	0.4126	0.3909	0.3689
	DINOHist-S	✓	**0.6718**	**0.5678**	0.4714	0.4211	0.3810	0.3490
	DINOHist-B	✓	0.6589	0.5826	**0.4886**	**0.4289**	0.3900	0.3652
PathoSpotter MultiContraste	ResNet50		0.5390	0.4815	0.4271	0.3924	0.3736	0.3571
	DINOv2 (ViT-S)		0.5490	0.4996	0.4371	0.4099	0.3799	0.3621
	DINOv2 (ViT-B)		0.5576	0.4880	0.4286	0.3989	0.3797	0.3641
	RetCCL	✓	0.6490	0.5595	0.4837	0.4438	0.4151	0.3990
	Phikon ViT-B	✓	**0.6839**	0.5828	0.4929	0.4427	0.4141	0.3912
	DINOHist-S	✓	0.6599	0.5919	0.5235	**0.4948**	**0.4711**	**0.4515**
	DINOHist-B	✓	0.6720	**0.5977**	**0.5315**	0.4946	0.4674	0.4444
Kather	ResNet50		0.9825	0.9759	0.9684	0.9622	0.9577	0.9540
	DINOv2 (ViT-S)		0.9913	0.9840	0.9764	0.9703	0.9664	0.9616
	DINOv2 (ViT-B)		0.9915	0.9846	0.9783	0.9716	0.9666	0.9630
	RetCCL	✓	0.9977	0.9962	0.9937	0.9922	0.9906	0.9895
	Phikon ViT-B	✓	**0.9989**	**0.9985**	**0.9958**	**0.9937**	**0.9925**	**0.9912**
	DINOHist-S	✓	0.9842	0.9769	0.9644	0.9547	0.9464	0.9403
	DINOHist-B	✓	0.9821	0.9721	0.9586	0.9498	0.9419	0.9362
RCC	ResNet50		0.8065	0.7706	0.7408	0.7104	0.6965	0.6785
	DINOv2 (ViT-S)		0.8735	0.8735	0.8291	0.8022	0.7785	0.7462
	DINOv2 (ViT-B)		0.8651	0.8651	0.8107	0.7850	0.7611	0.7442
	RetCCL	✓	0.9178	0.9178	0.9022	0.8802	0.8571	0.8385
	Phikon ViT-B	✓	**0.9425**	**0.9425**	**0.9443**	**0.9399**	**0.9301**	**0.9296**
	DINOHist-S	✓	0.9210	0.9191	0.8723	0.8460	0.8111	0.8004
	DINOHist-B	✓	0.9009	0.9009	0.8505	0.8167	0.7925	0.7638
NCT-CRC-HE	ResNet50		0.8257	0.8084	0.7881	0.7752	0.7667	0.7600
	DINOv2 (ViT-S)		0.8935	0.8813	0.8660	0.8562	0.8501	0.8446
	DINOv2 (ViT-B)		0.9110	0.8973	0.8790	0.8697	0.8625	0.8573
	RetCCL	✓	0.9197	0.9154	0.9055	0.9002	0.8980	0.8951
	Phikon ViT-B	✓	**0.9593**	**0.9564**	**0.9556**	**0.9544**	**0.9530**	**0.9531**
	DINOHist-S	✓	0.9046	0.8869	0.8668	0.8532	0.8430	0.8354
	DINOHist-B	✓	0.9070	0.8902	0.8696	0.8568	0.8477	0.8409

A Tabela 4.4 apresenta os resultados do teste estatístico de Wilcoxon, com 95% de confiança, onde comparou-se os modelos histopatológicos com o melhor modelo pré-treinado com imagens naturais (DINOv2 (ViT-B)). As células em verde indicam superioridade estatística do modelo adaptado ao domínio histopatológico, enquanto as rosas indicam inferioridade. Em branco estão as células onde os modelos foram considerados estatisticamente equivalentes. Observou-se que os modelos RetCCL e Phikon ViT-B foram consistentemente superiores ou equivalentes ao DINOv2 (ViT-B) considerando todas as bases de dados. O DINOHist-S e o DINOHist-B também foram estatisticamente superiores ou equivalentes em cinco das sete bases utilizadas. As únicas exceções foram para os *datasets* Kather e NCT-CRC-HE, ambos associados ao câncer colorretal. Este resultado indica a necessidade de realizar novas avaliações, que investiguem outras possibilidades de otimização de modelos histopatológicos.

A Tabela 4.5 apresenta os resultados do teste estatístico, comparando o Phikon ViT-B com os demais modelos histopatológicos. De modo geral, o Phikon ViT-B foi estatisticamente superior ou equivalente aos outros modelos para a maioria dos *datasets*. As únicas exceções foram os *datasets* vinculadas ao PathoSpotter. Para estas bases, o Phikon ViT-B foi estatisticamente inferior ao DINOHist-S e ao DINOHist-B para vários níveis de ranking, sobretudo, considerando as bases PathoSpotter-HE e PathoSpotter MultiContraste. Este é um resultado importante, pois demonstra o potencial dos modelos DINOHist em cenários desafiadores, como o caso de dados oriundos da nefropatologia

De modo geral, todos os modelos alcançaram resultados promissores, especialmente os auto-supervisionados Phikon ViT-B, DINOHist-S e DINOHist-B. Este resultado indica que esses modelos são mais aptos a capturar e recuperar as características sutis das imagens de tecidos, um aspecto crucial na identificação de padrões patológicos específicos. Essa capacidade pode permitir potencializar o auxílio ao diagnóstico, como também pode reduzir o tempo de análise, minimizar erros diagnósticos e, consequentemente, melhorar o atendimento ao paciente. Esses resultados reforçam a importância do desenvolvimento e evolução de modelos avançados de aprendizado de máquina em aplicações médicas, onde a eficácia e a confiabilidade são essenciais. Além disso, demonstram o potencial de modelos auto-supervisionados, mesmo pré-treinados com imagens naturais, em se adaptarem a outros domínios.

As Figuras (4.2 a 4.7) apresentam exemplos ilustrativos comparativos de busca efetuada para cada um dos *datasets* usando o DINOv2 (ViT-B) e o melhor modelo histopatológico para o *dataset* de exemplo. Foram selecionadas consultas que destacam a melhoria proporcionada pelos modelos adaptados. São apresentadas as dez primeiras imagens recuperadas pelo sistema. Na parte superior são apresentadas as imagens recuperadas a partir das *features* extraídas pelo DINOv2 (ViT-B), enquanto na parte inferior são apresentadas as imagens para o melhor modelo histopatológico encontrado para o *dataset* em questão. Imagens consideradas irrelevantes à imagem de consulta especificada são destacadas em vermelho. Como pode ser observado, para os cenários exemplificados, os modelos histopatológicos são altamente eficazes.

Tabela 4.4: Resultado do teste estatístico de Wilcoxon, com 95% de confiança, comparando os modelos histopatológicos com o melhor modelo pré-treinado com imagens naturais. As células em verde indicam superioridade estatística, enquanto as rosas indicam inferioridade. Em branco estão as células onde as diferenças entre os modelos são estatisticamente insignificantes

Dataset	Comparativo	MAP@5	MAP@10	MAP@20	MAP@30	MAP@40	MAP@50
AIDPATH	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						
PathoSpotter-HE	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						
PathoSpotter-PAS	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						
PathoSpotter MultiContraste	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						
Kather	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						
RCC	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						
NCT-CRC-HE	RetCCL vs DINOv2 (ViT-B)						
	Phikon ViT-B vs DINOv2 (ViT-B)						
	DINOHist-S vs DINOv2 (ViT-B)						
	DINOHist-B vs DINOv2 (ViT-B)						

Tabela 4.5: Resultado do teste estatístico de Wilcoxon, com 95% de confiança, comparando o Phikon ViT-B com os demais modelos histopatológicos. As células em verde indicam superioridade estatística, enquanto as rosas indicam inferioridade. Em branco estão as células onde as diferenças entre os modelos são estatisticamente insignificantes

Dataset	Comparativo	MAP@5	MAP@10	MAP@20	MAP@30	MAP@40	MAP@50
AIDPATH	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						
PathoSpotter-HE	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						
PathoSpotter-PAS	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						
PathoSpotter MultiContraste	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						
Kather	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						
RCC	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						
NCT-CRC-HE	Phikon ViT-B vs RetCCL						
	Phikon ViT-B vs DINOHist-S						
	Phikon ViT-B vs DINOHist-B						

Para estes modelos, nota-se que a primeira imagem exibida é sempre uma imagem relevante, isso para todos os *datasets* considerados. Ou seja, a imagem exibida pertence à classe de interesse. Em contrapartida, o DINOv2 (ViT-B) apresenta um alto número de imagens não relevantes para o mesmo cenário de avaliação. Por exemplo, tomando o PathoSpotter-HE, o DINOv2 (ViT-B) recuperou apenas uma imagem da classe de interesse, enquanto a sua versão ajustada recuperou nove imagens relevantes.

De modo geral, os resultados alcançados neste estudo permitiram demonstrar o potencial das arquiteturas auto-supervisionadas. Este resultado é de grande importância, especialmente considerando conjunto de dados reduzidos, assim como a dificuldade de obtenção de dados anotados. Esses problemas evidenciam a importância deste estudo, ao explorar o uso de modelos auto-supervisionados para enfrentá-los, buscando impulsionar o uso prático de tais recursos na prática clínica. Afinal, na prática clínica, a capacidade de identificar rapidamente imagens histopatológicas semelhantes a partir de grandes bancos de dados pode auxiliar os patologistas na comparação de casos, diagnóstico preciso e tomada de decisão informada.

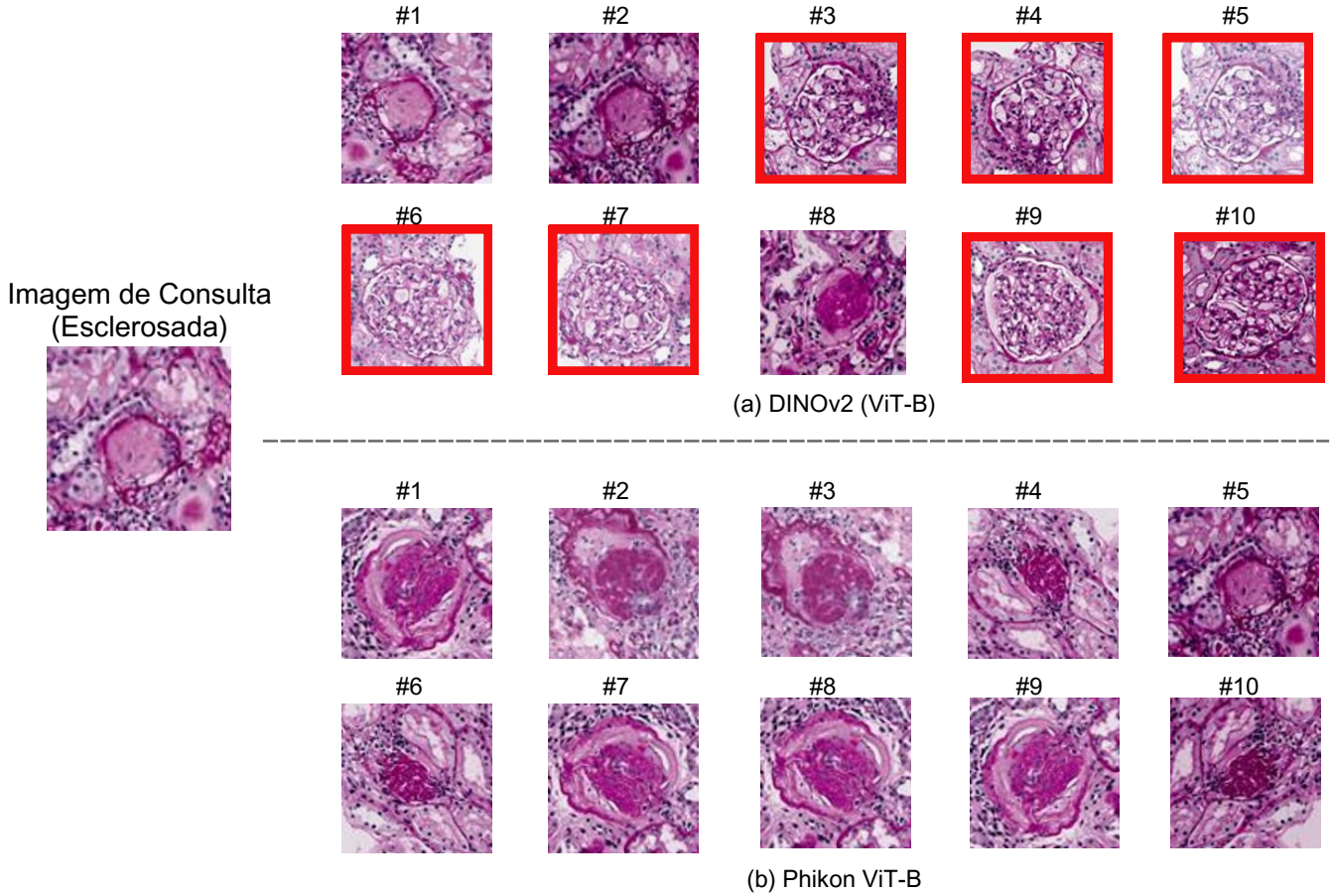


Figura 4.3: Top-10 resultados para a imagem de consulta especificada do dataset AIDPATH. No topo temos o resultado do DINOv2 (ViT-B) (MAP@10 = 0.2852). Na parte inferior o resultado do Phikon ViT-B (MAP@10 = 1.0000. Imagens não relevantes (que não apresentam a mesma lesão presente na imagem de consulta) são destacadas em vermelho.

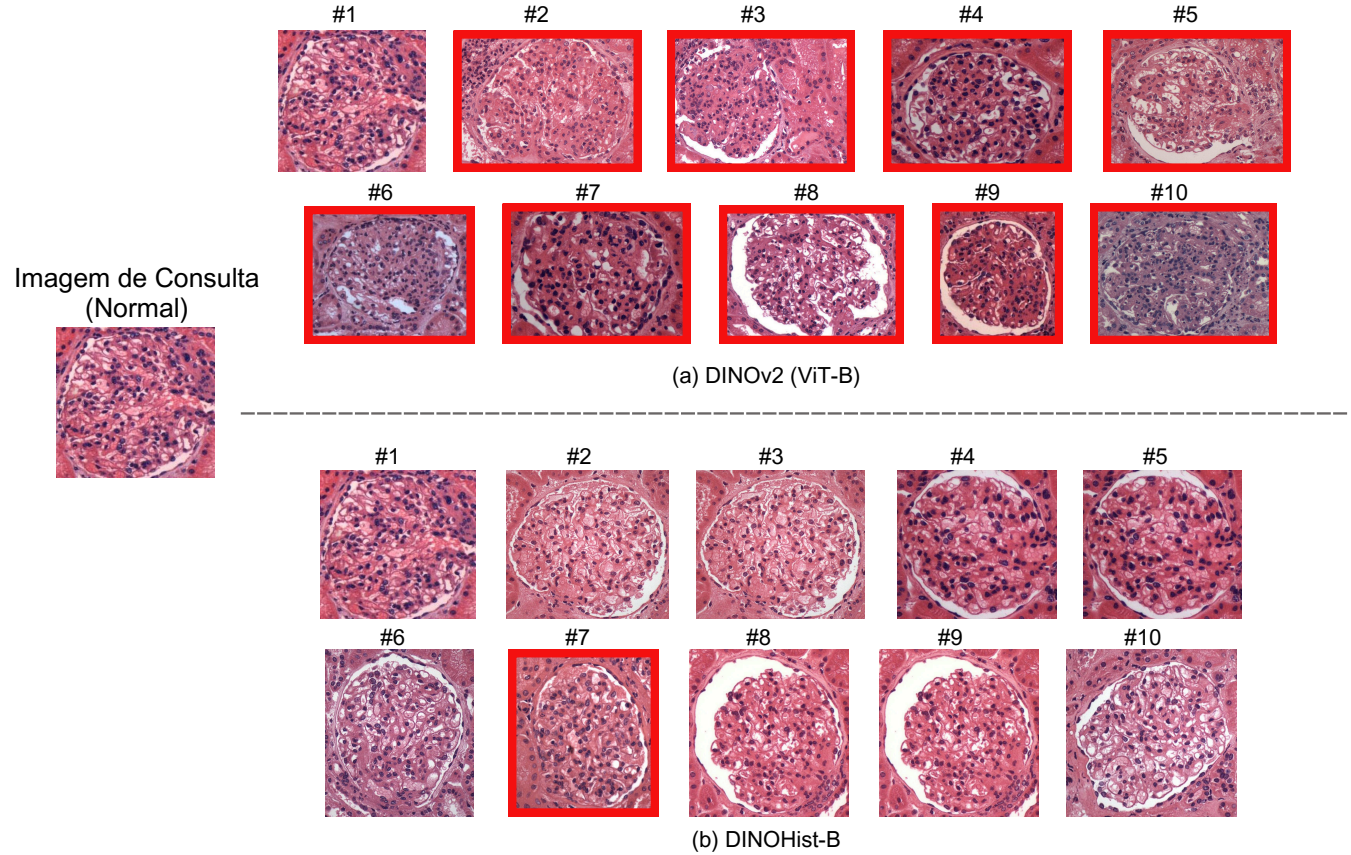


Figura 4.4: Top-10 resultados para a imagem de consulta especificada do dataset PathoSpotter-HE. No topo temos o resultado do DINOv2 (ViT-B) (MAP@10 = 0.1000). Na parte inferior o resultado do DINOHist-B (MAP@10 = 0.9060. Imagens não relevantes são destacadas em vermelho

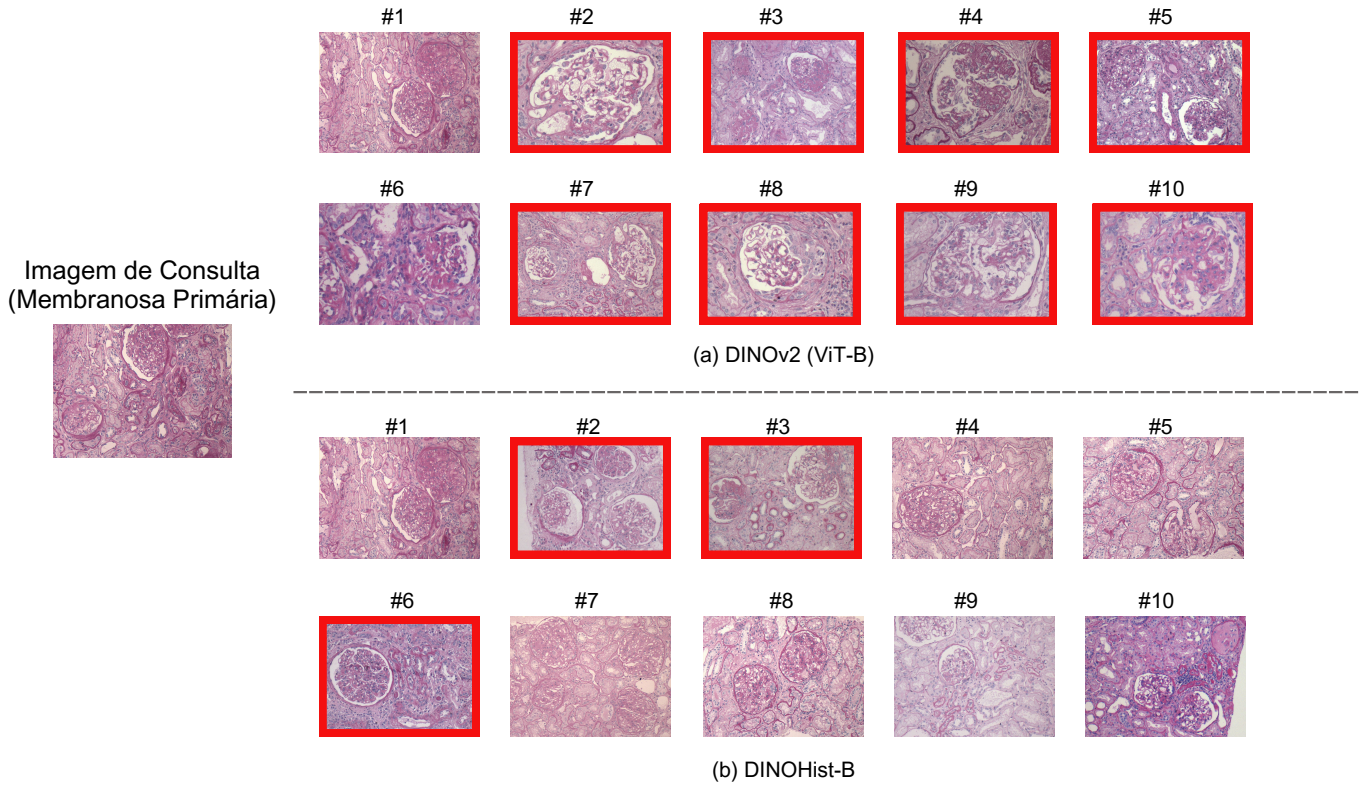


Figura 4.5: Top-10 resultados para a imagem de consulta especificada do dataset PathoSpotter-PAS. No topo temos o resultado do DINOv2 (ViT-B) (MAP@10 = 0.2000). Na parte inferior o resultado do DINOHist-B (MAP@10 = 0.9129. Imagens não relevantes são destacadas em vermelho

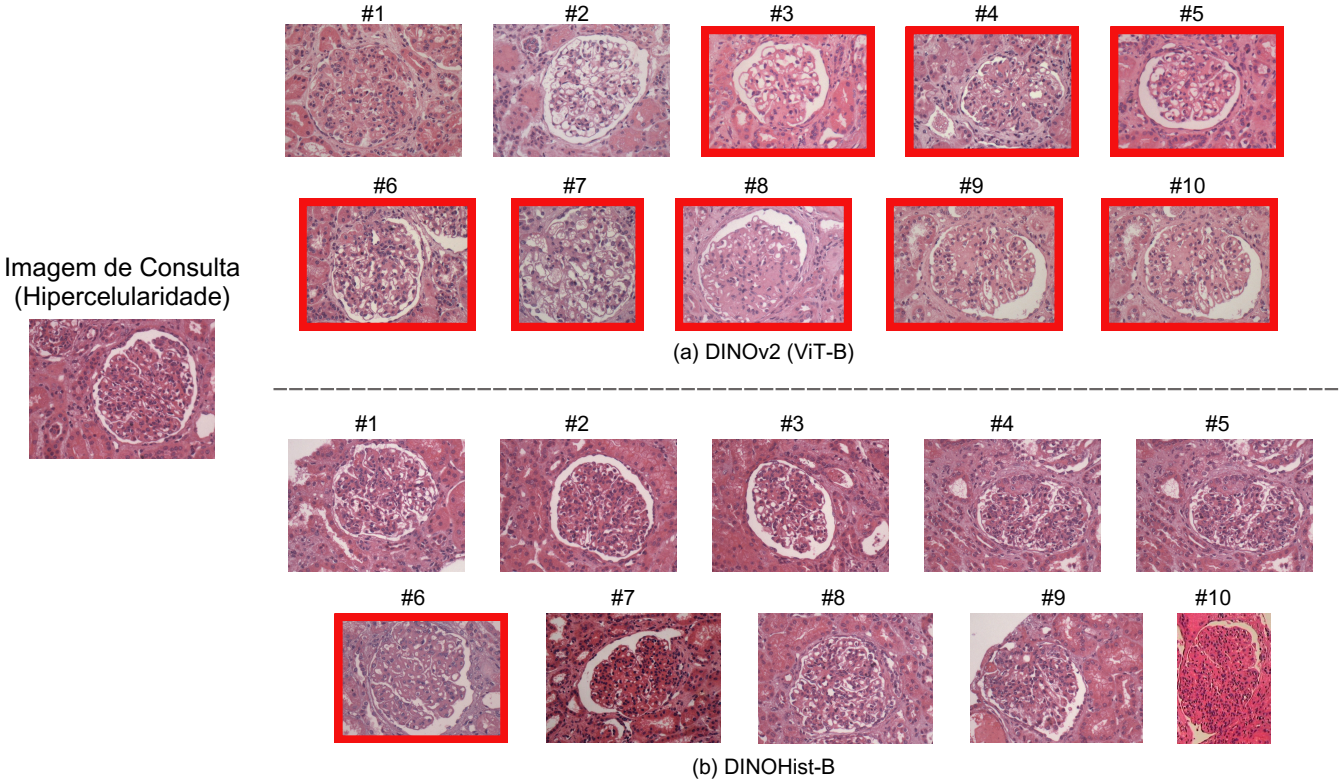


Figura 4.6: Top-10 resultados para a imagem de consulta especificada do dataset PathoSpotter-MultiContraste. No topo temos o resultado do DINOv2 (ViT-B) (MAP@10 = 0.1556). Na parte inferior o resultado do DINOHist-B (MAP@10 = 0.9283. Imagens não relevantes são destacadas em vermelho

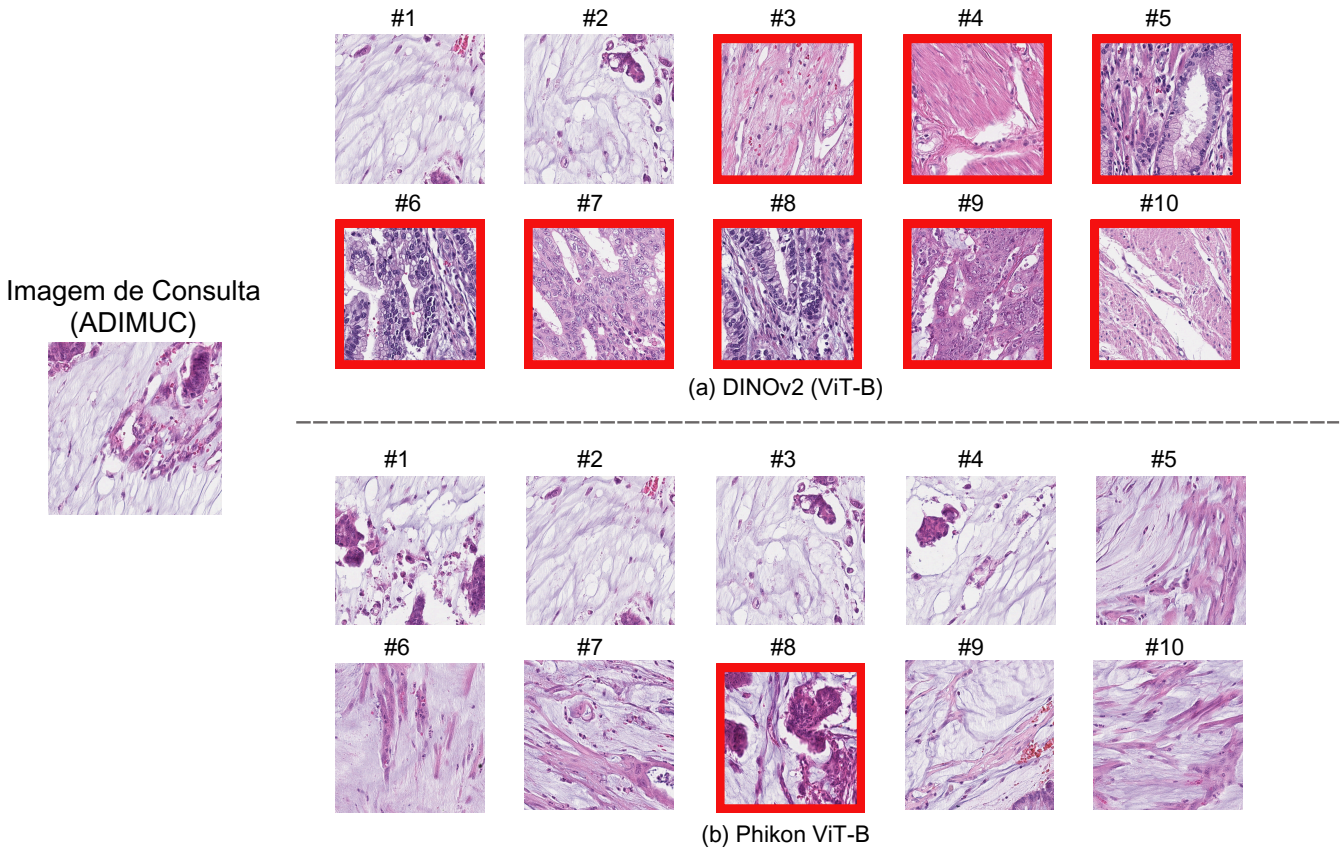


Figura 4.7: Top-10 resultados para a imagem de consulta especificada do dataset Kather. No topo temos o resultado do DINOv2 (ViT-B)(MAP@10 = 0.1556). Na parte inferior o resultado do Phikon ViT-B (MAP@10 = 0.8783. Imagens não relevantes são destacadas em vermelho

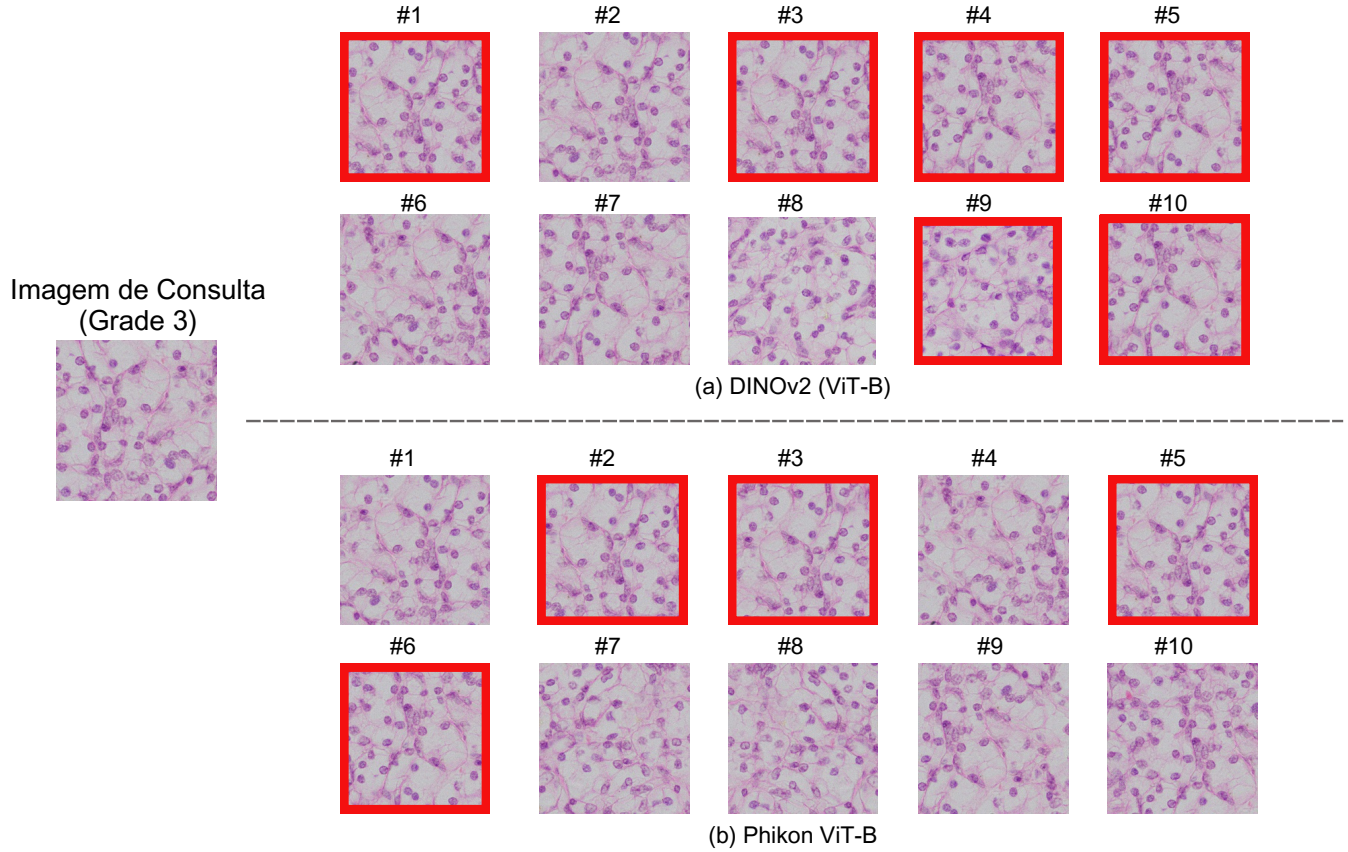


Figura 4.8: Top-10 resultados para a imagem de consulta especificada do dataset RCC. No topo temos o resultado do DINOv2 (ViT-B)(MAP@10 = 0.4694). Na parte inferior o resultado do Phikon ViT-B (MAP@10 = 0.8857. Imagens não relevantes são destacadas em vermelho

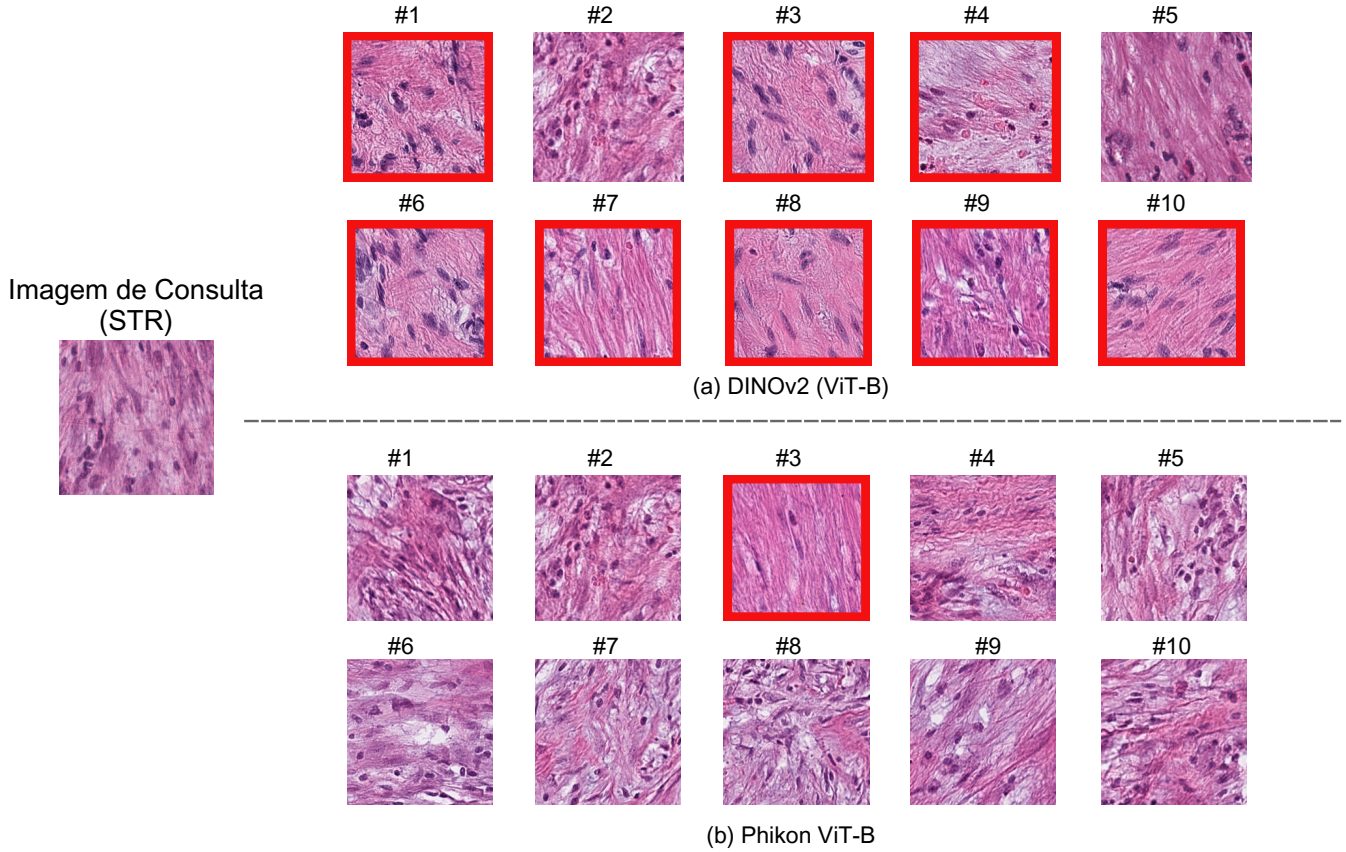


Figura 4.9: Top-10 resultados para a imagem de consulta especificada do dataset NCT-CRC-HE. No topo temos o resultado do DINOv2 (ViT-B)(MAP@10 = 0.1944). Na parte inferior o resultado do Phikon ViT-B (MAP@10 = 0.9468. Imagens não relevantes são destacadas em vermelho

Capítulo 5

Conclusões

"O importante é não parar de questionar. A curiosidade tem sua própria razão de existir."
– Albert Einstein

A histopatologia está associada ao estudo e análise de células e tecidos, em geral com o auxílio de um microscópio. Esse processo permite ao patologista realizar diagnóstico mediante amostras de tecido. Embora comum, essa avaliação manual é trabalhosa, demorada, subjetiva e suscetível a erros. Para reduzir a carga de trabalho e melhorar a análise, sistemas de diagnóstico auxiliados por computador podem ser utilizados como complemento. Para este fim, duas categorias de sistemas são comumente estudados, àqueles voltados para classificação de alguma doença/lesão ou àqueles relacionados a recuperação de imagens similares de casos anteriores em uma base e dados de larga escala. Contudo, para desenvolver esses sistemas, geralmente é necessário utilizar *datasets* rotulados, que são empregados no treinamento de modelos preditivos ou para gerar modelos de extração de *features*. No entanto, rotular estas bases costuma ser um processo demorado e bastante oneroso. Para mitigar esse problema, é fundamental investigar abordagens alternativas, como o aprendizado auto-supervisionado, que apresenta potencial para lidar com essa limitação. Diante desse cenário, este estudo desenvolveu e avaliou experimentalmente modelos auto-supervisionados, aplicados em tarefas de classificação e recuperação de imagens baseada em conteúdo, no contexto histopatológico.

As abordagens auto-supervisionadas investigadas alcançaram resultados promissores ao serem adaptadas para o contexto de imagens histopatológicas. Os modelos foram eficazes para vários tipos de imagens (tecidos/órgãos), considerando várias medidas de avaliação. Foi verificado que o *fine-tuning* do DINOv2 proporcionou melhorias expressivas, permitindo considerá-lo equivalente ou melhor do que os extratores de *features* de última geração específicos do domínio histopatológico. Apesar disso,

observou-se que a eficácia para alguns *datasets* não atingiu níveis satisfatórios, mais especificamente para bases de glomérulos. No entanto, os modelos submetidos ao *fine-tuning* apresentaram melhorias, especialmente quando comparados à sua versão original, sem a aplicação do *fine-tuning*, ou seja, os modelos DINOv2 (ViT-S) e DINOv2 (ViT-B).

Em resposta à **Q1: Considerando modelos auto-supervisionados do estado da arte pré-treinados com imagens histopatológicas e com imagens naturais, quais seus níveis e diferença em termos de eficácia no contexto de classificação de doenças histopatológicas?** os resultados demonstraram que os modelos pré-treinados com imagens associadas ao domínio de interesse foram numericamente superiores. No entanto, é importante destacar que, mesmo os modelos pré-treinados com imagens naturais, alcançaram resultados expressivos. Isso foi verificado para a maioria dos *datasets*, exceto para as bases de dados de nefropatologia. Em relação à **Q2: Considerando modelos auto-supervisionados do estado da arte pré-treinados com imagens histopatológicas e com imagens naturais, qual seu nível de eficácia no contexto de recuperação de imagens histopatológicas e quais demonstram melhor adaptação para cada um dos sub-domínios da histopatologia considerados neste estudo?** também foi verificado que os modelos histopatológicos foram superiores aos pré-treinados com imagens naturais. As únicas exceções foram os *datasets* Kather e NCT-CRC-HE. Para esses *datasets*, o modelo DINOv2 (ViT-B) superou o DINOHist-S e o DINOHist-B, mas foi inferior ao RetCCL e ao Phikon ViT-B. Como o DINOHist-S e o DINOHist-B são versões ajustadas por *fine-tuning*, esse resultado sugere a necessidade de aprimorar a adaptação do DINOv2 para este tipo de dado.

Em resposta à **Q3: Em que medida modelos pré-treinados em imagens naturais, ajustados auto-supervisionadamente com poucos dados histopatológicos, impactam a eficácia na classificação e recuperação em comparação com extratores de estado da arte?** foi verificado que os modelos ajustados (DINOHist-S e DINOHist-B) foram superiores aos extratores de estado da arte específicos do domínio histopatológico (Phikon ViT-B e RetCCL). Essa melhoria foi mais acentuada nos *datasets* relacionados a patologias renais, isto é, o PathoSpotter-HE, PathoSpotter-PAS e PathoSpotter MultiContraste. Aqui é importante destacar este achado, pois, mesmo para estes *datasets*, cuja eficácia geral alcançada foi baixa, os métodos ajustados foram capazes de proporcionar melhorias significativas. Em relação à tarefa de recuperação de imagens, o Phikon ViT-B obteve um melhor desempenho para a maioria das bases, com exceção das três bases de nefropatologia. Para estas bases, os modelos adaptados foram estatisticamente superiores.

Em relação à **Q4: Como se comportam os extratores de características pré-treinados com imagens naturais quando aplicados ao domínio histopatológico?**, nota-se que, mesmo os modelos pré-treinados com imagens naturais, alcançaram resultados expressivos para quatro dos sete *datasets* investigados. Ressalta-se, ainda, que dentre os modelos pré-treinados com imagens naturais, os

modelos DINOv2 foram os mais eficazes, apresentando desempenho superior ao ResNet50 na maioria dos casos investigados, com ganhos superiores a 12%.

Por fim, em reposta a **Q5: Qual o impacto do *fine-tuning* em modelos auto-supervisionados no aprimoramento da representação de dados e na melhoria da eficácia em tarefas específicas, como classificação e recuperação de imagens?**, considerando a tarefa de classificação, verificou-se que o *fine-tuning* melhorou a eficácia para todos os *datasets* avaliados. Este resultado é notório para os *datasets* vinculados ao PathoSpotter, apresentando ganhos acima de 30% para algumas medidas. Em relação à recuperação de imagens baseada em conteúdo, também verificou-se uma maior eficácia dos modelos que utilizaram o *fine-tuning* na maioria dos *datasets*. Essa tendência não foi observada apenas nos datasets Kather e NCT-CRC-HE. Para estas bases o Phikon ViT-B foi superior.

Em resumo, espera-se que este trabalho sirva de base para outros pesquisadores aprofundarem as discussões e experimentos reportados. De modo especial, que estes trabalhos estejam direcionados para o paradigma auto-supervisionado, tendo em vista que ainda são incipientes os estudos que buscam soluções que ajudem a resolver problemas associados as bases não rotuladas neste domínio.

5.1 Direcionamentos de Pesquisas Futuras

Além dos resultados promissores, existem direcionamentos de pesquisa que podem ser explorados para estender este estudo. Foram organizadas cinco direções de trabalhos futuros, que são descritas detalhadamente a seguir.

- **Avaliação de novas arquiteturas:** Neste estudo avaliamos a eficácia de vários modelos, alguns treinados com imagens naturais, outros com imagens histopatológicas. No entanto, a avaliação da melhoria da eficácia proporcionada pelo *fine-tuning* foi realizada a partir da seleção de um dos modelos, isto é, o DINOv2. Existem outros modelos na literatura que podem ser explorados, como o iBOT (Zhou et al., 2022), EsVIT (Li et al., 2022) e o IJEPA (Assran et al., 2023). Além disso, pode-se explorar o *fine-tuning* a partir de modelos pré-treinados no mesmo domínio de interesse, como o Phikon ViT-B (Filiot et al., 2023).
- **Utilização de outros conjuntos de dados:** Para expandir a avaliação conduzida neste estudo, outras bases de dados podem ser utilizadas. Como sugestão, as seguintes bases podem ser exploradas: Camelyon[1], BreakHis[2], BACH[3], ou alguma base do Atlas de Patologia Digital[4]. Como demonstrado nos experimentos executados, verificou-se que os conjuntos de dados associados as

[1]https://camelyon17.grand-challenge.org/Data/
[2]https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/
[3]https://zenodo.org/records/3632035
[4]https://www.dsp.utoronto.ca/projects/ADP/

lesões renais foram as mais desafiadoras. Deste modo, explorar outras bases de dados pertencentes a este mesmo domínio configura-se como uma abordagem promissora a ser considerada. Exemplos de bases públicas associadas a este domínio incluem: NephroNet[5] e a TCGA-KIRC[6].

- **Investigar a influência dos hiperparâmetros:** Considerando que a variação dos hiperparâmetros pode afetar a convergência do modelo e, consequentemente, a sua eficácia, é preciso conduzir estudos que expandam a avaliação dos seus efeitos no domínio deste estudo. Como foi obervado, para alguns dos *datasets* utilizados, variar o número de *crops* aplicados pelo DINOv2 afetava a eficácia do modelo. Deste modo, é preciso estender essa análise para outros hiperparâmetros, de modo a gerar modelos mais eficazes e robustos.

- **Avaliar diferentes estratégias de *fine-tuning*:** Aplicar diferentes estratégias de *fine-tuning* pode permitir identificar modelos que sejam eficientes, eficazes e adaptáveis a uma variedade de situações e domínios. Além disso, o *fine-tuning* possibilita gerar tais modelos com recursos computacionais limitados. Afinal, o acesso a recursos computacionais robustos nem sempre está disponível. Adicionalmente, dependendo da arquitetura utilizada, há estratégias que são mais aderentes do que outras. Por exemplo, neste trabalho foi realizado o *fine-tuning* completo, ou seja, ajustando-se os pesos em todas as camadas do *backbone*. Uma estratégia alternativa é o congelamento de algumas dessas camadas para avaliar o impacto da abordagem no desempenho do modelo. Outras alternativas, como a *Low-Rank Adaptation* (LoRA) (Hu et al., 2021), também pode ser explorada e adaptada para o cenário de treinamento auto-supervisionado.

- **Redução da dimensionalidade:** O tamanho do vetor de características pode impactar negativamente na eficácia dos modelos. Vetores de alta dimensionalidade podem resultar em um problema denominado com maldição da dimensionalidade (Crespo Márquez, 2022). Nesta situação, à medida que as dimensões aumentam, a diferença entre a maior e a menor distância entre os pontos tende a diminuir. Em termos práticos, isso significa que todos os pontos tendem a ficar quase equidistantes entre si, tornando as medidas de distância menos úteis para discriminar entre instâncias. Deste modo, algoritmos que utilizam cálculo de distância entre amostras, como o kNN, podem ter sua eficácia reduzida. Assim, novos trabalhos podem ser conduzidos para investigar a aplicação de técnicas de redução de dimensionalidade.

- **Avaliação de outros classificadores**: Neste estudo, a avaliação dos extratores de *features* foi conduzida utilizando um classificador kNN, seguindo o que tem sido feito na literatura correlata. Contudo, é importante avaliar também, esses mesmos extratores com outros classificadores amplamente difundidos.

[5]https://zenodo.org/records/7498108
[6]https://www.cancerimagingarchive.net/collection/tcga-kirc/

Deste modo, trabalhos futuros podem explorar o uso de classificadores como: Logistic regression, SVM, Decision Tree, Random forest, entre outros.

Referências

Abels, E. et al. (2019). Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *J Pathol*, 249(3):286–294.

Anwar, S. M. et al. (2018). Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems*, 42(11):226.

Assran, M. et al. (2023). Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of CVPR*, páginas 15619–15629.

Azizi, S., Culp, L., Freyberg, J., e et al. (2023). Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nature Biomedical Engineering*, 7:756–779.

Baharoon, M., Qureshi, W., Ouyang, J., Xu, Y., Aljouie, A., e Peng, W. (2024). Evaluating general purpose vision foundation models for medical image analysis: An experimental study of dinov2 on radiology benchmarks.

Bao, H. et al. (2021). Beit: BERT pre-training of image transformers. *CoRR*, abs/2106.08254.

Barata, C. e Santiago, C. (2021). Improving the explainability of skin cancer diagnosis using cbir. páginas 550–559, Cham. Springer International Publishing.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Brondolo, F. e Beaussant, S. (2024). Dinov2 rocks geological image analysis: Classification, segmentation, and interpretability.

Bueno, G., Gonzalez-Lopez, L., Garcia-Rojo, M., Laurinavicius, A., e Deniz, O. (2020). Data for glomeruli characterization in histopathological images. *Data in Brief*, 29:105314.

Calumby, R. T., Duarte, A. A., Angelo, M. F., Santos, E., Sarder, P., dos Santos, W. L. C., e Oliveira, L. R. (2023). Toward real-world computational nephropathology. *Clinical Journal of the American Society of Nephrology*, 18(6):809–812.

Caron, M. et al. (2020). Unsupervised learning of visual features by contrasting cluster assignments. NIPS'20, Red Hook, NY, USA. Curran Associates Inc.

Chagas, P., Souza, L., Araújo, I., Aldeman, N., Duarte, A., Angelo, M., Dos-Santos, W., e Oliveira, L. (2020). Classification of glomerular hypercellularity using convolutional features and support vector machine. *Artificial Intelligence in Medicine*, 103:101808.

Chan, H.-P. et al. (2020). Deep learning in medical image analysis. *Adv Exp Med Biol*, 1213:3–21.

Chen, C. et al. (2022). Fast and scalable search of whole-slide images via self-supervised deep learning. *Nat Biomed Eng*, 6(12):1420–1434.

Chen, L. et al. (2019). Self-supervised learning for medical image analysis using image context restoration. *Medical Image Analysis*, 58:101539.

Chen, X. e He, K. (2021). Exploring simple siamese representation learning. In *CVPR*, páginas 15745–15753.

Chen, X., Xie, S., e He, K. (2021). An empirical study of training self-supervised vision transformers. In *ICCV*, páginas 9620–9629, Los Alamitos, CA, USA. IEEE Computer Society.

Crespo Márquez, A. (2022). *The Curse of Dimensionality*, páginas 67–86. Springer International Publishing, Cham.

Damm, S., Laszkiewicz, M., Lederer, J., e Fischer, A. (2025). Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2.

Doersch, C., Gupta, A., e Efros, A. A. (2015). Unsupervised visual representation learning by context prediction. In *2015 IEEE International Conference on Computer Vision (ICCV)*, páginas 1422–1430.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*, 31(4-5):198–211.

Dosovitskiy, A. et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*. OpenReview.net.

Dromain, C. et al. (2012). Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol*, 82(3):417–423.

Erfankhah, H. et al. (2019). Heterogeneity-aware local binary patterns for retrieval of histopathology images. *IEEE Access*, 7:18354–18367.

Filiot, A. et al. (2023). Scaling self-supervised learning for histopathology with masked image modeling. *medRxiv*.

Gudivada, V. e Raghavan, V. (1995). Content based image retrieval systems. *Computer*, 28(9):18–22.

Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., e Tao, D. (2024). A survey on self-supervised learning: Algorithms, applications, and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9052–9071.

Hameed, I. M. et al. (2021). Content-based image retrieval: A review of recent trends. *Cogent Engineering*, 8(1):1927469.

Han, J., Kamber, M., e Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edição.

Haq, N. F. et al. (2021). A deep community based approach for large scale content based x-ray image retrieval. *Medical Image Analysis*, 68:101847.

He, K., Zhang, X., Ren, S., e Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, páginas 770–778.

Hegde, N. et al. (2019). Similar image search for histopathology: Smily. *npj Digital Medicine*, 2(1):56.

Hemmatirad, K., Babaie, M., Hodgin, J., Pantanowitz, L., e Tizhoosh, H. R. (2023). An investigation into glomeruli detection in kidney he and pas images using yolo.

Hosseini, M. S., Bejnordi, B. E., Trinh, V. Q.-H., Chan, L., Hasan, D., Li, X., Yang, S., Kim, T., Zhang, H., Wu, T., Chinniah, K., Maghsoudlou, S., Zhang, R., Zhu, J., Khaki, S., Buin, A., Chaji, F., Salehi, A., Nguyen, B. N., Samaras, D., e Plataniotis, K. N. (2024). Computational pathology: A survey review and the way forward. *Journal of Pathology Informatics*, 15:100357.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., e Chen, W. (2021). Lora: Low-rank adaptation of large language models.

Jing, L. e Tian, Y. (2021). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):4037–4058.

K., C. A., S., L., R., K., T., K. J., e J., K. (2023). A novel dataset and efficient deep learning framework for automated grading of renal cell carcinoma from kidney histopathology images. *Scientific Reports*, 13(1):5728.

Kalra, S. et al. (2020). Yottixel – an image search engine for large archives of histopathology whole slide images. *Medical Image Analysis*, 65:101757.

Kang, C., Lee, C., Song, H., Ma, M., e Pereira, S. (2023a). Variability matters: Evaluating inter-rater variability in histopathology for robust cell detection. In Karlinsky, L., Michaeli, T., e Nishino, K., editores, *Computer Vision – ECCV 2022 Workshops*, páginas 552–565, Cham. Springer Nature Switzerland.

Kang, M. et al. (2023b). Benchmarking self-supervised learning on diverse pathology datasets. páginas 3344–3354, Los Alamitos, CA, USA. IEEE Computer Society.

Kataria, T. et al. (2023). To pretrain or not to pretrain? a case study of domain-specific pretraining for semantic segmentation in histopathology. In *Medical Image Learning with Limited and Noisy Data*, páginas 246–256, Cham. Springer Nature Switzerland.

Kather, J. N., Halama, N., e Marx, A. (2018). 100,000 histological images of human colorectal cancer and healthy tissue.

Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.-A., Gaiser, T., Marx, A., Valous, N. A., Ferber, D., Jansen, L., Reyes-Aldasoro, C. C., Zörnig, I., Jäger, D., Brenner, H., Chang-Claude, J., Hoffmeister, M., e Halama, N. (2019). Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOS Medicine*, 16(1):1–22.

Komura, D. e Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16:34–42.

Koohbanani, N. A., Unnikrishnan, B., Khurram, S. A., Krishnaswamy, P., e Rajpoot, N. (2021). Self-path: Self-supervision for classification of pathology images with limited annotations. *IEEE Transactions on Medical Imaging*, 40(10):2845–2856.

Kumar, N., Gupta, R., e Gupta, S. (2020). Whole slide imaging (wsi) in pathology: Current perspectives and future directions. *Journal of Digital Imaging*, 33(4):1034–1040.

Li, C. et al. (2022). Efficient self-supervised vision transformers for representation learning.

Li, X. et al. (2021). Recent developments of content-based image retrieval (cbir). *Neurocomputing*, 452:675–689.

L'Imperio, V. et al. (2021). Digital pathology for the routine diagnosis of renal diseases: a standard model. *Journal of Nephrology*, 34(3):681–688.

Majumdar, S. et al. (2023). Gamma function based ensemble of cnn models for breast cancer detection in histopathology images. *Expert Systems with Applications*, 213:119022.

Manning, C. D., Raghavan, P., e Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, USA.

Mohammad Alizadeh, S. et al. (2023). A novel siamese deep hashing model for histopathology image retrieval. *Expert Systems with Applications*, 225:120169.

Moor, M., Huang, Q., Wu, S., Yasunaga, M., Dalmia, Y., Leskovec, J., Zakka, C., Reis, E. P., e Rajpurkar, P. (2023). Med-flamingo: a multimodal medical few-shot learner. In Hegselmann, S., Parziale, A., Shanmugam, D., Tang, S., Asiedu, M. N., Chang, S., Hartvigsen, T., e Singh, H., editores, *Proceedings of the 3rd Machine Learning for Health Symposium*, volume 225 of *Proceedings of Machine Learning Research*, páginas 353–367. PMLR.

Müller-Franzes, G., Khader, F., Siepmann, R., Han, T., Kather, J. N., Nebelung, S., e Truhn, D. (2024). Medical slice transformer: Improved diagnosis and explainability on 3d medical images with dinov2.

Navid Farahani, A. et al. (2015). Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathology and Laboratory Medicine International*, 7:23–33.

Oquab, M. et al. (2023). Dinov2: Learning robust visual features without supervision.

Orchard, G. e Nation, B. (2012). *Histopathology*. Fundamentals of Biomedical Science. OUP Oxford.

Ozen, Y. et al. (2021). Self-supervised learning with graph neural networks for region of interest retrieval in histopathology. In *ICPR*, páginas 6329–6334.

Pathak, D., Krähenbühl, P., Donahue, J., Darrell, T., e Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 2536–2544.

Pouyanfar, S. et al. (2018). A survey on deep learning: Algorithms, techniques, and applications. 51(5).

Shi, X. et al. (2018). Pairwise based deep ranking hashing for histopathology image classification and retrieval. *Pattern Recognition*, 81:14–22.

Shurrab, S. e Duwairi, R. (2022). Self-supervised learning methods and applications in medical imaging analysis: a survey. *PeerJ Comput Sci*, 8:e1045.

Souid, A. et al. (2023). Improving diagnosis accuracy with an intelligent image retrieval system for lung pathologies detection: a features extractor approach. *Scientific Reports*, 13(1):16619.

Torres, R. e Falcão, A. (2006). Content-based image retrieval: Theory and applications. *RITA*, 13:161–185.

Veeling, B. S., Linmans, J., Winkens, J., Cohen, T., e Welling, M. (2018). Rotation equivariant CNNs for digital pathology.

Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., Chen, B., e Wu, Y. (2014). Learning fine-grained image similarity with deep ranking. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, página 1386–1393, USA. IEEE Computer Society.

Wang, X., Du, Y., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., e Han, X. (2023). Retccl: Clustering-guided contrastive learning for whole-slide image retrieval. *Medical Image Analysis*, 83:102645.

Wang, X. et al. (2020). Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, 50(9):3950–3962.

Wickstrøm, K., Trosten, D., Løkse, S., et al. (2023a). Relax: Representation learning explainability. *International Journal of Computer Vision*, 131(10):1584–1610.

Wickstrøm, K. K. et al. (2023b). A clinically motivated self-supervised approach for content-based image retrieval of ct liver images. *Computerized Medical Imaging and Graphics*, 107:102239.

Yamaguchi, R. et al. (2021). Glomerular classification using convolutional neural networks based on defined annotation criteria and concordance evaluation among clinicians. *Kidney International Reports*, 6(3):716–726.

Yang, P. et al. (2020). A deep metric learning approach for histopathological image retrieval. *Methods*, 179:14–25. Interpretable machine learning in bioinformatics.

Yao, T., Lu, Y., Long, J., Jha, A., Zhu, Z., Asad, Z., Yang, H., Fogo, A. B., e Huo, Y. (2022). Glo-in-one: holistic glomerular detection, segmentation, and lesion characterization with large-scale web image mining. *Journal of Medical Imaging*, 9(5):052408.

Yoshinobu, Y. et al. (2020). Deep learning method for content-based retrieval of focal liver lesions using multiphase contrast-enhanced computer tomography images. páginas 1–4.

Zheng, Y. et al. (2022). Encoding histopathology whole slide images with location-aware graphs for diagnostically relevant regions retrieval. *Medical Image Analysis*, 76:102308.

Zhou, J. et al. (2022). ibot: Image bert pre-training with online tokenizer.