



UNIVERSIDADE FEDERAL DO MARANHÃO
Programa de Pós-Graduação em Ciência da Computação

José Ribamar Durand Rodrigues Júnior

**HistAttentionNAS: Uma CNN construída via
NAS para o diagnóstico do câncer do pênis
usando imagens histopatológicas**

São Luís - MA

2025

José Ribamar Durand Rodrigues Júnior

HistAttentionNAS: Uma CNN construída via NAS para o diagnóstico do câncer do pênis usando imagens histopatológicas

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Geraldo Braz Junior

São Luís - MA

2025

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

Rodrigues Junior, José Ribamar Durand.

HistAttentionNAS: Uma CNN construída via NAS para o diagnóstico do câncer do pênis usando imagens histopatológicas / José Ribamar Durand Rodrigues Junior. - 2025.

93 p.

Orientador(a): Geraldo Braz Junior.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, São Luís, 2025.

1. Câncer de Pênis. 2. Imagens Histopatológicas. 3. Aprendizado Profundo. 4. Redes Neurais Convolucionais. 5. Pesquisa de Arquitetura Neural. I. Braz Junior, Geraldo. II. Título.

José Ribamar Durand Rodrigues Júnior

HistAttentionNAS: Uma CNN construída via NAS para o diagnóstico do câncer do pênis usando imagens histopatológicas

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Trabalho _____ São Luís - MA, 27 de fevereiro de 2025:

Prof. Dr. Geraldo Braz Junior

Orientador

Universidade Federal do Maranhão

Prof. Dr. João Dallyson Sousa de Almeida

Examinador Interno

Universidade Federal do Maranhão

Prof. Dr. Darlan Bruno Pontes Quintanilha

Examinador Interno

Universidade Federal do Maranhão

Prof. Dr. Cláudio de Souza Baptista

Examinador Externo

UFCG

São Luís - MA

2025

Dedico este trabalho à minha mãe, cuja dedicação e sacrifício foram fundamentais para que minha irmã e eu tivéssemos acesso a uma educação de qualidade. Sua força, amor e apoio incondicional sempre me guiaram, tanto na trajetória acadêmica quanto em todos os aspectos da minha vida. Esta conquista é, em grande parte, fruto do seu esforço e da sua crença no nosso potencial. Com toda a gratidão e amor, esta dedicatória é para você.

Agradecimentos

A minha mãe, Maria de Nazaré, expresso minha profunda gratidão por seu constante incentivo aos estudos e por ter valorizado, desde sempre, minha educação. Você é meu maior exemplo de coragem, força, perseverança e caráter, ensinando-me o valor do esforço, da honestidade e da determinação — princípios que levarei comigo por toda a vida.

A minha irmã, Anna Beatriz, agradeço de coração por todo o apoio prestado ao longo destes anos, desde a graduação até o mestrado. Sua presença e incentivo foram fundamentais para que eu pudesse alcançar meus objetivos, e sou extremamente grato por tudo o que fez por mim.

À minha noiva, Gilvana Siqueira, agradeço pelo carinho e companheirismo que me acompanham há sete anos de relacionamento. Obrigado pelo amor e pelo suporte incondicional, que tornaram esta trajetória mais serena e feliz.

Aos meus amigos de laboratório, Lucas Araujo, Marcos Melo, Caio Falcão, Mackele Jurema e Matheus Levy, sou grato por todo conhecimento compartilhado e pelas trocas valiosas que vivenciamos. Vocês contribuíram imensamente para meu crescimento pessoal e acadêmico durante esta etapa.

Agradeço ainda a todos os professores do PPGCC, cuja dedicação e atenção foram fundamentais para esclarecer dúvidas, oferecer sugestões em apresentações e trabalhos, e ampliar meus horizontes acadêmicos.

Ao meu estimado orientador, Geraldo Braz, expresso um sincero reconhecimento por toda a atenção e apoio desde o período de graduação e pelas experiências no PET. Sou grato pelas oportunidades e orientações que tanto enriqueceram minha formação acadêmica e profissional.

*"Por mais que eu compreenda bem, minha compreensão pode ser apenas uma fração
infinitesimal de tudo o que desejo entender."*

(Ada Lovelace)

Resumo

O câncer de pênis (CP) é uma neoplasia rara que, embora pouco frequente em países desenvolvidos, apresenta incidência alarmante em determinadas regiões de países em desenvolvimento. Seu diagnóstico precoce é fundamental para melhorar o prognóstico e reduzir a necessidade de intervenções cirúrgicas mutiladoras. A análise histopatológica permanece como o padrão ouro para diagnóstico, porém envolve processos subjetivos que podem atrasar o laudo e gerar divergências entre especialistas. Nesse sentido, o aprendizado profundo tem se mostrado promissor ao agilizar e padronizar avaliações, auxiliando patologistas na identificação de padrões relevantes. Este trabalho propõe uma metodologia para a classificação de CP em imagens histopatológicas digitais. Inicia-se com a preparação dos dados, balanceando as classes por meio de transformações geométricas e aplicando diferentes técnicas de pré-processamento e perturbações de cor. Em seguida, um processo de Pesquisa de Arquitetura Neural é empregado para definir, de forma automatizada, a melhor combinação de blocos iniciais (*stem*) e mecanismos de atenção nos estágios principais (*backbone*) da rede, otimizando a extração de texturas e informações discriminativas. Nos experimentos realizados, as arquiteturas otimizadas apresentaram valores de F1-Score de 0,927 em 40x e 0,958 em 100x, confirmando a eficácia da estratégia. Comparações com trabalhos relacionados na mesma base mostram resultados competitivos, em particular na ampliação de 100x, mesmo sem o uso de grandes bases externas ou transferência de aprendizado. Conclui-se que o balanceamento de classes, a configuração do pré-processamento e aumento de representatividade, aliados à automação via pesquisa de arquitetura neural (NAS), potencializam o desempenho da classificação de câncer de pênis.

Palavras-chave: Câncer de pênis, Imagens histopatológicas, Aprendizado profundo, Redes neurais convolucionais, Pesquisa de arquitetura neural, Mecanismos de atenção.

Abstract

Penile cancer (PC) is a rare neoplasm which, although infrequent in developed countries, has an alarming incidence in some areas of developing countries. Early diagnosis is essential to improve prognosis and reduce the need for mutilating surgical interventions. Histopathological analysis remains the gold standard for diagnosis, but it involves subjective processes that can delay the report and lead to disagreements between specialists. In this sense, deep learning has shown promise in speeding up and standardizing assessments, helping pathologists to identify relevant patterns. This paper proposes a methodology for classifying PC in digital histopathology images. It begins by preparing the data, balancing the classes using geometric transformations and applying different pre-processing and colour disturbance techniques. Next, a Neural Architecture Search process is used to automatically define the best combination of stem blocks and attention mechanisms in the backbone stages of the network, optimizing the extraction of textures and discriminative information. In the experiments, the optimized architectures showed F1-Score values of 0.927 at 40x and 0.958 at 100x, confirming the strategy's effectiveness. Comparisons with related work on the same basis show competitive results, particularly at 100x magnification, even without using large external bases or transfer learning. It is concluded that class balancing, pre-processing configuration and increased representativeness, combined with automation via neural architecture search (NAS), boost the performance of penile cancer classification.

Keywords: Penile cancer, Histopathological images, Deep learning, Convolutional neural networks, Neural architecture research, Attention mechanisms.

Lista de ilustrações

Figura 1 – Imagens do Dataset	16
Figura 2 – Imagem de Lâmina Inteira	24
Figura 3 – Rede Neural	26
Figura 4 – Operação de Convolução	28
Figura 5 – Camada de <i>Pooling</i> Máximo	29
Figura 6 – Blocos Resnet	30
Figura 7 – Mecanismo de Atenção BAM	32
Figura 8 – Mecanismo de Atenção CBAM	33
Figura 9 – Metodologia NAS	34
Figura 10 – Equalização de canais YUV	37
Figura 11 – Equalização de canais cieLAB	38
Figura 12 – Equalização de canais H&E	38
Figura 13 – Normalização Vahadane	40
Figura 14 – Normalização stainGAN	40
Figura 15 – Normalização stainNet	41
Figura 16 – Transformações Rotação e Inversão.	42
Figura 17 – Transformações de Brilho e Contraste.	42
Figura 18 – Transformações no Espaço de Cor HSV	43
Figura 19 – Transformações no Espaço de Cor H&E	44
Figura 20 – Processo de Validação Cruzada utilizada no trabalho	46
Figura 21 – Matriz de Confusão.	47
Figura 22 – Imagens do Dataset	55
Figura 23 – Fluxograma da metodologia proposta.	56
Figura 24 – Transformações Aplicadas no Balanceamento de Dados.	58
Figura 25 – Pré-processamentos aplicados no estudo	59
Figura 26 – Transformações Aplicadas no Aumento de Representatividade	60
Figura 27 – blocos do Stem	61
Figura 28 – Células do Backbone	62
Figura 29 – Arquitetura e Espaço de Busca.	63
Figura 30 – Melhor arquitetura para imagens com ampliação de 40X.	74
Figura 31 – Melhor arquitetura para imagens com ampliação de 100X.	76

Lista de tabelas

Tabela 1 – Trabalhos relacionados, com base de dados usada, resumo do método e resultados obtidos.	53
Tabela 2 – Distribuição das imagens conforme ampliação e classificação patológica.	55
Tabela 3 – Hiper-parâmetros de Treinamento	65
Tabela 4 – Quantidade de imagens geradas para o primeiro <i>fold</i>	66
Tabela 5 – Resultados do Balanceamento de Dados para imagens 40X	67
Tabela 6 – Resultados do Balanceamento de Dados para imagens 100X	68
Tabela 7 – F1-Score médio para Pré-processamento e Aumento em 40x.	69
Tabela 8 – F1-Score médio para Pré-processamento e Aumento em 100x	70
Tabela 9 – Arquiteturas Otimizadas para Ampliação de 40x – Top 10 Configurações	73
Tabela 10 – Arquiteturas Otimizadas para Ampliação de 100x – Top 10 Configurações.	75
Tabela 11 – Comparação entre subetapas.	77
Tabela 12 – Comparação com trabalhos relacionados.	78

Lista de abreviaturas e siglas

BAM	<i>Bottleneck Attention Module</i>
CBAM	<i>Convolutional Block Attention Module</i>
CCE	<i>Carcinoma de Células Escamosas</i>
cieLAB	<i>Espaço de cor CIE Lab</i>
CLAHE	<i>Contrast Limited Adaptive Histogram Equalization</i>
CNN	<i>Convolutional Neural Network (Rede Neural Convolutacional)</i>
CP	<i>Câncer de Pênis</i>
H&E	<i>Hematoxilina e Eosina</i>
HED	<i>Hematoxilina, Eosina e Diaminobenzidina</i>
HPV	<i>Papilomavírus Humano</i>
IA	<i>Inteligência Artificial</i>
IARC	<i>International Agency for Research on Cancer (Agência Internacional de Pesquisa em Câncer)</i>
ISTs	<i>Infecções Sexualmente Transmissíveis</i>
MLP	<i>Multilayer Perceptron</i>
NAS	<i>Neural Architecture Search (Pesquisa de Arquitetura Neural)</i>
SGD	<i>Stochastic Gradient Descent (Descida de Gradiente Estocástica)</i>
TPE	<i>Tree-structured Parzen Estimator</i>
WSI	<i>Whole Slide Imaging (Imagem de lâmina inteira)</i>
YUV	<i>Espaço de cor formado por Luminância (Y) e Crominâncias (U e V)</i>

Sumário

1	INTRODUÇÃO	14
1.1	Objetivos	17
1.2	Questões de Pesquisa.	18
1.3	Contribuições	18
1.4	Organização do Trabalho	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Câncer de Pênis	21
2.2	Análise Histopatológica	22
2.3	Patologia Digital	23
2.4	Aprendizado Profundo	25
2.4.1	Redes Neurais Convolucionais	27
2.4.2	Arquitetura Resnet	29
2.4.3	Mecanismos de atenção	31
2.4.3.1	BAM: <i>Bottleneck Attention Module</i>	32
2.4.3.2	CBAM: <i>Convolutional Block Attention Module</i>	32
2.5	Pesquisa de Arquitetura Neural	33
2.6	Preprocessamento e Aumento de Dados	35
2.6.1	Pré-processamento	37
2.6.2	Aumento de Dados	41
2.7	Avaliação	45
2.8	Considerações Finais	48
3	TRABALHOS RELACIONADOS	49
3.1	Mecanismos de Atenção na Histopatologia Digital	49
3.2	Aprendizado Profundo para Detecção de Câncer de Pênis	50
3.3	Diferenciais da Metodologia Proposta	52
4	MATERIAIS E MÉTODO	54
4.1	Ambiente e Configurações experimentais	54
4.2	Dataset	54
4.3	Método	55
4.3.1	Preparação dos Dados	56
4.3.1.1	Balanceamento de Classes	57
4.3.1.2	Pré-processamento e Aumento de Representatividade	58
4.3.2	NAS: Pesquisa de Arquitetura Neural	60

4.3.2.1	Definição do Espaço de Busca	60
4.3.2.1.1	Blocos Stem	60
4.3.2.1.2	Blocos Backbone	61
4.3.2.1.3	Definição da Arquitetura	62
4.3.2.2	Estratégia de Busca	63
4.3.2.3	Estimativa de Desempenho	64
4.4	Considerações Finais	64
5	RESULTADOS	65
5.1	Definição de Hiperparâmetros	65
5.2	Preparação dos Dados	66
5.2.1	Balanceamento de Classes	66
5.2.2	Pré-processamento e Aumento de Representatividade	69
5.3	NAS: Pesquisa de Arquitetura Neural	72
5.3.1	Arquiteturas 40x	72
5.3.2	Arquiteturas 100x	74
5.3.3	Análise Comparativa das Arquiteturas	75
5.4	Desempenho nas etapas do Método	76
5.5	Comparação com trabalhos relacionados	77
5.6	Questões de Pesquisa	79
5.7	Limitações	81
5.8	Trabalhos Futuros	82
6	CONCLUSÃO	84
	REFERÊNCIAS	86

1 Introdução

O câncer de pênis (CP) é uma neoplasia maligna rara, com baixas incidências em países desenvolvidos. Este câncer representa 0,4% a 0,6% de todas as neoplasias malignas que atingem homens nos Estados Unidos e na Europa (KORKES et al., 2020). Em contrapartida, em países em desenvolvimento, como o Brasil, a incidência é consideravelmente mais elevada, alcançando 2,1% de todas as neoplasias que acometem os homens, com maior frequência em indivíduos com mais de 50 anos (COELHO et al., 2018). Diversos fatores clínicos estão associados ao seu desenvolvimento, incluindo fimose, tabagismo, infecções por ISTs e comportamentos sexuais de risco. Contudo, fatores socioeconômicos também desempenham papel relevante, com destaque para a educação formal precária, o contexto rural e o trabalho no campo, que têm mostrado correlação com a incidência da doença (JÚNIOR et al., 2022).

Um estudo realizado nos Estados Unidos entre 1998 e 2003 constatou que quase todos os pacientes diagnosticados com o tipo mais comum de câncer de pênis, o carcinoma de células escamosas (CCE), precisaram passar por alguma intervenção cirúrgica, com 72% deles submetidos a uma amputação parcial ou total do membro (HERNANDEZ et al., 2008). A cirurgia de amputação é mais comum em pacientes com estágio mais avançado da doença e em indivíduos mais idosos, destacando a importância do diagnóstico precoce dessa condição. Embora a amputação melhore as taxas de cura, os benefícios terapêuticos devem ser cuidadosamente equilibrados com os impactos psicossociais. Cerca de 40% dos pacientes relataram efeitos negativos no bem-estar após a cirurgia; além disso, sintomas psiquiátricos foram observados em 50% dos pacientes (RICHTER et al., 2013).

O diagnóstico do câncer de pênis (CP) é tradicionalmente realizado por meio de biópsia do tecido, seguido de uma avaliação criteriosa conduzida por um médico especialista, que busca identificar anormalidades no tecido coletado (OLIVEIRA et al., 2020). Embora o exame histopatológico visual seja amplamente reconhecido como o padrão ouro para diversos tipos de câncer, essa abordagem apresenta limitações significativas, como subjetividade, tempo prolongado para obtenção de resultados e a possibilidade de discordâncias entre observadores (GHANATE et al., 2011a). Essas limitações incentivam o uso de técnicas de inteligência artificial (IA), que oferecem soluções promissoras devido à eficiência em metodologias de reconhecimento de padrões. Essas tecnologias podem auxiliar na identificação das características morfológicas relevantes, tradicionalmente analisadas por especialistas, como anormalidades morfológicas (LAAK; LITJENS; CIOMPI, 2021).

As primeiras aplicações de patologia computacional eram fundamentadas em modelos manualmente projetados para reconhecer características nas imagens, que

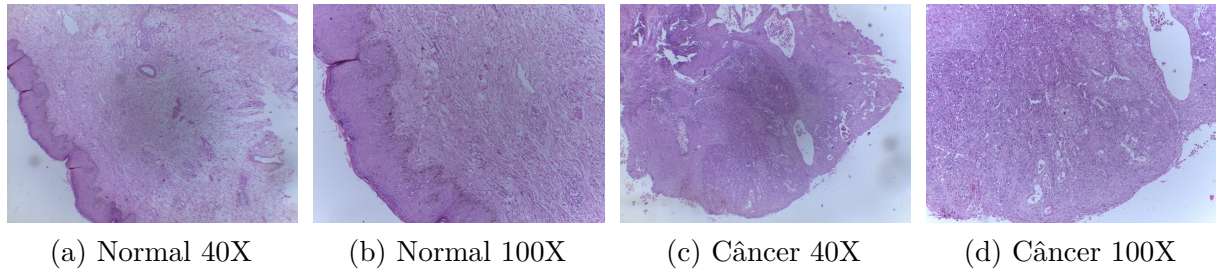
posteriormente foram substituídos por extratores de características quantitativos associados a classificadores, visando identificar relações entre as características extraídas e os diagnósticos das amostras (LAAK; LITJENS; CIOMPI, 2021). Com o avanço das redes neurais convolucionais (CNNs), essas abordagens passaram por uma transformação significativa. As CNNs demonstraram desempenho superior em diversas tarefas de visão computacional, superando métodos tradicionais, e destacaram-se por sua capacidade de aprender características diretamente dos dados brutos, eliminando a necessidade de intervenção de especialistas para definição de características específicas. Além disso, a crescente disponibilidade de conjuntos de dados de histopatologia anotados impulsionou ainda mais o interesse no uso de aprendizado profundo aplicado à área (SRINIDHI; CIGA; MARTEL, 2021).

A transição quase completa da engenharia de características manuais para o aprendizado profundo deve-se a vários fatores. A capacidade de construir algoritmos que dependem predominantemente de treinamento em vez de programação explícita ou filtros predefinidos possibilitou a criação de representações hierárquicas poderosas, que frequentemente superam os métodos tradicionais de análise de imagens. Além disso, essa abordagem reduz substancialmente a necessidade de conhecimento especializado em domínio para o desenvolvimento de modelos, democratizando o acesso às técnicas e ampliando as possibilidades de aplicação (LAAK; LITJENS; CIOMPI, 2021).

As texturas desempenham um papel crucial na classificação de imagens histopatológicas (TSAKU et al., 2019; HIREN et al., 2021). Em redes neurais convolucionais (CNNs), as características texturais são extraídas principalmente nas primeiras camadas convolucionais, destacando-se como elementos fundamentais para a análise histopatológica. A Figura 1 apresenta exemplos de imagens do tecido peniano com câncer e normais. Essas características, que capturam padrões locais e variações de intensidade nas imagens, fornecem informações essenciais para a diferenciação de tecidos normais e anormais. Dada a importância das primeiras camadas para extração de características de baixo nível fundamentais para diferenciação dos tecidos nas imagens histopatológicas, a otimização dessas camadas pode melhorar significativamente o desempenho de arquiteturas desenvolvidas. Os mecanismos de atenção em redes neurais convolucionais (CNNs) emergiram como uma estratégia eficaz para suprimir ou realçar características com base em sua relevância para a distinção entre classes. Esses mecanismos são geralmente acoplados à saída convolucional dos blocos da rede, refinando as características extraídas e, assim, promovendo uma melhoria no desempenho das arquiteturas sem acarretar um aumento significativo no número de parâmetros (WOO et al., 2018a; PARK et al., 2018). Estudos recentes destacaram o uso de mecanismos de atenção para análise de imagens histopatológicas, apresentando avanços significativos no desempenho dos modelos (LI et al., 2022a; CHEN et al., 2022c; CHEN et al., 2022a). Nos trabalhos mencionados, os autores delinearam arquiteturas onde o posicionamento e o tipo dos mecanismos de

atenção foram escolhidos estrategicamente, com base nas hipóteses formuladas para cada aplicação. Esses achados sugerem que a incorporação de mecanismos de atenção, desde que devidamente adicionados, pode ser uma ferramenta poderosa para otimizar a extração e a utilização de características em imagens histopatológicas.

Figura 1 – Imagens do Dataset



Fonte: elaborada pelo Autor.

Observa-se que os desempenhos obtidos por muitos modelos de aprendizado profundo ainda dependem amplamente de arquiteturas projetadas manualmente, que exigem um nível considerável de especialização humana e conhecimento específico do domínio. O processo de projetar manualmente arquiteturas de redes neurais profundas é uma tarefa intensiva e trabalhosa, exigindo esforço significativo para alcançar arquiteturas de última geração (OYELADE; EZUGWU, 2021).

Com o intuito de reduzir essa problemática, surgiram metodologias de Pesquisa de Arquitetura Neural (Neural Architecture Search, NAS), que buscam automatizar o processo de projeto de arquiteturas de aprendizado profundo por meio de métodos de otimização (OYELADE; EZUGWU, 2021). O NAS é estruturado em três componentes principais: espaço de busca, estratégia de busca e estratégia de avaliação (REN et al., 2021). O espaço de busca é definido como o conjunto de todas as arquiteturas possíveis num conjunto predefinido de regras ou operadores. A estratégia de busca determina como o espaço será explorado, guiando o processo de seleção de arquiteturas candidatas. Por fim, a estratégia de avaliação mede o desempenho dessas arquiteturas sugeridas, fornecendo métricas para otimizar a busca.

O NAS foi criado para mitigar ainda mais essa dependência do ser humano no processo, automatizando o processo de projeto de arquiteturas. No entanto, o sucesso do NAS depende da definição adequada do espaço de busca, que delimita o conjunto de arquiteturas possíveis a serem exploradas.

A incorporação de conhecimentos específicos do domínio na definição do espaço de busca é uma estratégia essencial para restringi-lo eficazmente (CHEN et al., 2022d). Essa abordagem permite otimizar as buscas, direcionando os esforços para configurações mais promissoras, reduzindo consideravelmente o custo computacional do processo. Além disso, um espaço de busca bem delimitado não apenas acelera o processo de pesquisa, mas

também contribui para a melhoria do desempenho final dos modelos gerados, evidenciando a importância do equilíbrio entre automação e o conhecimento do domínio investigado.

Neste trabalho, foi desenvolvida uma metodologia para a detecção de câncer de pênis (CP) baseada em aprendizado profundo, visando aprimorar a eficácia das arquiteturas a partir da pesquisa de arquitetura neural (NAS) e preparação dos dados. Inicialmente, foi delimitado e avaliado o desempenho de diferentes combinações de balanceamento de classes, pré-processamento e aumento de representatividade aplicadas a uma arquitetura base, para identificar o conjunto que proporcionasse os melhores resultados. Essa etapa visa estabelecer um ponto de partida robusto para a subsequente pesquisa de arquitetura neural. A partir dessa otimização, foram exploradas modificações em suas camadas iniciais (*stem layers*) e nos blocos principais (*backbone*). A otimização das camadas iniciais teve como objetivo aprimorar a extração de características texturais de forma eficiente, enquanto as alterações nos blocos do *backbone* foram direcionadas à incorporação estratégica de mecanismos de atenção.

1.1 Objetivos

O objetivo geral deste trabalho é desenvolver e validar uma metodologia para a detecção de câncer de pênis em imagens histopatológicas digitais, baseado em pesquisa de arquitetura neural (NAS) para a criação de uma arquitetura otimizada para este domínio específico. Além disso, o trabalho explora técnicas como balanceamento de classes, aumento de representatividade e pré-processamento, visando aumentar a robustez do modelo, aprimorar sua capacidade de lidar com o desbalanceamento de classes e mitigar os desafios decorrentes da escassez de dados.

Para alcançar tal objetivo geral, foram propostos alguns objetivos específicos:

- Realizar o balanceamento de classes para mitigar o impacto de distribuições desiguais no conjunto de dados;
- Avaliar o impacto de diferentes estratégias de pré-processamento e aumento de dados na melhoria da qualidade e diversidade das imagens histopatológicas;
- Analisar o papel de diferentes configurações de camadas de *stem* na extração de características de baixo nível, como texturas, e sua influência no desempenho global do modelo.
- Investigar o impacto de diferentes mecanismos de atenção e sua aplicação estratégica nos estágios do *backbone* das arquiteturas.

- Desenvolver e aplicar uma metodologia de pesquisa de arquitetura neural (NAS) baseada em células para explorar configurações eficientes e adaptadas ao diagnóstico de câncer de pênis.
- Adquirir arquiteturas otimizadas para o contexto histopatológico de câncer de pênis, a partir da metodologia desenvolvida.

1.2 Questões de Pesquisa.

Com base nos objetivos delineados neste trabalho, as seguintes questões de pesquisa foram elaboradas para guiar a investigação, considerando os desafios específicos do domínio de imagens histopatológicas de câncer de pênis.

- Qual a proporção ideal entre as classes no conjunto de dados que maximiza o desempenho do modelo, evitando viés de aprendizado e assegurando a generalização para dados desbalanceados, como observado no problema em questão?
- Qual a combinação mais eficiente de estratégias de pré-processamento e aumento de representatividade que melhora a qualidade das imagens histopatológicas de câncer de pênis, potencializa a generalização do modelo e mitiga os efeitos de variações indesejadas nas características visuais?
- Quais configurações de camadas de *stem* melhor extraem características de baixo nível, como texturas, nas imagens histopatológicas de câncer de pênis, e como essas configurações impactam o desempenho final do modelo?
- Como a adoção de mecanismos de atenção influenciam a capacidade dos modelos de extrair representações discriminativas e relevantes nas imagens histopatológicas de câncer de pênis?
- Como a pesquisa de arquitetura neural pode melhorar os resultados a partir de uma arquitetura otimizada para o domínio histopatológico de câncer de pênis?

1.3 Contribuições

Destacam-se como principais contribuições:

- A proposição de uma metodologia estruturada para a preparação de dados, abrangendo etapas de balanceamento de classes, pré-processamento e aumento de representatividade em imagens histopatológicas de câncer de pênis. A abordagem demonstrou impacto significativo no desempenho do modelo, evidenciando como a preparação criteriosa dos dados pode aumentar a robustez e a capacidade

de generalização das redes neurais. Além disso, essa metodologia fornece uma abordagem base que pode ser utilizada e otimizada em futuros trabalhos neste domínio.

- A investigação estratégica de mecanismos de atenção e extração de características, avaliando diferentes configurações de camadas *stem* para otimizar a extração de características de baixo nível, como texturas, e explorando o posicionamento de mecanismos de atenção ao longo da arquitetura. Essa análise permitiu identificar configurações de arquiteturas competitivas para a classificação histopatológica de câncer de pênis.
- O desenvolvimento de uma arquitetura otimizada para o problema de classificação histopatológica de câncer de pênis, obtida por meio da aplicação de uma metodologia de Neural Architecture Search (NAS) baseada em células. Essa contribuição não apenas resultou em uma arquitetura competitiva e adaptada ao conjunto de dados, mas também evidenciou o valor da própria metodologia NAS como ferramenta de descoberta de modelos otimizados.

1.4 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 aborda a fundamentação teórica do problema e das técnicas utilizadas. Inicialmente, são discutidos os conceitos de análise histopatológica e o processamento de lâminas para biópsias. Em seguida, são apresentados os fundamentos do aprendizado profundo, com foco em redes neurais convolucionais (CNNs), Neural Architecture Search (NAS) e mecanismos de atenção como BAM e CBAM, destacando suas aplicações em diagnósticos histopatológicos. Também são apresentadas as métricas de avaliação utilizadas neste trabalho.
- O Capítulo 3 descreve trabalhos relacionados à análise de imagens histopatológicas e ao câncer de pênis em imagens histopatológicas, abordando as metodologias existentes que utilizam redes neurais convolucionais (CNNs), aprendizado profundo e mecanismos de atenção para diagnóstico médico.
- O Capítulo 4 detalha os materiais e a metodologia proposta, incluindo o ambiente experimental, a aquisição e descrição dos dados, balanceamento inicial do dataset, e análise do pré-processamento com aumento de representatividade. Também são descritas as etapas da pesquisa de arquitetura neural (NAS): a definição do espaço de busca, estratégia de busca e estratégia de avaliação, além dos hiper-parâmetros utilizados no treinamento e otimização das arquiteturas.

- O Capítulo 5 trata sobre os resultados obtidos e discussões em relação aos experimentos realizados, comparando as arquiteturas propostas com as arquiteturas-base utilizadas e trabalhos relacionados que utilizam o mesmo *dataset*, analisando as métricas de desempenho em diferentes ampliações das imagens histopatológicas, além da frequência de escolha do otimizador nas melhores arquiteturas candidatas. Também apresenta as limitações do trabalho desenvolvido e perspectivas futuras para sanar as limitações apresentadas.
- O Capítulo 6 apresenta as considerações finais sobre os resultados e artigos científicos resultantes desta pesquisa.

2 Fundamentação Teórica

Este capítulo apresenta os conceitos explorados para o desenvolvimento do estudo, fornecendo uma base teórica para compreender as técnicas utilizadas e as escolhas metodológicas realizadas. São discutidos aspectos relacionados ao câncer de pênis, análise histopatológica, patologia digital, redes neurais convolucionais, pesquisa de arquitetura neural e técnicas de aumento de dados, visando contextualizar e justificar as abordagens adotadas.

2.1 Câncer de Pênis

O câncer de pênis (CP) pode ser definido como uma neoplasia, ou seja, uma massa celular anormal apresentada no órgão sexual masculino. Dentre as neoplasias masculinas que afetam o pênis, o carcinoma de células escamosas (CCE) é responsável por mais de 95% das malignidades penianas, geralmente surgindo do epitélio do prepúcio interno ou da glândula (BROUWER et al., 2024). Outros tipos de neoplasias que afetam o órgão, todas menos comuns que o CCE, são lesões melanocíticas, tumores mesenquimais, linfomas e metástases que normalmente são de origem prostática, da bexiga urinária ou colorretal.

O câncer de pênis é um tipo de câncer raro em países desenvolvidos e mais frequente em países em desenvolvimento, podendo constituir até 10% das neoplasias malignas masculinas em algumas regiões africanas, asiáticas e sul-americanas (THOMAS et al., 2021). Sua ocorrência é altamente correlacionada com fatores socioculturais, sendo mais comum em países e regiões onde os padrões de renda e higiene são mais baixos (GIONA, 2022). Os principais fatores correlacionados ao CP são o tabagismo, alcoolismo, higiene precária do pênis, condições de inflamação crônica e infecção pelo Papilomavírus Humano (HPV) (JÚNIOR et al., 2022).

Conforme a Agência Internacional de Pesquisa em Câncer (IARC), a maior incidência de câncer de pênis (CP) foi registrada em Uganda, na África, com 5,6 casos por 100.000 homens. O Brasil ocupa a 13ª posição no ranking mundial, com uma taxa de incidência de 1,6 por 100.000 homens. No entanto, essa análise nacional não contempla adequadamente as disparidades regionais, especialmente nas áreas mais pobres do país. Estudos realizados entre 2006 e 2018 nas regiões norte e nordeste apontam uma taxa significativamente maior, de 5,4 por 100.000 homens. Particularmente, no estado do Maranhão, a incidência atinge 6,1 por 100.000 homens, superando os índices nacionais e internacionais mais elevados reportados pela IARC, destacando uma situação alarmante de CP no Maranhão (VIEIRA et al., 2020; JÚNIOR et al., 2022).

O diagnóstico do CP é realizado por meio de uma biópsia incisional de qualquer lesão peniana suspeita, permitindo a diferenciação entre lesões malignas, pré-cancerosas e benignas (INCA, 2022a). Esse procedimento consiste na remoção de um fragmento da lesão, os quais são posteriormente enviados para análise anatomopatológica. A avaliação inclui uma análise macroscópica e microscópica de células e tecidos, possibilitando a identificação precisa do tipo de lesão (benigna ou maligna) e suas características. É importante destacar que o exame histopatológico é o único método definitivo para diagnosticar tumores malignos, fornecendo informações cruciais para a escolha do tratamento mais adequado (INCA, 2022b; BROUWER et al., 2024).

O tratamento do CP consiste, principalmente, na remoção cirúrgica da lesão, por meio de penectomia parcial ou total, frequentemente associada à linfadenectomia inguinal bilateral (INCA, 2022a). A decisão entre a penectomia parcial ou total é determinada pela extensão da lesão, visando remover completamente o tumor, preservando o máximo possível do órgão sem comprometer o controle oncológico (BROUWER et al., 2024). Em casos de recidiva ou tumores irresssecáveis, a radioterapia e a quimioterapia podem ser utilizadas como tratamento paliativo, embora ainda não existam protocolos claramente benéficos para essas abordagens no CP (INCA, 2022a; BROUWER et al., 2024). Um estudo retrospectivo realizado no estado do Maranhão revelou que 39,5% dos pacientes apresentaram recidiva local e/ou metástase linfonodal, enquanto 60,5% não demonstraram evidência da doença após o tratamento cirúrgico (JÚNIOR et al., 2022).

O CP apresenta alta incidência em regiões subdesenvolvidas, com destaque alarmante para o estado do Maranhão, que pode figurar como a maior taxa de incidência no mundo. Esse cenário evidencia a urgência de intervenções eficazes tanto no diagnóstico quanto no tratamento da doença. A penectomia, associada à linfadenectomia, continua sendo o tratamento padrão para a remoção do tumor. Contudo, a confirmação por biópsia é obrigatória para autorização do procedimento, sendo o único método definitivo para distinguir lesões malignas de benignas.

2.2 Análise Histopatológica

Histopatologia é a ciência que estuda a patologia da doença por meio de alterações morfológicas que podem ser vistas microscopicamente (DEWI; PURWANTO; WIDJIATI, 2023). O processo inicia-se com a obtenção de uma amostra de tecido, sendo devidamente processada e preparada para análise microscópica. Essa análise é realizada por um patologista especializado, que busca identificar anormalidades morfológicas, alterações estruturais nos tecidos e características específicas que possam embasar um diagnóstico.

O processamento da amostra histopatológica segue etapas bem definidas. Primeiramente, a fixação preserva a integridade do tecido, prevenindo autólise e

degradação celular pós-morte. Em seguida, o tecido é embebido em parafina, conferindo rigidez necessária para cortes ultrafinos. Posteriormente, a amostra é seccionada em fatias de 3 a 5 micrômetros e corada para destacar suas estruturas internas. Por fim, a lâmina é montada com a remoção do excesso de corante e coberta com uma lamínula fixada com um adesivo impermeável, ficando pronta para a análise microscópica (DEWI; PURWANTO; WIDJIATI, 2023).

As imagens histopatológicas permitem identificar diferentes tipos de tecido e distinguir alterações significativas, como as diferenças entre tecido saudável e canceroso, ou entre tecido necrótico e sobrevivente. Além de fornecer suporte essencial para o diagnóstico, classificação e estadiamento de tumores malignos, a histopatologia orienta os cirurgiões durante a excisão do tecido canceroso, garantindo a preservação da função do órgão sempre que possível. Adicionalmente, os achados histopatológicos auxiliam na definição de tratamentos adjuvantes, desempenhando um papel central no planejamento terapêutico e no manejo adequado da doença (WANG et al., 2023).

O diagnóstico histopatológico, apesar de ser fundamental para a identificação e classificação de doenças, apresenta limitações importantes. Trata-se de um processo subjetivo e demorado, frequentemente suscetível a variações entre especialistas (GHANATE et al., 2011b). Cada patologista analisa minuciosamente as estruturas do tecido, buscando características específicas que embasem o diagnóstico. No entanto, esse processo demanda tempo, depende da experiência do especialista e pode apresentar reprodutibilidade variável, influenciada tanto pelo tipo de tecido quanto pelas características analisadas (DANO et al., 2020; SANZ-MOTILVA et al., 2024; ACANFORA et al., 2024).

Além disso, há uma distribuição desigual de especialistas em histopatologia, com maior concentração em regiões desenvolvidas. Em países subdesenvolvidos, observa-se um número consideravelmente menor de profissionais, agravando a dificuldade de acesso a diagnósticos precisos (WILSON; FLEMING, 2016). Essas limitações evidenciam a necessidade de métodos complementares que reduzam a subjetividade, acelerar o processo e ampliar o acesso ao diagnóstico em regiões carentes de recursos.

2.3 Patologia Digital

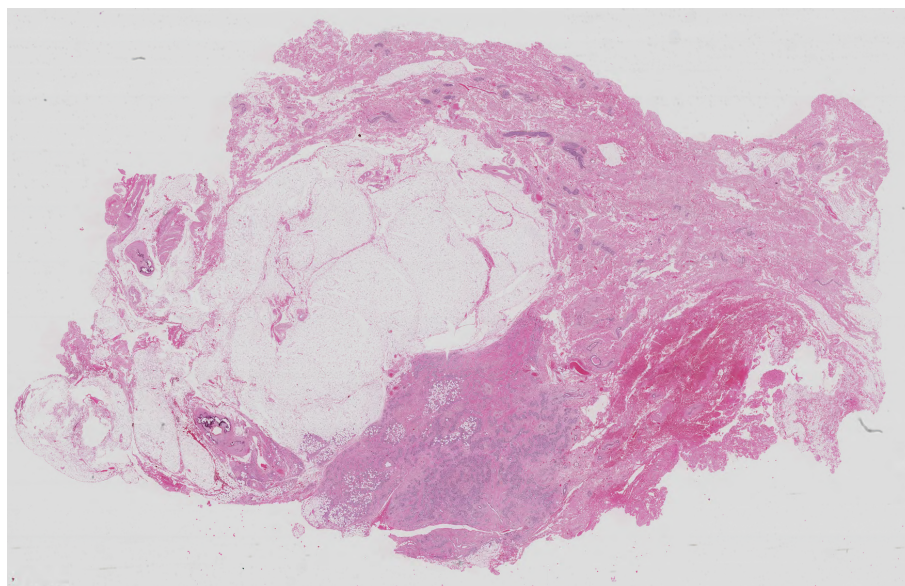
Os avanços tecnológicos contínuos remodelaram consistentemente as práticas de saúde. Na patologia, a implementação de tecnologias digitais desenvolve um novo campo interdisciplinar, a patologia digital. A patologia digital combina tecnologia, análise de dados e experiência clínica para transformar significativamente o diagnóstico médico, substituindo práticas convencionais baseadas em lâminas de vidro por abordagens digitais (KIRAN et al., 2023).

A patologia tradicional enfrenta diversas limitações, centrada na análise de lâminas

histológicas de vidro. A natureza física dessas lâminas apresenta desafios logísticos relacionados ao armazenamento e transporte, além de restringir o acesso a diagnósticos especializados devido à necessidade de presença física em locais específicos. Essa limitação afeta particularmente regiões carentes, onde o acesso a patologistas experientes é limitado. A análise manual das lâminas também está sujeita à variabilidade interobservadores e ao potencial de erro humano, introduzindo subjetividade que pode impactar diretamente a precisão do diagnóstico e as decisões de tratamento ([KIRAN et al., 2023](#)).

A patologia digital fornece abordagens digitais para superar essas limitações, desde a digitalização das lâminas em imagens de lâminas inteiras (WSI, do inglês *Whole slide imaging*) com resolução na ordem de gigapixels, armazenamento, recuperação e compartilhamento das informações necessárias ao diagnóstico, permitindo obter segundas opiniões com maior facilidade entre outros especialistas não limitando a proximidade física ([DAWSON, 2022](#)). A Figura 2 ilustra uma imagem de lâmina inteira frequentemente utilizada na patologia digital.

Figura 2 – Imagem de Lâmina Inteira



Fonte: [Aresta et al. \(2019\)](#).

O conjunto de dados gerados pelo fluxo de trabalho da patologia digital abre espaço para a patologia computacional. A Patologia Computacional (CPath) é uma ciência interdisciplinar que aumenta o desenvolvimento de abordagens computacionais para analisar e modelar imagens de histopatologia médica. O principal objetivo do CPath é desenvolver infraestrutura e fluxos de trabalho de diagnósticos digitais para patologia clínica, permitindo melhorias no diagnóstico e tratamento do câncer que as áreas mais abordadas pelas ferramentas CPath ([HOSSEINI et al., 2024](#)).

Com desenvolvimentos sucessivos em algoritmos de aprendizado profundo e visão computacional, e a facilidade do fluxo de dados da patologia digital, atualmente o CPath

está testemunhando uma mudança de paradigma (HOSSEINI et al., 2024). Os algoritmos de IA podem analisar conjuntos de dados extensos, detectar padrões e auxiliar os patologistas a fazer diagnósticos precisos. A integração da experiência humana e do aumento da IA podem resultar em uma diminuição nos erros de diagnóstico e uma aceleração do processo de diagnóstico (KIRAN et al., 2023).

2.4 Aprendizado Profundo

A inteligência artificial (IA) compreende técnicas que permitem aos computadores imitar o comportamento humano, reproduzindo ou superando a capacidade de tomada de decisão para resolver tarefas complexas de forma independente ou com mínima intervenção humana (RUSSELL; NORVIG, 2021). As primeiras aplicações de IA baseavam-se em abordagens fundamentadas em bases de conhecimento, nas quais regras explícitas eram codificadas em linguagens formais para o computador poder inferir conclusões. No entanto, essa abordagem apresentava limitações significativas, pois dependia de conjuntos de regras explícitas, que precisavam ser formuladas manualmente (JANIESCH; ZSCHECH; HEINRICH, 2021).

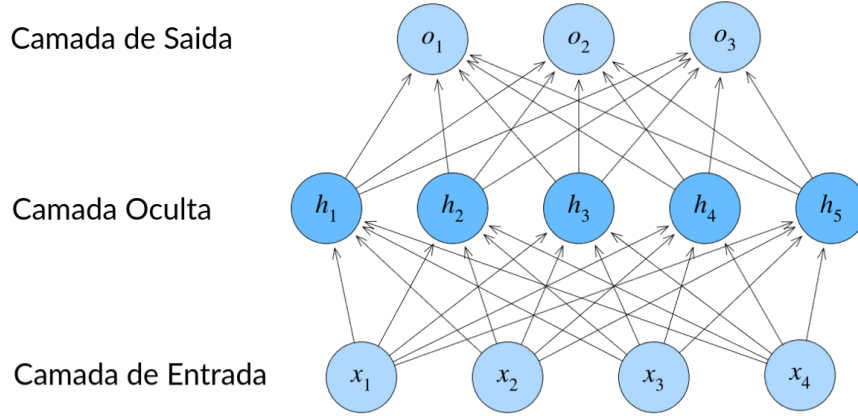
Dentro do campo da IA, o aprendizado de máquina busca estudar algoritmos e métodos que possam aprender regras e associações a partir dos dados, sem que elas precisem ser explicitamente programadas. Isso permite que os modelos aprendam automaticamente com os dados fornecidos (N; GUPTA, 2020). Entretanto, muitas técnicas de aprendizado de máquina possuem limitações na capacidade de lidar com dados complexos. O aprendizado profundo supera essas limitações ao utilizar redes neurais profundas, que empregam aprendizado hierárquico para extrair representações mais avançadas dos dados (SHARMA; SHARMA; JINDAL, 2021).

O aprendizado profundo é baseado em redes neurais artificiais compostas por múltiplas camadas. Essas camadas extraem representações simples dos dados nas camadas iniciais e, gradualmente, criam representações mais complexas em camadas subsequentes. Esse processo hierárquico aumenta a capacidade do modelo de aprender características complexas dos dados (SHARMA; SHARMA; JINDAL, 2021).

Na Figura 3, é apresentada uma representação gráfica de uma rede neural. A camada de entrada representa as dimensões dos dados que serão distribuídos para todos os neurônios da camada seguinte. Cada neurônio realiza uma combinação linear dos dados recebidos, gerando uma representação processada por uma função de ativação. Essa função introduz não-linearidade, permitindo que a rede capture relações complexas entre os dados e normalize os valores para melhorar o treinamento e a representatividade do modelo.

O processo de cálculo das representações avança camada por camada, até chegar à camada de saída, onde o modelo apresenta sua predição. Esse procedimento é chamado

Figura 3 – Rede Neural



Fonte: [Zhang et al. \(2023\)](#), adaptado pelo Autor.

de *feed forward propagation*. O aprendizado em redes neurais profundas é alcançado por meio do processo de *backpropagation*. Após o *feed forward propagation*, calcula-se a função de perda (também chamada de função custo), que quantifica a diferença entre a predição do modelo e o valor alvo esperado. Em seguida, essa perda é retro-propagada através das camadas, permitindo a atualização dos parâmetros da rede (pesos e *bias*).

Os pesos presentes nos neurônios costumam ser ajustados utilizando o método de descida de gradiente, o qual consiste em realizar iterações sucessivas para encontrar os valores de parâmetros que minimizam a função de perda ([LECUN et al., 1998](#)). Nesse processo, para cada peso w é calculada a derivada parcial da função de perda em relação a esse peso (o gradiente $\frac{\partial L}{\partial w}$). A partir dessa derivada, o peso é atualizado conforme a Equação 2.1:

$$w_{\text{novo}} \leftarrow w_{\text{antigo}} - \eta \frac{\partial L}{\partial w}, \quad (2.1)$$

onde η é a taxa de aprendizado (learning rate), um hiper-parâmetro que controla o tamanho do passo na direção do gradiente. Valores muito altos de η podem fazer com que o algoritmo não convirja, pois os passos podem ser largos demais, enquanto valores muito baixos podem tornar o aprendizado extremamente lento.

Adicionalmente, em muitos casos práticos utiliza-se uma variante do método de descida de gradiente chamada descida de gradiente estocástica (SGD, do inglês *Stochastic Gradient Descent*) ou *mini-batch gradient descent*, onde apenas um subconjunto dos dados de treinamento (*batch* ou *mini-batch*) é usado a cada iteração. Essas estratégias visam acelerar o aprendizado e atenuar possíveis problemas de convergência em grandes conjuntos de dados ([LECUN et al., 1998](#)).

Durante a retro-propagação, a regra da cadeia é aplicada para decompor o gradiente

em cada camada, permitindo o ajuste apropriado dos pesos. A regra da cadeia é apresentada na Equação 2.2, onde $a^{(l)}$ é a ativação na camada l , $z^{(l)}$ é a entrada ponderada na camada l .

$$\frac{\partial L}{\partial w^{(l)}} = \frac{\partial L}{\partial a^{(l)}} \cdot \frac{\partial a^{(l)}}{\partial z^{(l)}} \cdot \frac{\partial z^{(l)}}{\partial w^{(l)}}, \quad (2.2)$$

Esse ajuste iterativo faz com que a rede reduza gradualmente o erro, melhorando sua capacidade de generalizar com base nos dados de treinamento. O processo prossegue até que um critério de parada seja atingido, como um número máximo de épocas ou a estabilização da função de perda.

2.4.1 Redes Neurais Convolucionais

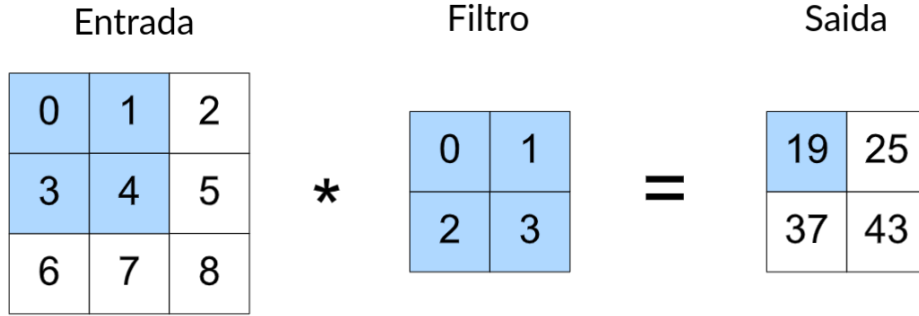
As redes neurais convolucionais (CNNs, do inglês *convolutional neural network*) são um tipo específico de rede neural projetada para processar dados organizados em grade, como imagens e vídeos. Inspiradas na organização do córtex visual humano, as CNNs têm ampla aplicação na extração e análise de características complexas em dados visuais (LI et al., 2022b). Sua arquitetura baseia-se na aplicação de convoluções para identificar padrões locais, como bordas, texturas e formas, essenciais para tarefas como classificação de imagens, detecção de objetos e segmentação (LECUN; BENGIO; HINTON, 2015). Tradicionalmente, as CNNs são compostas por camadas convolucionais, camadas de pooling, normalização em lote, funções de ativação (como ReLU) e camadas totalmente conectadas (KRICHEN, 2023).

As camadas convolucionais são o alicerce das CNNs, responsáveis por extrair características geométricas das imagens, como bordas e texturas. Essas camadas utilizam filtros, que são pequenas matrizes aplicadas de forma deslizante sobre a imagem. Quando o padrão presente no filtro é semelhante ao padrão na imagem, a multiplicação escalar gera um valor positivo significativo. O resultado dessas operações forma uma matriz denominada mapa de características, enviada para camadas subsequentes, onde características mais complexas são extraídas ao longo da rede. A operação da camada convolucional é ilustrada na Figura 4.

As camadas convolucionais podem produzir valores negativos, que não são úteis para o aprendizado. Para resolver isso, utiliza-se uma função de ativação, que suprime valores indesejados e adiciona não-linearidade ao modelo. A introdução de não-linearidade é essencial para que a rede possa aprender padrões complexos e representações mais abstratas dos dados. As funções de ativação também ajudam a normalizar os valores, melhorando a eficiência do treinamento e a representatividade do modelo.

Outro desafio nas camadas convolucionais é a alta covariância dos resultados, que pode dificultar o aprendizado nas camadas subsequentes, pois os valores dos mapas de

Figura 4 – Operação de Convolução



Fonte: [Zhang et al. \(2023\)](#), adaptado pelo Autor.

características podem variar constantemente devido a mudanças nos dados de entrada. A normalização em lote (*Batch Normalization*) mitiga esse problema ao normalizar as ativações de cada camada. O *Batch Normalization* está descrito na Equação 2.3, onde x representa a entrada da camada que está sendo normalizada, $\hat{\mu}_B$ é a média da amostra ao longo do *batch* de treinamento, $\hat{\sigma}_B$ representa o desvio padrão da amostra.

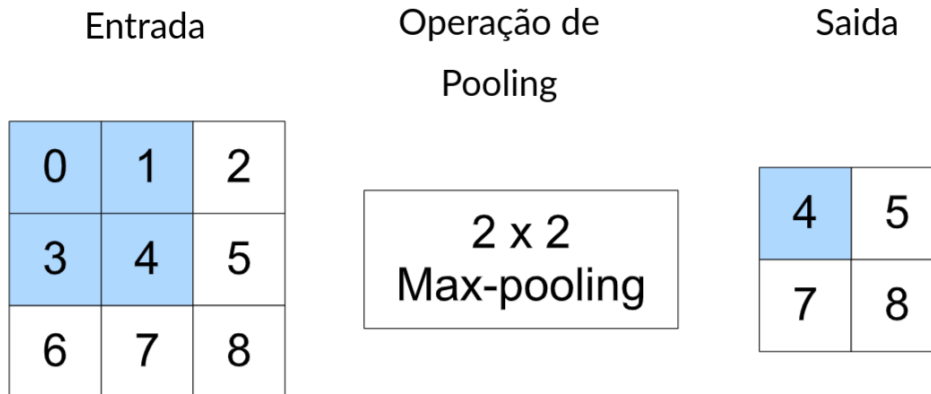
$$\text{BN}(\mathbf{x}) = \gamma \odot \frac{\mathbf{x} - \hat{\mu}_B}{\hat{\sigma}_B} + \beta. \quad (2.3)$$

Os parâmetros γ e β são aprendidos e permitem à rede ajustar a escala e o deslocamento das ativações normalizadas, respectivamente. Essa etapa de normalização assegura que as ativações tenham média zero e variância unitária, enquanto os parâmetros γ e β permitem que a rede recupere a representatividade original dos dados, se necessário.

As camadas de *pooling* são usadas para resumir as características extraídas, reduzindo sua dimensionalidade ao calcular o máximo ou a média de regiões sem sobreposição. Essa operação reduz a complexidade computacional, aumenta a robustez da rede a pequenas translações e amplia o campo receptivo das camadas subsequentes. A camada de *pooling* máximo é representada graficamente na Figura 5.

As camadas totalmente conectadas são compostas por neurônios interligados, responsáveis por sintetizar as características extraídas pelas camadas convolucionais e produzir o resultado da rede. Essas camadas mapeiam as características extraídas para o conjunto de classes esperado na saída, desempenhando um papel crucial na etapa de predição.

Com o avanço das CNNs, arquiteturas cada vez mais profundas foram desenvolvidas. Contudo, a profundidade excessiva introduziu o problema de vanishing gradient, no qual o sinal do gradiente se perde ao longo da retropropagação, limitando a capacidade de

Figura 5 – Camada de *Pooling* Máximo

Fonte: [Zhang et al. \(2023\)](#), adaptado pelo Autor.

aprendizado em redes muito profundas. Para mitigar esse problema, foram propostas soluções como saídas intermediárias, que fornecem supervisão adicional ao longo das camadas, e conexões residuais (*skip connections*), que permitem que o gradiente contorne camadas específicas, facilitando a propagação do sinal e melhorando o desempenho da rede ([NOH, 2021](#)).

2.4.2 Arquitetura Resnet

A arquitetura ResNet (*Residual Network*) é uma das mais importantes no campo do aprendizado profundo, especialmente para redes neurais convolucionais. Proposta por [He et al. \(2016\)](#), a ResNet foi projetada para superar problemas relacionados ao treinamento de redes muito profundas, como o *vanishing gradient* (desaparecimento do gradiente). Esse problema ocorre quando o gradiente se torna extremamente pequeno, dificultando a atualização eficiente dos pesos nas camadas iniciais da rede.

O diferencial da ResNet está na introdução das conexões residuais ou *skip connections*. Essas conexões permitem que a saída de uma camada seja somada diretamente à saída de uma camada subsequente, contornando uma ou mais camadas intermediárias. O conceito pode ser descrito pela Equação 2.4, onde x é a entrada do bloco, $F(x)$ é a transformação aprendida pelo bloco intermediário, e y é a saída final:

$$y = F(x) + x \quad (2.4)$$

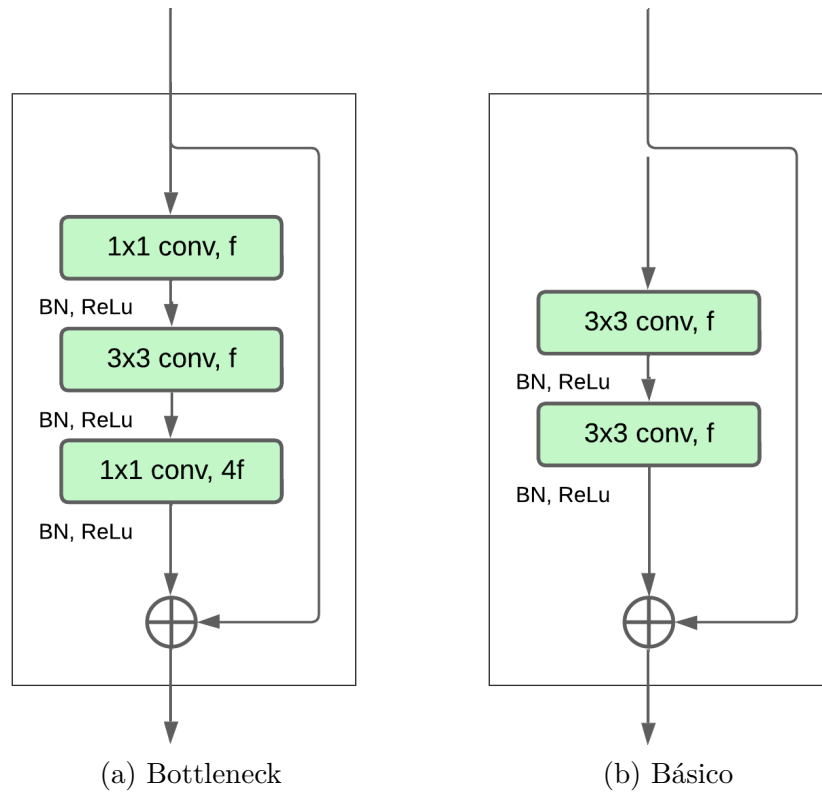
Essas conexões auxiliam na propagação do gradiente durante o treinamento, permitindo que a informação flua diretamente pelas camadas da rede. Isso reduz significativamente o impacto do *vanishing gradient*, possibilitando o treinamento eficaz de redes com centenas ou até milhares de camadas.

A ResNet é construída com base em blocos residuais, cuja característica central é a presença das conexões de atalho. Existem dois tipos principais de blocos:

- Bloco Básico: Utilizado em redes mais rasas, como a ResNet-18 e a ResNet-34. Esse bloco é composto por duas camadas convolucionais seguidas de normalização em lote e ativação ReLU, conectadas pela soma direta da entrada do bloco com sua saída;
- Bloco *Bottleneck*: Empregado em redes mais profundas, como a ResNet-50, ResNet-101 e ResNet-152. Esse bloco introduz convoluções 1×1 para reduzir e expandir a dimensionalidade, facilitando o treinamento de redes maiores ao diminuir a complexidade computacional.

A estrutura modular dos blocos residuais permite à rede extrair e combinar características em diferentes níveis de complexidade ao longo da arquitetura. A Figura 6 ilustra as estruturas básicas dos blocos residuais.

Figura 6 – Blocos Resnet



Fonte: elaborada pelo Autor.

A ResNet é organizada em cinco estágios convolucionais, começando pelo bloco de *stem*, responsável por capturar características de baixo nível, como bordas e linhas. O bloco de *stem* inclui uma camada convolucional 7×7 , seguida por normalização em lote, ativação ReLU e *max pooling*. Além de extrair características iniciais, o *stem* reduz a resolução da entrada, diminuindo a complexidade computacional das etapas subsequentes.

Os estágios subsequentes são compostos por sequências de blocos residuais, com uma camada de *pooling* ao final de cada estágio para reduzir as dimensões dos mapas de características. O número de blocos em cada estágio varia conforme a variante da ResNet:

- ResNet-18 e ResNet-34: Utilizam blocos básicos, adequados para problemas menos complexos devido à menor profundidade;
- ResNet-50, ResNet-101 e ResNet-152: Utilizam blocos *bottleneck*, sendo mais profundas e projetadas para lidar com problemas complexos e conjuntos de dados maiores.

Cada estágio da ResNet aumenta o número de filtros aplicados, dobrando a quantidade a cada novo estágio. Essa hierarquia permite que as camadas iniciais capturem características simples, enquanto as camadas mais profundas processam informações complexas, culminando na camada de saída, onde o modelo realiza a predição final.

2.4.3 Mecanismos de atenção

Os mecanismos de atenção são métodos utilizados para direcionar o foco do modelo para as regiões mais relevantes de uma imagem, desconsiderando partes irrelevantes. Esse conceito é inspirado no sistema visual humano, que seleciona áreas de interesse em cenas complexas para facilitar sua interpretação. Semelhantemente, os mecanismos de atenção foram introduzidos nos sistemas de visão computacional para melhorar o desempenho em tarefas como classificação, segmentação e detecção de objetos (GUO et al., 2022).

Um mecanismo de atenção funciona ponderando dinamicamente as características extraídas da entrada, atribuindo maior peso às regiões mais relevantes. Esse processo é realizado por meio de camadas que aprendem a importância relativa das características e geram um mapa de atenção. O mapa resultante é multiplicado pelas características originais, suprimindo elementos irrelevantes e destacando os mais importantes para a predição.

Os mecanismos de atenção variam conforme o objetivo da rede neural. Neste trabalho, destaca-se o uso de atenção por canal e atenção espacial. Em mapas de características, diferentes canais frequentemente representam diferentes tipos de objetos ou padrões. A atenção por canal recalibra os pesos atribuídos a cada canal, funcionando como um processo seletivo de objetos, decidindo quais canais merecem maior destaque (GUO et al., 2022). Já a atenção espacial atua como um mecanismo adaptativo que seleciona regiões específicas da entrada, determinando “onde” prestar atenção.

A atenção espacial é tipicamente implementada por meio de camadas convolucionais progressivamente menores, que produzem um único mapa de características indicando

as regiões mais importantes. Esse mapa é replicado para corresponder à quantidade de canais originais e, em seguida, multiplicado pelas características de entrada, resultando em mapas de características ponderadas espacialmente.

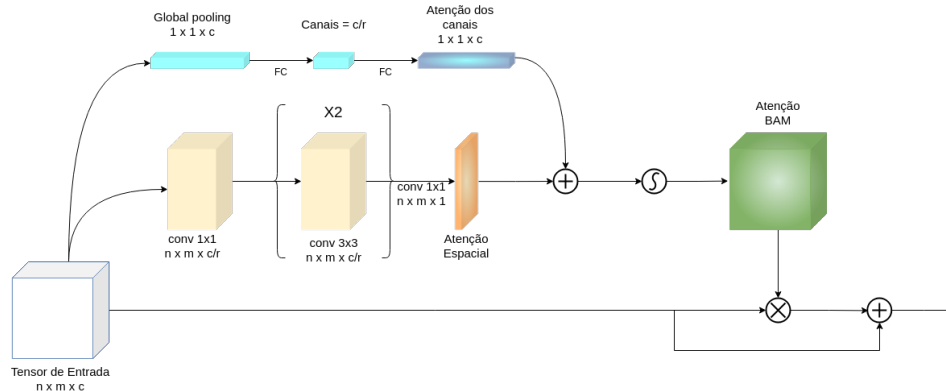
Neste trabalho, utilizam-se os mecanismos de atenção BAM e CBAM, ambos foram selecionados por utilizarem a combinação de atenção por canal e atenção espacial. A principal diferença entre eles está na forma como essas operações são estruturadas: no BAM, as atenções são aplicadas em paralelo, enquanto no CBAM são aplicadas sequencialmente.

2.4.3.1 BAM: *Bottleneck Attention Module*

O *Bottleneck Attention Module* (BAM) foi proposto por [Park et al. \(2018\)](#) e projetado para ser integrado com arquiteturas residuais. Sua adição reduz o erro do modelo e melhora a ativação do gradiente em áreas relevantes para a predição.

O BAM implementa a atenção por canal e a atenção espacial paralelamente, como ilustrado na Figura 7. A atenção por canal é baseada em um mecanismo similar ao *Squeeze-and-Excitation* ([HU et al., 2019](#)), no qual uma transformação de *pooling* global seguida de uma *multilayer perceptron* (MLP) gera os pesos de atenção por canal. Por outro lado, a atenção espacial é obtida por camadas convolucionais sucessivas, culminando em uma convolução 1×1 para gerar o mapa de atenção. Ambas as atenções são expandidas para ter dimensões compatíveis antes de serem combinadas por soma e multiplicação.

Figura 7 – Mecanismo de Atenção BAM



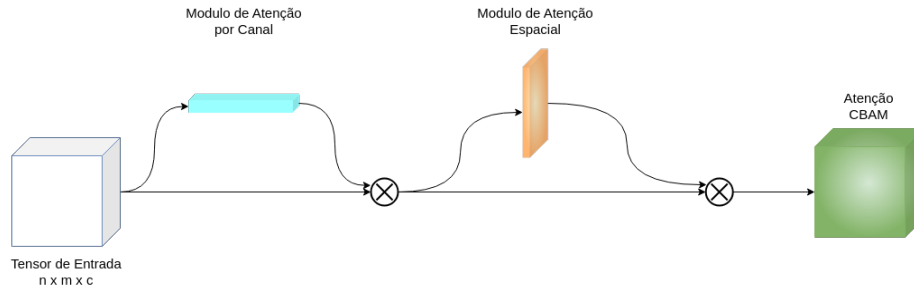
Fonte: elaborada pelo Autor.

2.4.3.2 CBAM: *Convolutional Block Attention Module*

O CBAM, proposto por [Woo et al. \(2018b\)](#), também combina atenção por canal e atenção espacial, mas organiza essas operações sequencialmente. Essa sequência é representada graficamente na Figura 8.

O CBAM calcula a atenção por canal aplicando *pooling* médio e máximo às características, cujos resultados são passados por uma MLP compartilhada. Os vetores

Figura 8 – Mecanismo de Atenção CBAM



Fonte: elaborada pelo Autor.

resultantes são somados e ativados por uma função sigmoide, gerando o mapa de atenção por canal. A atenção espacial, por sua vez, é derivada do *pooling* médio e máximo aplicados ao longo dos canais, criando matrizes que, quando concatenadas e convoluídas por 1×1 , produzem o mapa de atenção espacial. Assim como no BAM, os mapas de atenção são ajustados para compatibilidade dimensional antes da multiplicação com as características originais.

2.5 Pesquisa de Arquitetura Neural

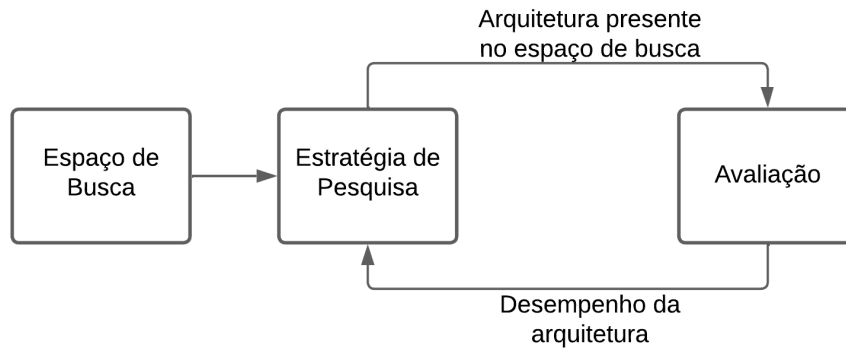
O desempenho das redes neurais profundas depende de dois fatores principais: a arquitetura da rede e os pesos associados. Ambos precisam atingir condições ideais para alcançar resultados satisfatórios. Contudo, a obtenção de arquiteturas ótimas não pode ser diretamente formulada como uma função contínua ao necessitar de uma métrica explícita que guie o processo de design (LIU et al., 2023). A Pesquisa de Arquitetura Neural (NAS, do inglês *neural architecture search*) surge como uma solução para automatizar o design de arquiteturas de redes neurais, adaptando-as a tarefas específicas e conjuntos de dados, minimizando a intervenção humana e substituindo processos manuais dependentes do conhecimento dos desenvolvedores (CHITTY-VENKATA et al., 2022).

A metodologia NAS é composta por três etapas fundamentais: delimitação do espaço de busca, estratégia de busca e estratégia de avaliação (ELSKEN; METZEN; HUTTER, 2019). A Figura 9 apresenta uma visão geral do processo.

O espaço de busca engloba todas as arquiteturas possíveis que podem ser geradas pela metodologia NAS. Essa etapa desempenha um papel crítico no desempenho final ao definir os limites superiores da otimização que o NAS pode alcançar (REN et al., 2021). A qualidade do espaço de busca é diretamente proporcional à eficiência e eficácia da metodologia. No entanto, um espaço muito amplo pode introduzir complexidade excessiva e alternativas pouco promissoras.

A correta delimitação do espaço de busca pode ser auxiliada pela experiência no

Figura 9 – Metodologia NAS



Fonte: elaborada pelo Autor.

domínio da tarefa em questão, permitindo a exclusão de arquiteturas menos relevantes. Além disso, a abordagem de busca modular, como o uso de espaços de busca baseados em células ou blocos, é amplamente adotada. Nessas estratégias, pequenas estruturas modulares são projetadas e repetidamente empilhadas para compor a arquitetura final. Essa modularidade reduz significativamente a complexidade do processo de busca, tornando-o mais eficiente (REN et al., 2021).

Após a definição do espaço de busca, é necessário aplicar uma estratégia para identificar as arquiteturas mais promissoras. A maioria dos métodos NAS utiliza abordagens como busca em grade, busca aleatória, otimização baseada em gradiente ou Otimização Bayesiana (KANG et al., 2023).

A Otimização Bayesiana destaca-se por sua capacidade de explorar eficientemente funções cujo custo de observação é elevado, como ocorre no problema do NAS. Durante o processo, cada arquitetura avaliada envolve treinamento e validação, etapas que consomem muitos recursos computacionais. Essa abordagem é especialmente útil para reduzir o número de avaliações necessárias, maximizando o retorno do processo de busca (JIN; SONG; HU, 2019).

Na presente pesquisa, foi adotado o algoritmo bayesiano TPE (*Tree-structured Parzen Estimator*) introduzido em Bergstra et al. (2011), um método de otimização bayesiana amplamente utilizado para ajustar hiper-parâmetros em diversos domínios, incluindo inteligência artificial, ciência da computação e física. Ao contrário de métodos tradicionais que usam Processos Gaussianos para prever diretamente o desempenho $p(y|x)$, o TPE segue uma abordagem inversa, modelando a relação $p(x|y)$ entre configurações e desempenho, além de $p(y)$. O espaço de configurações X , que possui uma estrutura hierárquica (como escolher primeiro o número de camadas de uma rede neural e depois seus parâmetros), é adaptado pelo TPE substituindo distribuições fixas por distribuições adaptativas mais flexíveis. Para variáveis numéricas ou categóricas, o algoritmo ajusta automaticamente modelos probabilísticos, capazes de representar melhor

regiões promissoras do espaço de busca.

A densidade $p(x|y)$ é definida Equação 2.5, onde $\ell(x)$ é construída a partir de observações $\{x^{(i)}\}$ com perda $y < y^*$ (configurações promissoras) e $g(x)$ utiliza as demais observações. O limiar y^* , crítico para separar as densidades, é definido como o quantil γ das perdas observadas, garantindo $p(y < y^*) = \gamma$. Essa abordagem permite ao TPE priorizar regiões de alto desempenho enquanto mantém eficiência computacional, escalando linearmente com o número de observações $|H|$ e dimensões do espaço.

$$p(x|y) = \begin{cases} \ell(x) & \text{se } y < y^* \\ g(x) & \text{se } y \geq y^* \end{cases} \quad (2.5)$$

Inicialmente são realizadas observações aleatórias para explorar de maneira ampla o espaço de busca e fornecer uma base inicial para o modelo $\ell(x)$ e $g(x)$. Na etapa de otimização, o TPE utiliza essas densidades para amostrar configurações com maior probabilidade de minimizar a perda y . Esse processo iterativo ajusta progressivamente as recomendações, concentrando-se em regiões promissoras identificadas pelo contraste entre $\ell(x)$ e $g(x)$, o que o torna especialmente adequado para problemas de alta dimensionalidade e orçamentos reduzidos de avaliação, como no treinamento de redes neurais profundas.

A avaliação das arquiteturas candidatas é um componente essencial para orientar o processo de busca. O método mais simples consiste em realizar o treinamento e a validação padrão da arquitetura nos dados. No entanto, esse procedimento é computacionalmente caro, limitando o número de arquiteturas que podem ser exploradas (ELSKEN; METZEN; HUTTER, 2019).

Além de medir o desempenho das arquiteturas, a estratégia de avaliação também desempenha um papel crucial ao fornecer um retorno ao algoritmo de busca. Esse feedback orienta a otimização e auxilia na priorização de arquiteturas promissoras.

O NAS, ao automatizar a geração de arquiteturas otimizadas, visa identificar redes que apresentem alto desempenho em dados não vistos, atendendo às demandas de tarefas específicas com eficiência e escalabilidade. No entanto, o sucesso dessa metodologia depende de decisões criteriosas em todas as suas etapas, desde a definição do espaço de busca até as estratégias de avaliação.

2.6 Preprocessamento e Aumento de Dados

Sistemas de aprendizado profundo utilizam redes neurais artificiais para uma ampla variedade de aplicações, sendo treinados e testados com conjuntos de dados específicos. No entanto, as amostras de imagens brutas presentes nesses conjuntos frequentemente contêm ruídos e informações pouco claras, decorrentes de fatores diversos. O pré-processamento

é uma etapa fundamental que auxilia as redes de aprendizado profundo a interpretar melhor as características das imagens, reduzindo ruídos e inconsistências. Entretanto, há um risco significativo de perda de informações relevantes durante o pré-processamento, especialmente devido à aplicação de filtros inadequados ou excessivos (KORESH, 2024).

O aumento de dados (*data augmentation*) é uma técnica essencial para enriquecer a diversidade de conjuntos de dados e melhorar a robustez dos modelos de aprendizado profundo. Ele pode ser classificado em duas categorias principais: distorção de dados e sobreamostragem. Exemplos de técnicas de distorção incluem cortes aleatórios, inversão horizontal e ajustes de cor. Já a sobreamostragem é frequentemente usada para balanceamento de classes, criando dados sintéticos das classes minoritárias para aumentar sua representatividade durante o treinamento (SHORTEN; KHOSHGOFTAAR, 2019; DABLAIN et al., 2023).

Na histopatologia digital, as técnicas de normalização de manchas são amplamente utilizadas para minimizar artefatos de cores que podem surgir durante o manuseio das lâminas histológicas. Fatores como diferenças entre fabricantes de corantes, variações em lotes, alterações na concentração ou envelhecimento do corante podem impactar o padrão de coloração das lâminas. Essas variações podem interferir na consistência das análises realizadas por patologistas. A normalização de manchas busca padronizar o esquema de cores em todas as imagens, melhorando a confiabilidade e a uniformidade das análises (HOQUE et al., 2024). Além disso, abordagens baseadas em aprendizado de máquina são empregadas para realçar características específicas de imagens histopatológicas, otimizando o processo de diagnóstico (LAUANDE. et al., 2022; BELFORT et al., 2023).

Técnicas tradicionais de aumento de dados, como rotação, translação, inversão e perturbações de cor, são aplicadas com sucesso em imagens histopatológicas. No entanto, métodos específicos para este domínio também foram desenvolvidos. Uma dessas abordagens é o aumento de manchas (*stain augmentation*), que simula as variações de cor frequentemente observadas nesse tipo de imagem. Esse método permite que os modelos de aprendizado profundo generalizem esses artefatos, tornando-os mais robustos às variações cromáticas inerentes às lâminas histológicas (TELLEZ et al., 2019).

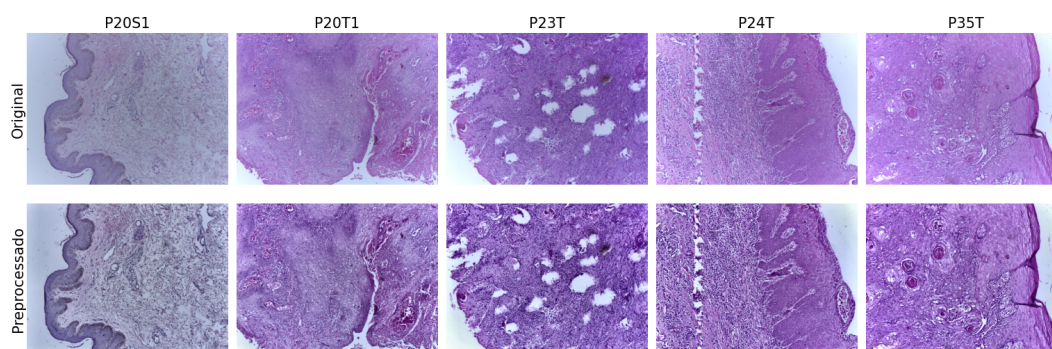
A seleção de uma combinação ideal de pré-processamento e aumento de dados é um desafio significativo, pois ambas as etapas afetam o desempenho dos modelos de maneiras distintas. Além disso, seu impacto pode variar conforme o conjunto de dados utilizado. Uma combinação específica pode apresentar um desempenho superior em um conjunto de dados, mas não necessariamente em outro, exigindo testes experimentais para identificar a abordagem mais adequada. Assim, a escolha da melhor tupla pré-processamento/aumento de dados requer experimentação e análise cuidadosa para atender às necessidades específicas de cada aplicação.

2.6.1 Pré-processamento

O pré-processamento dos dados é uma etapa fundamental para garantir a qualidade e a consistência das imagens histopatológicas antes de sua análise e classificação. Diversas técnicas de pré-processamento descritas na literatura foram aplicadas neste trabalho para aprimorar as características das imagens, melhorar o contraste, padronizar as cores e realçar estruturas relevantes. As abordagens utilizadas neste trabalho foram a equalização de canais YUV, equalização de canais cieLAB, equalização de canais H&E, normalização Vahadane, normalização stainGAN e normalização stainNet.

O método de **Equalização de canais YUV**, proposto em [Lauande et al. \(2022\)](#), foi desenvolvido para normalizar diferenças na distribuição de luz nas imagens histopatológicas. A técnica utiliza o algoritmo CLAHE (Contrast Limited Adaptive Histogram Equalization), originalmente descrito em [Zuiderveld \(1994\)](#). Nesse método, as imagens são inicialmente convertidas do espaço de cores RGB para o espaço YUV. Essa transformação é justificada pela característica do espaço RGB, que distribui informações de cor em todos os canais, podendo gerar artefatos indesejados após a equalização. No espaço YUV, a equalização é aplicada apenas ao canal Y, responsável pela intensidade luminosa, preservando as informações de cor nos canais U e V. Após a equalização, as imagens são reconvertidas para o espaço RGB. Na Figura 10 são apresentados exemplos de imagens preprocessadas utilizando a equalização de canais YUV.

Figura 10 – Equalização de canais YUV

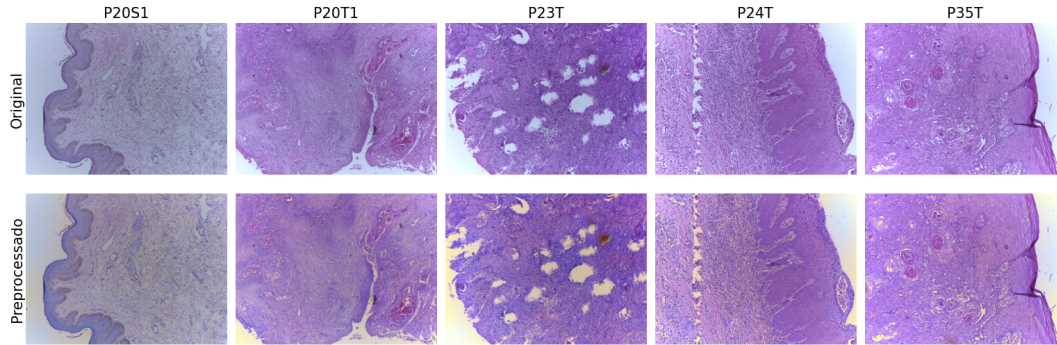


Fonte: elaborada pelo Autor.

Outra técnica empregada é a **Equalização de canais cieLAB**, descrita em [Belfort et al. \(2023\)](#), que visa realçar o contraste das estruturas relevantes para classificação histopatológica, como bordas em imagens negativas para câncer. Essas bordas, em imagens sem câncer, apresentam tamanho e forma mais consistentes em relação às imagens de casos positivos. A técnica consiste em converter as imagens do espaço RGB para o espaço cieLAB, composto pelos componentes L (luminosidade), A (verde a vermelho) e B (azul a amarelo). A equalização de histograma é aplicada exclusivamente ao componente B, realçando as variações cromáticas que contribuem para o diagnóstico. Posteriormente, a imagem é

convertida novamente para o espaço RGB, destacando áreas essenciais para a análise [Belfort et al. \(2023\)](#). Na Figura 11 são apresentados exemplos de imagens preprocessadas utilizando a equalização de canais cieLAB.

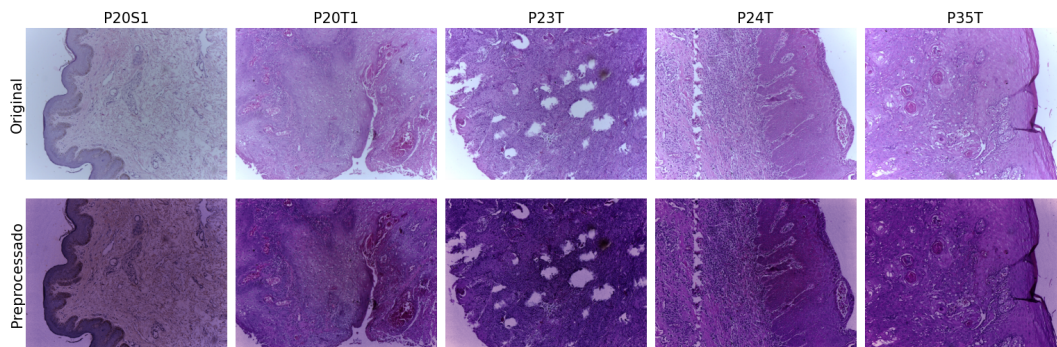
Figura 11 – Equalização de canais cieLAB



Fonte: elaborada pelo Autor.

Foi explorada a **Equalização de Canais H&E**. Nesse método, as imagens são decompostas em um espaço de cor onde os canais representam Hematoxilina, Eosina e Diaminobenzidina (DAB). A técnica aplicada no presente trabalho consiste na equalização de histograma individualmente a cada canal, destacando as características estruturais do tecido. Após a equalização, as imagens são reconvertidas para o espaço RGB, com o intuito de melhorar a visualização das áreas de interesse. Na Figura 12 são apresentados exemplos de imagens preprocessadas utilizando a equalização de canais H&E.

Figura 12 – Equalização de canais H&E



Fonte: elaborada pelo Autor.

A equalização de canais H&E consiste na conversão da imagem do espaço RGB para o espaço H&E, onde os canais representam Hematoxilina (H), Eosina (E) e Diaminobenzidina (D). A conversão é realizada conforme a Equação 2.6, onde M é a matriz de transformação do espaço RGB para H&E, estabelecida em [Ruifrok e Johnston \(2001\)](#).

$$I_{HED} = M \cdot I_{RGB} \quad (2.6)$$

Após a conversão, a equalização de histograma CLAHE (Contrast Limited Adaptive Histogram Equalization) é aplicada individualmente a cada canal H, E, D , conforme descrito na Equação 2.7, onde $\text{CLAHE}(\cdot)$ representa a aplicação do algoritmo de equalização adaptativa:

$$H' = \text{CLAHE}(H), \quad E' = \text{CLAHE}(E), \quad D' = \text{CLAHE}(D) \quad (2.7)$$

A aplicação do CLAHE pretende realçar as estruturas relevantes do tecido, melhorando o contraste e facilitando a distinção entre regiões de interesse. Para garantir que os valores resultantes permaneçam em um intervalo válido, os canais são normalizados conforme a Equação 2.8:

$$H'' = \text{clip}(H', 0, 255), \quad E'' = \text{clip}(E', 0, 255), \quad D'' = \text{clip}(D', 0, 255) \quad (2.8)$$

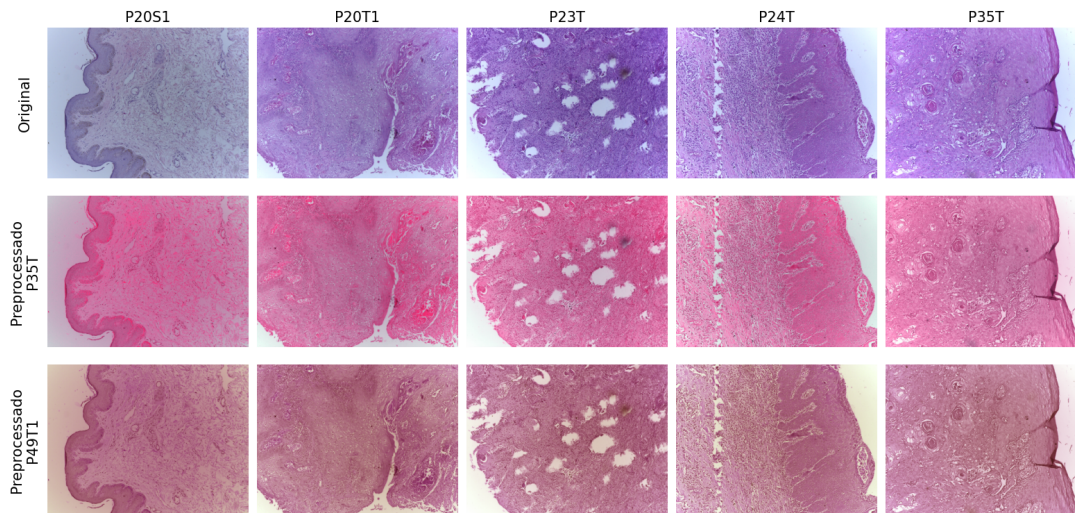
Por fim, a imagem é reconvertida para o espaço RGB utilizando a transformação inversa descrita na Equação 2.9, onde M^{-1} é a matriz de conversão inversa do espaço H&E para RGB:

$$I'_{RGB} = M^{-1} \cdot I''_{HED} \quad (2.9)$$

A **Normalização Vahadane**, proposta em Vahadane et al. (2016), foi implementada para padronizar artefatos de cor em imagens histopatológicas, permitindo análises consistentes e comparáveis. Esse método requer uma imagem de referência para obter os parâmetros de normalização, aplicando o padrão de cores da referência em outras imagens. No estudo de Vale et al. (2024), que propõe arquiteturas baseadas em DenseNet para classificação de câncer de pênis, foi realizada uma busca exaustiva pela melhor imagem de referência. Cada imagem do conjunto foi testada como alvo de normalização em validação cruzada, demonstrando a influência do padrão de cores no desempenho dos modelos. No entanto, devido ao caráter exploratório do presente estudo, a metodologia foi simplificada, aplicando-se o pré-processamento apenas às melhores imagens identificadas para cada arquitetura, especificamente P49T1 e P35T. Na Figura 13 são apresentados exemplos de imagens préprocessadas utilizando a normalização Vahadane com as imagens de referência P49T1 e P35T.

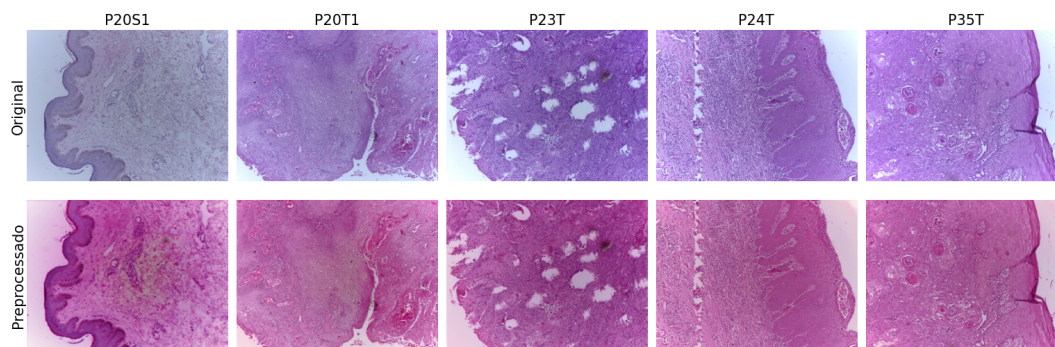
Outra abordagem utilizada foi a **Normalização stainGAN**, proposta por Shaban et al. (2019), que utiliza CycleGANs para transferência de estilo, permitindo a normalização de manchas sem a necessidade de uma imagem de referência. Durante o treinamento, a

Figura 13 – Normalização Vahadane



arquitetura aprende um padrão de cores e aplica essa padronização às imagens de entrada. Esse método foi testado em um sistema CAD para diagnóstico de câncer de mama, resultando em um aumento de 12% no AUC. Além disso, superou métodos tradicionais de normalização, como os propostos por [Vahadane et al. \(2016\)](#) e [Macenko et al. \(2009\)](#), em métricas de similaridade estrutural, preservando características do tecido enquanto padroniza as cores. Na Figura 14 são apresentados exemplos de imagens preprocessadas utilizando a normalização stainGAN.

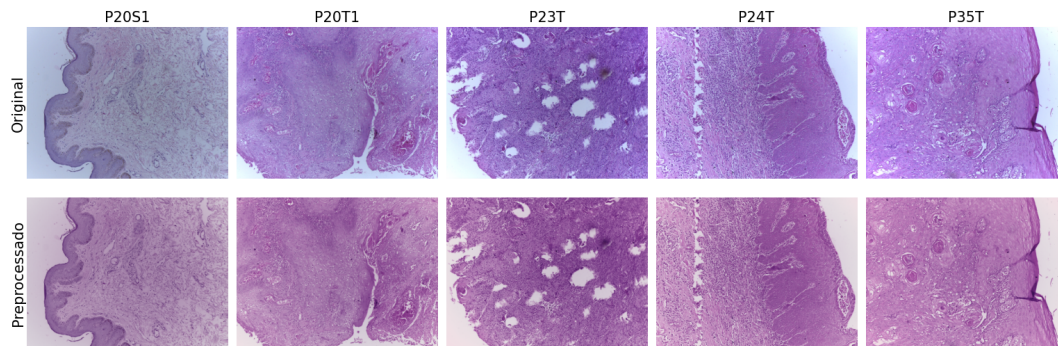
Figura 14 – Normalização stainGAN



Fonte: elaborada pelo Autor.

Por fim, foi empregada a **Normalização stainNet**, proposta por [Kang et al. \(2021\)](#), que segue a mesma premissa da StainGAN, mas com melhorias de eficiência computacional. Utilizando apenas camadas convolucionais 1×1 , a StainNet reduz significativamente o tempo de predição, sendo 40 vezes mais rápida que a StainGAN, sem comprometer a qualidade dos resultados. Em algumas situações, a StainNet apresentou desempenho superior, tornando-se uma alternativa prática para normalização de manchas em tempo real. Na Figura 15 são apresentados exemplos de imagens preprocessadas utilizando a normalização stainNet.

Figura 15 – Normalização stainNet



Fonte: elaborada pelo Autor.

2.6.2 Aumento de Dados

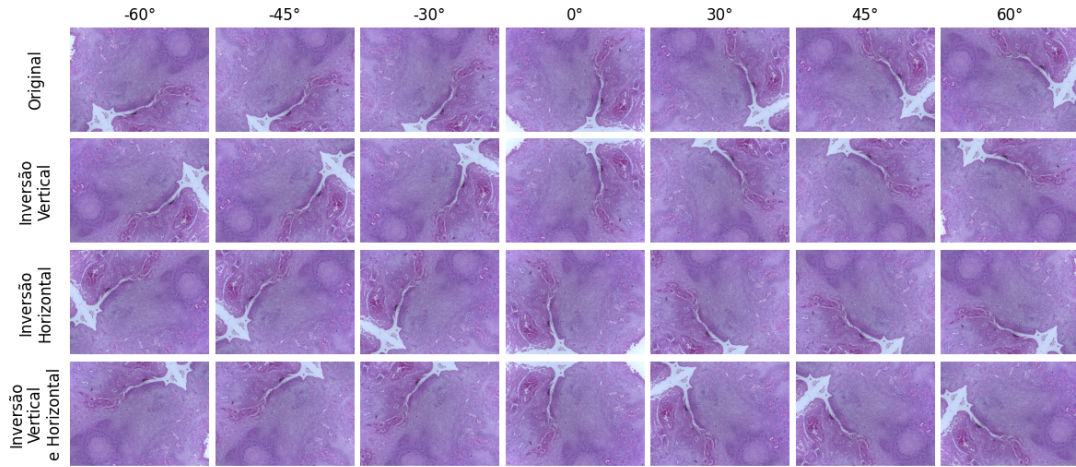
O aumento de dados é uma etapa importante no treinamento de modelos de aprendizado profundo, especialmente quando se trabalha com conjuntos de dados limitados. Essa técnica consiste em aplicar transformações nas imagens originais para aumentar a diversidade do conjunto de treinamento, o que pode melhorar a capacidade de generalização e a robustez do modelo. As técnicas utilizadas neste trabalho foram rotação de imagem, inversão de imagem, transformação de brilho e contraste, transformações no espaço de cor HSV e transformações no espaço de cor H&E.

O método de **Rotação e Inversão de Imagem** consiste em girar as imagens em ângulos variados e refletir suas orientações. A rotação é realizada em intervalos de 0 a 360 graus, sendo o ângulo de rotação um hiper-parâmetro definido com base nas características do conjunto de dados (KUMAR et al., 2024). Em imagens histopatológicas, a orientação da amostra geralmente é irrelevante para o diagnóstico, o que torna essa transformação amplamente aplicável e garante maior diversidade nos dados gerados. Além disso, a inversão horizontal e vertical reflete os píxeis da imagem, aumentando ainda mais a variabilidade sem introduzir artefatos indesejados. Na Figura 16 são apresentados exemplos da aplicação das transformações de rotação e inversão.

Outra técnica empregada é a **Transformação de Brilho e Contraste**, que consiste em aplicar perturbações aleatórias nesses atributos dentro de intervalos pré-definidos. Essa abordagem simula condições de iluminação variáveis, permitindo que os modelos se tornem mais robustos a essas variações. Ao expor o modelo a diferentes níveis de iluminação, a transformação de brilho e contraste aumenta a capacidade de generalização, preparando-o para cenários de aquisição de imagens em condições diversas. Na Figura 17 são apresentados exemplos das transformações de brilho e contraste.

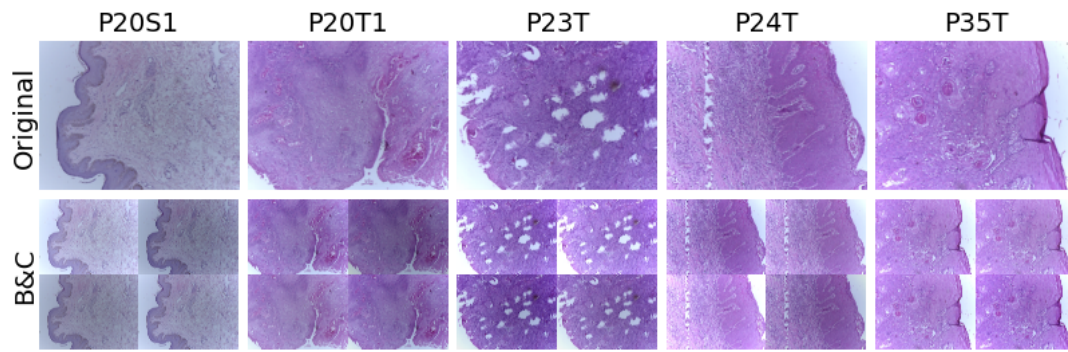
A transformação de brilho e contraste é aplicada a cada pixel x da imagem para simular variações de iluminação e melhorar a capacidade de generalização dos modelos. Inicialmente, o contraste da imagem é ajustado subtraindo a média da intensidade dos

Figura 16 – Transformações Rotação e Inversão.



Fonte: elaborada pelo Autor.

Figura 17 – Transformações de Brilho e Contraste.



Fonte: elaborada pelo Autor.

pixels e escalando os valores por um fator α , conforme descrito na Equação 2.10, onde μ representa a média da intensidade dos pixels, α é o fator de ajuste de contraste e $\text{clip}(\cdot)$ garante que os valores permaneçam dentro do intervalo válido para o tipo de dado da imagem.

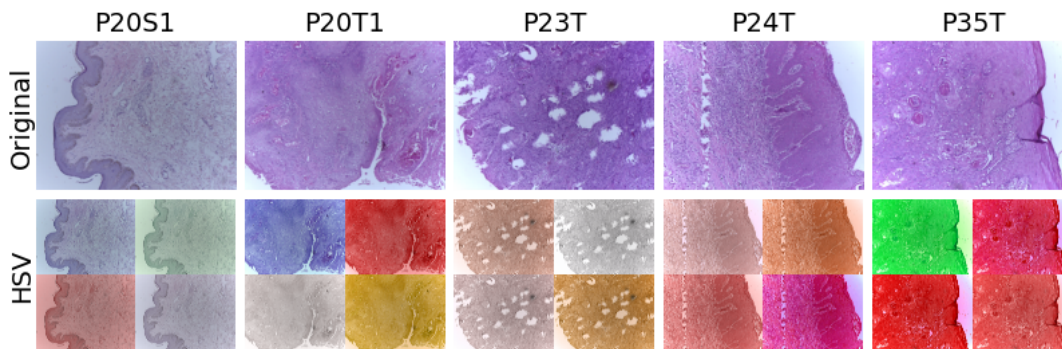
$$x' = \text{clip}((x - \mu) \cdot (1 + \alpha) + \mu, 0, \text{max}) \quad (2.10)$$

Após o ajuste de contraste, o brilho da imagem é modificado multiplicando os valores dos píxeis por um fator β , conforme descrito na Equação 2.11, onde β é o fator de ajuste de brilho e max representa o valor máximo permitido para os píxeis da imagem. Essa abordagem ajusta o brilho proporcionalmente à intensidade original da imagem, garantindo que as variações simuladas preservem a relação entre as intensidades dos píxeis.

$$x'' = \text{clip}(x' \cdot (1 + \beta), 0, \text{max}) \quad (2.11)$$

As **Transformações no Espaço de Cor HSV** são aplicadas para simular alterações cromáticas, permitindo que os modelos generalizem diante de variações de cor. Nesse processo, as imagens são convertidas para o espaço HSV, onde os canais de matiz (hue) e saturação são alterados aleatoriamente. Essas modificações geram novos padrões de cor, aumentando a diversidade do conjunto de dados. Entretanto, uma limitação dessa técnica é o risco de criar padrões de cor irrealistas, o que pode confundir o modelo e prejudicar sua capacidade de generalização para imagens reais (KUMAR et al., 2024). Por isso, os parâmetros de perturbação devem ser definidos com cautela. Na Figura 18 são apresentados exemplos das transformações no espaço de cor HSV.

Figura 18 – Transformações no Espaço de Cor HSV



Fonte: elaborada pelo Autor.

As transformações no espaço de cor HSV consistem na conversão da imagem do espaço RGB para o espaço HSV, onde os canais de matiz (H) e saturação (S) são perturbados aleatoriamente em um intervalo predefinido, enquanto o canal de valor (V) é mantido inalterado. Após a conversão, são aplicadas perturbações aleatórias nos canais de matiz e saturação, conforme a Equação 2.12, onde r_H e r_S são fatores aleatórios amostrados uniformemente de um intervalo $[1 - \alpha, 1 + \alpha]$, com α representando a intensidade da perturbação. O canal de valor (V) não é alterado.

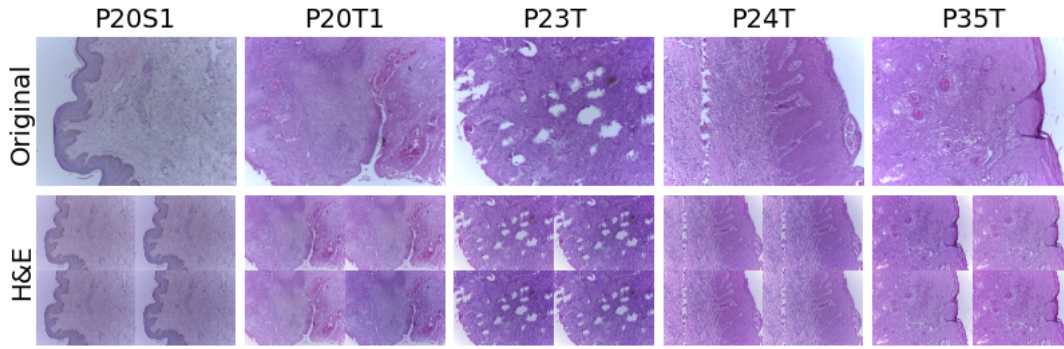
$$H' = \text{clip}(H \cdot r_H, 0, 1), \quad S' = \text{clip}(S \cdot r_S, 0, 1), \quad V' = V \quad (2.12)$$

Os valores são então normalizados para garantir que permaneçam em um intervalo válido. Por fim, a imagem é reconvertida para o espaço RGB utilizando a transformação inversa.

Por fim, as **Transformações no Espaço de Cor H&E** pretendem simular as variações de coloração frequentemente observadas em imagens histopatológicas. Diferentemente do espaço HSV, essa técnica trabalha diretamente com os canais de Hematoxilina e Eosina, principais corantes usados em lâminas histológicas. O processo envolve a conversão da imagem para o espaço de cor H&E, seguida por perturbações nos

canais correspondentes e a reconversão para o espaço RGB. Essas alterações permitem que os modelos aprendam a lidar com variações naturais na coloração das lâminas, tornando-os mais robustos às diferenças inerentes ao processo de coloração. Na Figura 19 são apresentados exemplos das transformações no espaço de cor H&E.

Figura 19 – Transformações no Espaço de Cor H&E



Fonte: elaborada pelo Autor.

As transformações no espaço de cor H&E são implementadas convertendo a imagem do espaço RGB para o espaço H&E, no qual os canais de Hematoxilina (H), Eosina (E) e Diaminobenzidina (D) são separados. Em seguida, são aplicadas perturbações aleatórias nos canais H e E, simulando variações de coloração observadas em amostras histopatológicas reais. O canal D é mantido inalterado, pois não está presente nas colorações padrão de Hematoxilina e Eosina. Para garantir que os valores permaneçam dentro de intervalos válidos, as intensidades modificadas são normalizadas antes de serem reconvertidas para o espaço RGB.

A conversão para o espaço de cor H&E é descrita na Equação 2.13 Onde I_{RGB} é a imagem original no espaço RGB, M é a matriz de separação de manchas utilizada na conversão do espaço RGB para H&E.

$$I_{HED} = M \cdot I_{RGB} \quad (2.13)$$

Após a conversão, são aplicadas perturbações aleatórias nos canais de Hematoxilina (H) e Eosina (E) descritas na Equação 2.14, onde r_H e r_E são fatores aleatórios amostrados uniformemente de um intervalo $[1 - \alpha, 1 + \alpha]$, com α representando a intensidade da perturbação. O canal de Diaminobenzidina (D) permanece inalterado.

$$H' = H \cdot r_H, \quad E' = E \cdot r_E, \quad D' = D \quad (2.14)$$

Em seguida, os valores são normalizados para garantir que permaneçam em um intervalo válido descrito na Equação 2.15.

$$H'' = \text{clip}(H', 0, 1), \quad E'' = \text{clip}(E', 0, 1), \quad D'' = \text{clip}(D', 0, 1) \quad (2.15)$$

Por fim, a imagem é reconvertida para o espaço RGB utilizando a transformação inversa descrita na Equação 2.16, onde M^{-1} é a matriz de reconversão do espaço H&E para RGB.

$$I'_{RGB} = M^{-1} \cdot I''_{HED} \quad (2.16)$$

Esse processo simula variações na coloração das lâminas histológicas, tornando o modelo mais robusto às diferenças cromáticas comuns em amostras de diferentes fabricantes de corantes, lotes diferentes ou laboratórios diferentes. O processo de conversão do padrão RGB para o HED foi estabelecido por Ruifrok e Johnston (2001), onde utiliza deconvolução para extrair os canais de cada componente de cor.

2.7 Avaliação

A avaliação de modelos de aprendizado de máquina é essencial para determinar sua capacidade de generalização. Neste trabalho, a mesma abordagem de avaliação foi aplicada em diversas etapas da metodologia proposta. Essa abordagem consiste em uma validação cruzada para particionar os dados em conjuntos de treinamento, validação interna e validação externa. Para a avaliação, foi utilizada a matriz de confusão para calcular o F1-Score, métrica principal deste estudo, que depende do cálculo da precisão e do *recall*.

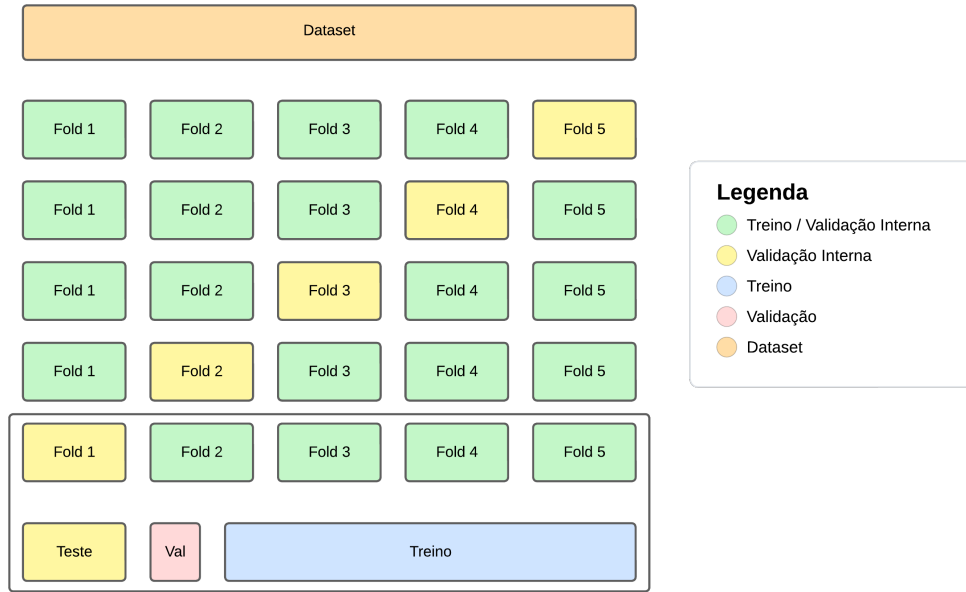
A validação cruzada é uma técnica que divide o conjunto de dados em várias partes (*folds*), permitindo que o modelo seja treinado em um subconjunto dos dados e testado em outro. Esse processo é repetido de forma que cada parte dos dados seja utilizada para validação pelo menos uma vez. A média dos resultados obtidos em todas as iterações oferecem uma avaliação mais robusta, mitigando problemas de sobreajuste e evitando a dependência de uma única divisão dos dados.

Neste estudo, foi utilizada validação cruzada com cinco *folds*, estratificada pelas classes. Durante cada iteração, uma partição foi reservada como validação interna para salvar o checkpoint do modelo. Em seguida, o modelo foi avaliado na validação externa (teste). A matriz de confusão resultante de cada iteração foi agregada para calcular as métricas de avaliação. A Figura 20 ilustra o processo de validação cruzada.

A matriz de confusão apresenta os acertos e erros do modelo, relacionando as classes preditas e reais. Em problemas binários, ela é composta por quatro elementos principais:

- **Verdadeiros Positivos (VP):** instâncias corretamente classificadas como positivas;

Figura 20 – Processo de Validação Cruzada utilizada no trabalho



Fonte: elaborada pelo Autor.

- **Falsos Positivos (FP):** instâncias incorretamente classificadas como positivas;
- **Verdadeiros Negativos (VN):** instâncias corretamente classificadas como negativas;
- **Falsos Negativos (FN):** instâncias positivas classificadas incorretamente como negativas.

A matriz de confusão é crucial para calcular métricas como precisão, recall e F1-Score, que ajudam a compreender o desempenho do modelo, especialmente em cenários de classes desbalanceadas. A Figura 21 apresenta um exemplo de matriz de confusão para um problema de classificação binária.

A precisão é uma métrica que avalia a acuracidade do modelo em relação às previsões positivas. Ela é calculada pela Equação 2.17, onde VP representa os verdadeiros positivos e FP os falsos positivos. Essa métrica indica a proporção de previsões positivas que foram de fato corretas, sendo particularmente relevante em problemas onde o custo de falsos positivos é significativo.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.17)$$

O *recall*, também chamado de sensibilidade ou taxa de verdadeiros positivos, avalia a capacidade do modelo de identificar corretamente todas as instâncias positivas. Ele é calculado pela Equação 2.18, onde VP representa os verdadeiros positivos e FN os falsos

Figura 21 – Matriz de Confusão.

		Valor Predito	
		Positivo	Negativo
Valor Esperado	Positivo	VP	FN
	Negativo	FP	VN

Fonte: elaborada pelo Autor.

negativos. O *recall* mede a proporção de instâncias positivas corretamente identificadas pelo modelo.

$$recall = \frac{VP}{VP + FN} \quad (2.18)$$

Embora ambas as métricas sejam fundamentais para a avaliação de modelos de classificação, cada uma apresenta limitações. A precisão pode ser alta mesmo que o modelo não consiga capturar um número significativo de instâncias positivas, enquanto o *recall* pode ser elevado se o modelo classificar muitas instâncias como positivas, incluindo aquelas que não são realmente positivas. Para cenários com classes desbalanceadas ou quando tanto falsos positivos quanto falsos negativos são indesejáveis, o F1-Score se torna uma métrica mais adequada.

O F1-Score é a média harmônica entre a precisão e o *recall*, penalizando modelos que apresentam alto desempenho em uma métrica, mas baixo em outra. Ele é calculado pela Equação 2.19, com valores variando entre 0 e 1, sendo 1 o desempenho ideal (quando a precisão e o *recall* atingem seus máximos).

$$F1\text{-Score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (2.19)$$

A abordagem de avaliação descrita nesta seção foi aplicada em várias etapas do método proposto. Ela foi utilizada durante a preparação dos dados, especialmente nas sub-etapas de balanceamento das classes e na escolha de técnicas de pré-processamento e aumento de dados. Além disso, foi empregada na etapa de pesquisa de arquitetura neural, permitindo a análise e comparação dos resultados.

2.8 Considerações Finais

O presente trabalho procura desenvolver uma metodologia robusta para a classificação de imagens histopatológicas de câncer de pênis. Para alcançar esse propósito, a metodologia apresentada no próximo capítulo incorpora as seguintes etapas: inicialmente, técnicas de aumento de dados são aplicadas para balancear as classes, reduzindo os efeitos de desbalanceamento do conjunto de dados. Em seguida, combinações de pré-processamento e aumento de representatividade são utilizadas para aprimorar a capacidade de generalização dos modelos, garantindo maior robustez frente a variações nas imagens. Por fim, a pesquisa de arquitetura neural é aplicada para projetar uma arquitetura otimizada, capaz de explorar eficientemente a distribuição específica dos dados disponíveis.

3 Trabalhos Relacionados

O avanço das técnicas de aprendizado profundo, aliado à crescente disponibilidade de dados em imagens médicas, impulsiona o desenvolvimento de métodos automatizados para diagnósticos histopatológicos. No contexto do câncer de pênis, embora ainda haja uma limitação de estudos específicos, a literatura apresenta abordagens promissoras baseadas em redes neurais convolucionais (CNNs) e mecanismos de atenção que otimizam a extração de características relevantes. Este capítulo examina os principais trabalhos relacionados, destacando métodos aplicados à análise histopatológica e as lacunas que justificam a metodologia proposta neste estudo.

Os trabalhos analisados estão organizados em duas vertentes principais. A primeira foca em estudos que aplicam mecanismos de atenção no contexto da histopatologia digital, ressaltando sua capacidade de aprimorar a extração de características relevantes em imagens médicas. A segunda aborda o uso de aprendizado profundo especificamente para a detecção de câncer de pênis, explorando diferentes arquiteturas e estratégias de treinamento. Ambas as vertentes são fundamentais para compreender os métodos aplicados à análise histopatológica e as lacunas que justificam a metodologia proposta neste estudo.

3.1 Mecanismos de Atenção na Histopatologia Digital

A metodologia proposta por [Chen et al. \(2022b\)](#) apresenta uma abordagem baseada na fusão de características extraídas de três arquiteturas distintas: VGG-16, Inception-v3 e Xception. Sendo acopladas a diferentes mecanismos de atenção às arquiteturas: SimAM (*A Simple, Parameter-Free Attention Module*), SE (*squeeze-and-excitation*) e ECA (*Efficient Channel Attention*), respectivamente. A hipótese defendida pelo autor é que a diversidade dos tipos de atenção reduz a redundância nas características extraídas e, conseqüentemente, minimiza os erros do modelo. Os blocos de atenção foram inseridos nas saídas dos blocos de backbone das arquiteturas, refinando as características ao longo do processo de extração. Além disso, experimentos de ablação conduzidos pelos autores apresentam um maior desempenho quando os mecanismos de atenção são acoplados na arquitetura. Esses resultados destacam o impacto positivo do uso de mecanismos de atenção na classificação de imagens histopatológicas.

A proposta [Ding et al. \(2024\)](#) consiste em uma arquitetura baseada na ResNet-50, aprimorada por modificações no bloco *bottleneck*. Essas alterações incluem o uso de convoluções *depth-wise* e *point-wise*, além da introdução estratégica de mecanismos de atenção no interior do bloco e no ramo residual, uma abordagem inovadora e pouco explorada em arquiteturas com atenção. Os experimentos de ablação conduzidos pelo

autor demonstraram que os mecanismos de atenção resultaram em um aumento de 0,9% na precisão do modelo. Esse incremento é atribuído à capacidade da arquitetura de focar seletivamente nos canais mais relevantes e em informações críticas nos canais individuais. Os achados do estudo ressaltam a importância de integrar mecanismos de atenção de forma estratégica, inclusive no ramo residual, demonstrando seu impacto na melhoria da extração de características relevantes.

A abordagem apresentada em [Li et al. \(2022b\)](#) consiste em uma arquitetura voltada para a extração e refinamento de características em imagens de histopatologia. O modelo inicia com uma rede convolucional rasa, responsável por realizar a extração inicial das características. Em seguida, um estágio de refinamento é implementado por meio de módulos DPA, que combinam convoluções *depth-wise* e *point-wise* com mecanismos de atenção, otimizando a captura de detalhes texturais e semânticos das imagens. Para aprimorar ainda mais a discriminação de características, uma unidade de fusão residual é incorporada após o estágio de refinamento. Nessa unidade, um mecanismo de atenção por canal é aplicado, permitindo que os mapas de ativação mais representativos sejam destacados para o módulo de classificação. É importante destacar a disposição estratégica de mecanismos de atenção na arquitetura, quanto ao seu tipo e o objetivo.

[Xu et al. \(2024\)](#) propõem uma arquitetura baseada na ResNet-50, modificada por meio da introdução da convolução SK no bloco bottleneck. As convoluções SK permitem a seleção adaptativa de diferentes tamanhos de *kernels* conforme as características da entrada, utilizando um mecanismo de atenção softmax interno. Além disso, o estudo explora a adição de diferentes mecanismos de atenção à arquitetura, incluindo *Squeeze-and-Excitation* (SE), *Coordinate Attention* (CA) e *Efficient Channel Attention* (ECA), melhorando o refinamento de características e a seleção de mapas de características mais úteis para a tarefa. Os experimentos apresentados demonstram que a eficácia do mecanismo de atenção varia conforme o grau de ampliação da imagem, sugerindo que a escolha estratégica do tipo de atenção é crucial para otimizar o desempenho da arquitetura.

3.2 Aprendizado Profundo para Detecção de Câncer de Pênis

A literatura apresenta uma quantidade limitada de estudos voltados especificamente para a classificação de imagens histopatológicas de câncer de pênis. Na revisão realizada, foram identificados apenas os trabalhos de [Lauande et al. \(2022\)](#), [Belfort et al. \(2023\)](#), [Vale et al. \(2024\)](#) e [Lauande et al. \(2024\)](#), todos utilizando o mesmo conjunto de dados empregado neste estudo. Essa lacuna bibliográfica pode ser atribuída à raridade do câncer de pênis, dificultando a obtenção de amostras representativas, além da limitada disponibilidade de dados devidamente rotulados, um fator crítico para o desenvolvimento e validação de modelos de aprendizado profundo.

O estudo conduzido por [Lauande et al. \(2022\)](#) propõe a classificação de imagens histopatológicas utilizando um modelo DenseNet-201 pré-treinado, associado a técnicas de *data augmentation* e pré-processamento. O *data augmentation* inclui rotações de até 90° e espelhamento vertical e horizontal, buscando aumentar a diversidade do conjunto de treinamento. Adicionalmente, o autor emprega um método de pré-processamento baseado no algoritmo CLAHE, aplicado ao canal Y do espaço de cores YUV, para normalizar a luminância das imagens. O estudo também realiza testes de ablação para avaliar o impacto individual do *data augmentation* e do pré-processamento sobre o desempenho do modelo, demonstrando que ambos os aspectos são significativos para o desempenho do método.

A proposta de [Belfort et al. \(2023\)](#) constrói um modelo DenseNet-121 pré-treinado, com um mecanismo de atenção suave acoplado à sua saída. O método inclui um *data augmentation*, com rotações de até 160° em intervalos de 4°, além de espelhamentos verticais e horizontais. Adicionalmente, um pré-processamento baseado na equalização de histograma é aplicado ao componente B do espaço de cor cieLAB, visando realçar áreas específicas da imagem críticas para a decisão diagnóstica. Experimentos comparativos entre configurações com e sem o uso de atenção demonstraram que a presença do mecanismo de atenção resulta em um desempenho superior.

O trabalho de [Vale et al. \(2024\)](#) utiliza a arquitetura DenseNet-201 como base, com a adição do mecanismo de atenção em múltiplas escalas (MSAG) e do mecanismo de atenção em escala única (USAG) nos ramos de concatenação. Contudo, o aumento no número de parâmetros decorrente dessas modificações tornou o treinamento completo da arquitetura final computacionalmente inviável, permitindo apenas a adição nos blocos iniciais da rede. Para lidar com o desbalanceamento do conjunto de dados, foram aplicadas transformações de matriz e saturação direcionadas à classe normal, além de transformações geométricas aplicadas a todas as imagens, para aumentar a diversidade do treinamento. O estudo também emprega o método de pré-processamento Vahadane, amplamente utilizado em análises histopatológicas para normalização de cores. No entanto, por requerer uma imagem de referência para ajustar as demais, foi necessário realizar uma avaliação exaustiva em todas as imagens, a fim de identificar referências adequadas para as duas arquiteturas propostas.

Em [Lauande et al. \(2024\)](#) é desenvolvida uma arquitetura híbrida, composta pelos dois blocos iniciais de uma DenseNet pré-treinada, combinada com blocos densos nos estágios intermediários e a inclusão de blocos *Transformers* e/ou MBConv nos estágios finais. Essa abordagem visa explorar o potencial complementar de diferentes tipos de blocos para aprimorar a extração e o processamento de características nas imagens. O autor utiliza o mesmo pré-processamento descrito em [Lauande et al. \(2022\)](#), além de pré-treinar o modelo em um dataset maior e mais diversificado, o LC25000, visando melhorar a capacidade de generalização da arquitetura. O estudo também analisa a relação

entre as configurações arquiteturais propostas e os resultados obtidos em diferentes níveis de ampliação das imagens histopatológicas, destacando como as escolhas arquitetônicas impactam o desempenho do modelo em cenários variados.

3.3 Diferenciais da Metodologia Proposta

As abordagens apresentadas por [Chen et al. \(2022b\)](#), [Li et al. \(2022b\)](#), [Ding et al. \(2024\)](#), [Xu et al. \(2024\)](#) e [Vale et al. \(2024\)](#) evidenciam a eficácia dos mecanismos de atenção no refinamento de características e no aprimoramento do desempenho de modelos em tarefas de classificação. Essa eficiência é particularmente destacada no trabalho de [Chen et al. \(2022b\)](#), no qual os experimentos de ablação demonstraram um aumento objetivo de 0,20% no F1-Score, totalizando 98,85%, como resultado da adição de mecanismos de atenção. Por sua vez, [Li et al. \(2022b\)](#) enfatizam a importância do posicionamento estratégico dos mecanismos de atenção ao longo da arquitetura, evidenciando que a escolha do tipo de atenção deve ser alinhada com o local de aplicação.

O estudo de [Ding et al. \(2024\)](#) apresenta uma contribuição diferenciada ao explorar a inclusão de mecanismos de atenção nos ramos residuais, uma abordagem pouco comum em arquiteturas baseadas em ResNet, oferecendo novas possibilidades para o design de futuras arquiteturas. Além disso, [Xu et al. \(2024\)](#) destacam como o desempenho dos mecanismos de atenção pode variar significativamente conforme o grau de ampliação das imagens, reforçando a necessidade de ajustar as escolhas arquiteturais conforme o contexto dos dados. Por fim, o trabalho [Vale et al. \(2024\)](#) ressalta que os resultados obtidos dependem diretamente do tipo de atenção aplicada, enfatizando a importância de selecionar mecanismos de atenção adequados às características específicas do problema abordado.

Além do refinamento de características, [Lauande et al. \(2022\)](#), [Belfort et al. \(2023\)](#) e [Vale et al. \(2024\)](#) destacam a importância de técnicas de pré-processamento e *data augmentation* na melhoria da análise de imagens histopatológicas. Em particular, o estudo [Lauande et al. \(2022\)](#) realiza uma ablação que mostra os ganhos obtidos com o uso de pré-processamento, enquanto a utilização de *data augmentation* não apresentou benefícios consistentes em todos os experimentos. Esses resultados ressaltam a importância de uma investigação mais detalhada na aplicação dessas técnicas, visando maximizar seu impacto no desempenho dos modelos.

A Tabela 1 resume os principais trabalhos relacionados, destacando suas metodologias, contribuições e resultados no contexto da classificação de imagens histopatológicas. Os estudos analisados enfatizam o impacto de estratégias como o uso de mecanismos de atenção, estratégias de pré-processamento e *data augmentation* na otimização de arquiteturas de aprendizado profundo. Embora as abordagens destacadas

Tabela 1 – Trabalhos relacionados, com base de dados usada, resumo do método e resultados obtidos.

Autor	Base de Imagens	Método	Resultados.
(CHEN et al., 2022b)	HE-CRC-DS	Fusão de características + mecanismos de atenção	Classe Normal: -F1-Score de 98,86 Classe Anormal: -F1-Score de 98,84 Acurácia de 98,85 (+/-0.27)
(DING et al., 2024)	PathoIMG	ResNet-50 + alterações no bottleneck + mecanismos de atenção	Classe Benigna: -F1-Score de 0,990 Classe InSitu: -F1-Score de 0,978 Classe Invasiva: -F1-Score de 0,980 Classe Normal: -F1-Score de 0,961
(LI et al., 2022b)	LC25000 Pcam BreakHis	Módulos de refinamento estrategicamente posicionados	LC25000: -F1-Score de 0,9992 Pcam: -F1-Score de 0,9754 BreakHis: -F1-Score de 0,9779
(XU et al., 2024)	BreakHis	ResNet-50 + convoluções SK + mecanismo de atenção	F1-Score de 97,13 Acurácia de 97,89
(LAUANDE. et al., 2022)	PCPam	CLAHE no canal Y do YUV + Densenet 201 + Transfer Learning	Ampliação 40X: -F1-Score de 97,39 (+/-2.13) Ampliação 100X: -F1-Score de 97,31 (+/-3.62)
(BELFORT et al., 2023)	PCPam	CLAHE no canal B do cieLAB + Densenet 121 + Modelo em Cascata	Ampliação 40X: -F1-Score de 91,2 (+/-2.3) Ampliação 100X: -F1-Score de 92,4 (+/-1.5)
(VALE et al., 2024)	PCPam	Pré-processamento Vahadane + Densenet 201 + Mecanismos de atenção	Ampliação 100X: -F1-Score de 93,1 (+/-4.6)
(LAUANDE et al., 2024)	LC25000 PCPam	CLAHE no canal Y do YUV + Transferência de domínio. + Arquitetura Híbrida	Ampliação 40X: -F1-Score de 95,78 (+/-4.39) Ampliação 100X: -F1-Score de 94,99 (+/-3.88)

Fonte: elaborada pelo Autor.

tenham alcançado resultados relevantes, o presente estudo diferencia-se ao propor uma abordagem estruturada em três etapas: inicialmente, são realizados experimentos para identificar as melhores combinações de transformações geométricas e espelhamento, visando balancear o dataset utilizando uma arquitetura base. Em seguida, é conduzida uma análise para determinar o melhor pré-processamento e a estratégia de *data augmentation* aplicada online, buscando aumentar a variabilidade dos dados e a robustez dos modelos. Por fim, aplica-se a metodologia NAS com a exploração de combinações de *stem layers*, mecanismos de atenção adicionados ao backbone, visando explorar e identificar arquiteturas otimizadas, adaptadas às especificidades do problema de classificação de imagens de câncer de pênis.

4 Materiais e Método

Este capítulo apresenta uma descrição detalhada da metodologia empregada no estudo, abrangendo todas as etapas fundamentais. Primeiramente, são especificados o ambiente experimental e as configurações utilizadas, incluindo os recursos computacionais e ferramentas empregadas. Em seguida, é descrito o conjunto de dados, destacando suas características, origem e os processos de pré-processamento aplicados. A seção de avaliação aborda os critérios, métricas e métodos utilizados para validar os resultados. Por fim, o método proposto é detalhado, explicando os passos adotados para alcançar os objetivos do estudo.

4.1 Ambiente e Configurações experimentais

Os experimentos deste estudo foram conduzidos utilizando aceleradoras gráficas NVIDIA RTX 3060 com 12 GB de memória. O ambiente foi configurado com base na imagem Docker `pytorch:2.4.1-cuda12.4-cudnn9-runtime`, incorporando o PyTorch na versão 2.4.1, CUDA 12.4 e Python 3.11.9. Para o rastreamento e gerenciamento dos experimentos, foi utilizada a ferramenta de MLOps `Weights & Biases (wandb)`, que permitiu o registro das métricas e o monitoramento dos resultados.

4.2 Dataset

O conjunto de dados utilizado neste trabalho foi desenvolvido pelo Projeto Câncer de Pênis Amazônia Legal, em colaboração com o Núcleo de Computação Aplicada da Universidade Federal do Maranhão. Esse dataset é composto por 194 imagens em codificação RGB, com resolução de 2048x1536 píxeis, capturadas em ampliações de 40x e 100x. Cada imagem possui sua correspondente em ambas as ampliações, garantindo consistência na análise.

A distribuição das classes no dataset é apresentada na Tabela 2. O conjunto inclui 55 imagens de tecidos com diagnóstico de câncer e 42 imagens de tecidos sem câncer para cada ampliação.

As imagens foram capturadas em 2021, a partir de amostras de tecidos obtidos por biópsias de pacientes em tratamento no estado do Maranhão, Brasil. Os tecidos foram corados com Hematoxilina e Eosina (H&E) e fotografados utilizando uma câmera de alta definição Leica ICC50 HD acoplada a um microscópio de campo claro Leica DM500. O processo de aquisição foi conduzido por dois estudantes de pós-graduação. Posteriormente,

Tabela 2 – Distribuição das imagens conforme ampliação e classificação patológica.

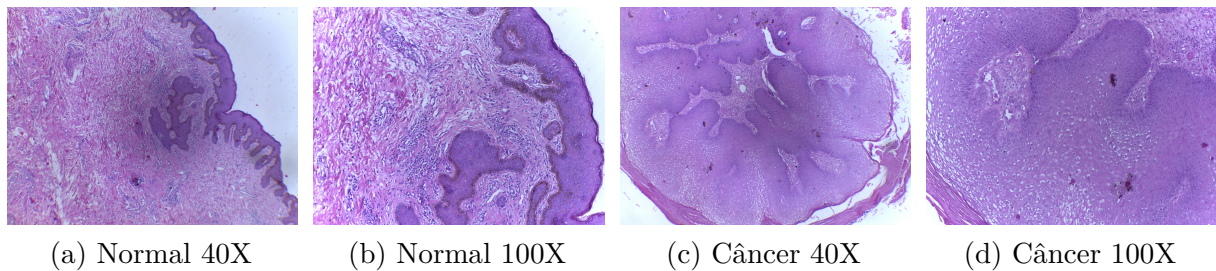
Categoria/Ampliação	40x	100x
Normal	42	42
Câncer	55	55
Total	97	97

Fonte: elaborada pelo Autor.

as imagens foram analisadas e classificadas por dois patologistas, com o suporte do software Leica Aperio ImageScope (LAUANDE. et al., 2022).

Na Figura 22, são apresentados exemplos representativos das imagens do dataset. As Figuras 22a e 22b correspondem a imagens de tecidos sem câncer capturadas em diferentes ampliações (40X e 100X, respectivamente). Já as Figuras 22c e 22d apresentam imagens de tecidos com câncer, também capturadas nas mesmas ampliações.

Figura 22 – Imagens do Dataset



Fonte: elaborada pelo Autor.

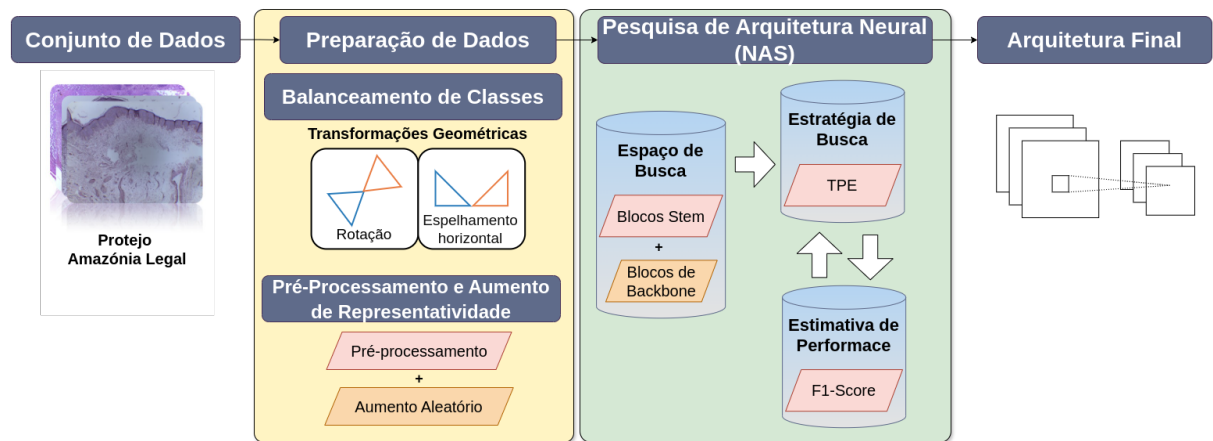
4.3 Método

Esta seção apresenta o método proposto para a classificação de imagens histopatológicas de câncer de pênis. O diagrama representando o fluxo das etapas realizadas neste estudo está ilustrado na Figura 23.

O fluxograma inicia com a aquisição do conjunto de dados, descrita na Seção 4.2. O conjunto de dados é composto por imagens histopatológicas capturadas em diferentes ampliações, contendo tanto tecidos com diagnóstico de câncer quanto sem câncer.

Em seguida, é realizado o balanceamento das classes, utilizando transformações geométricas. Essas transformações, aplicadas separadamente a cada classe, visam gerar dados adicionais para as classes minoritárias e melhorar a diversidade do conjunto de dados, sendo avaliado qual conjunto de aumentos apresenta o melhor desempenho para balancear as classes na arquitetura base que será otimizada na etapa NAS, a arquitetura base utilizada foi a ResNet-50.

Figura 23 – Fluxograma da metodologia proposta.



Fonte: elaborada pelo Autor.

Na próxima etapa, é feita a otimização do pré-processamento e do aumento de representatividade na arquitetura base, cujo objetivo é aprimorar a generalização do modelo. Durante esta fase, diferentes combinações de técnicas de pré-processamento, como normalização de cores e equalização de histograma, são testadas juntamente com técnicas de aumento de representatividade aplicadas dinamicamente durante o treinamento.

Com o conjunto de dados balanceado, pré-processado e aumentado, inicia-se a otimização da arquitetura neural. Nesse estágio, são definidos os blocos iniciais (*stem blocks*) e os blocos do *backbone* para cada estágio da arquitetura base. A escolha dessas configurações é realizada por meio de uma pesquisa de arquitetura neural (NAS) para encontrar uma estrutura otimizada que melhor se adapte às características do conjunto de dados.

Ao final, o método proposto resulta em uma arquitetura otimizada, configurada para maximizar o desempenho na classificação das imagens histopatológicas de câncer de pênis, considerando as especificidades do conjunto de dados utilizado.

4.3.1 Preparação dos Dados

A preparação dos dados pretende estabelecer uma distribuição que potencialize o desempenho da arquitetura base, a qual será posteriormente otimizada. Como a metodologia NAS encontra uma estrutura otimizada para um determinado conjunto de dados, qualquer mudança significativa nessa distribuição, seja durante o processo de otimização ou após a definição da arquitetura, pode prejudicar o desempenho do modelo.

O balanceamento foi realizado por meio de transformações geométricas aplicadas diretamente a ambas as classes, gerando maior diversidade de exemplos e equilibrando a quantidade de amostras. Dessa forma, buscou-se reduzir o risco de viés do modelo em relação à classe mais numerosa.

Concluído o balanceamento, procedeu-se a uma análise para identificar quais combinações de pré-processamento e aumento de representatividade oferecem melhor desempenho em termos de generalização. O objetivo foi preparar o conjunto de dados de maneira uniforme antes de otimizar a arquitetura neural, evitando variações que pudessem comprometer os resultados.

4.3.1.1 Balanceamento de Classes

Em dados médicos, o desbalanceamento entre classes é comum porque as incidências das patologias variam, dificultando obter conjuntos equilibrados. Em casos raros, como o câncer de pênis, esse desequilíbrio se acentua. Para este estudo, recorreu-se a transformações geométricas que aumentam artificialmente o número de amostras das classes, buscando estabelecer uma proporção favorável à classificação.

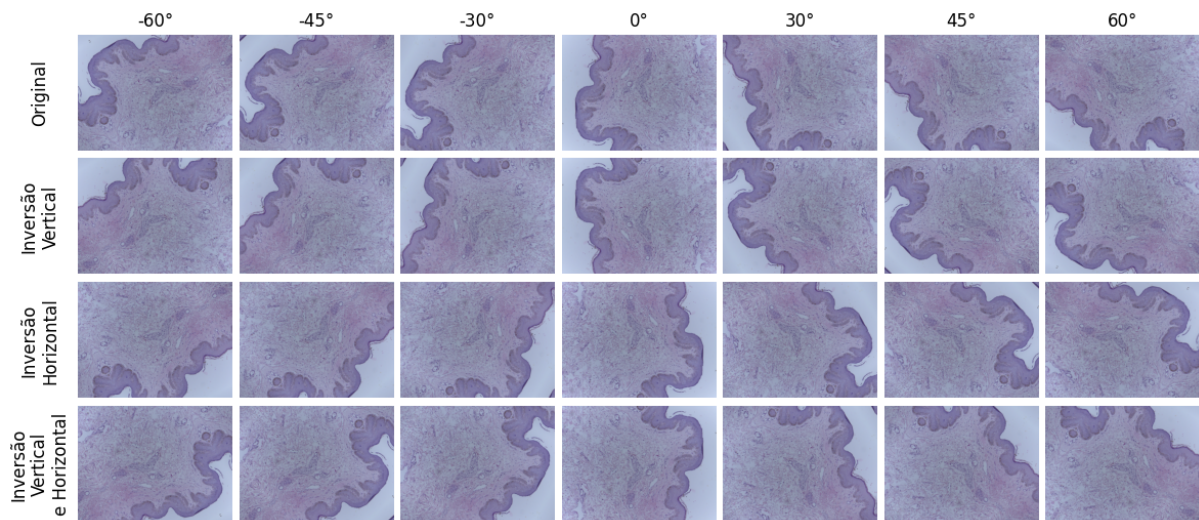
A estratégia baseou-se na criação de “perfis de aumento”, cada qual com um conjunto fixo de transformações a serem aplicadas separadamente em cada classe. Esses perfis foram então testados em uma busca em grade, avaliando todas as combinações possíveis entre as classes. Os perfis desenvolvidos incluem:

- **Sem aumento:** quantidade original de instâncias por classe, sem aplicação de transformações;
- **Todas as inversões:** aplicação de inversões horizontal, vertical e horizontal com vertical, quadruplicando o número de instâncias da classe;
- **Inversões e rotação de 45 graus:** imagens obtidas por inversões, combinadas com rotação de 45 graus positivas e negativas, resultando em um aumento de doze vezes no número de instâncias;
- **Inversões e rotações de 30 e 60 graus:** imagens obtidas por inversões, combinadas com rotações de 30 e 60 graus positivas e negativas, aumentando o número de instâncias em 20 vezes.

Somente rotações de 30, 45 e 60 graus foram aplicadas, pois combinações de inversões com rotações superiores a 90 graus resultam em imagens redundantes. A Figura 24 ilustra as diferentes transformações utilizadas nesta etapa.

Após a aplicação dos perfis, as combinações foram avaliadas utilizando a metodologia descrita na Seção 2.7. A melhor configuração identificada foi adotada como base para a próxima etapa, que envolve a otimização do pré-processamento e o aumento de representatividade.

Figura 24 – Transformações Aplicadas no Balanceamento de Dados.



Fonte: elaborada pelo Autor.

4.3.1.2 Pré-processamento e Aumento de Representatividade

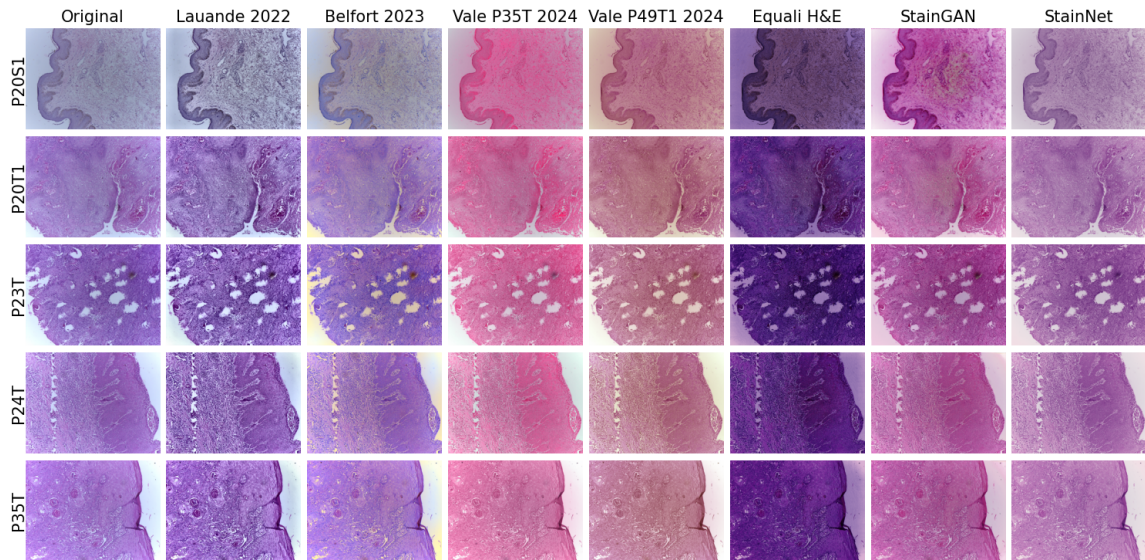
De acordo com [Tellez et al. \(2019\)](#), combinar técnicas de aumento de cor e normalização de cor pode melhorar o desempenho em imagens histopatológicas. O estudo indica que esse tipo de aumento ajuda a reduzir o erro de generalização, mas ressalta que tanto a categoria de aumento quanto a intensidade aplicada têm grande impacto no desempenho final do modelo.

No entanto, o estudo evidenciou não haver uma abordagem única que seja eficaz para todos os casos. Os resultados experimentais variaram amplamente dependendo do conjunto de dados analisado, indicando que as estratégias de pré-processamento e aumento de representatividade devem ser cuidadosamente ajustadas para cada problema específico. Isso ressalta a necessidade de explorar diferentes combinações e parâmetros para identificar a configuração ideal que maximize a capacidade de generalização do modelo em cenários variados.

Visando a normalização das imagens e o realce das estruturas relevantes, este estudo avaliou as abordagens propostas por [Lauande et al. \(2022\)](#) e [Belfort et al. \(2023\)](#), além da normalização Vahadane ([VAHADANE et al., 2016](#)) com as imagens referências encontradas por [Vale et al. \(2024\)](#) e das técnicas baseadas em redes adversárias, StainGAN ([SHABAN et al., 2019](#)) e StainNet ([KANG et al., 2021](#)). Essas soluções atuam na padronização de cores, com benefícios como a dispensa de imagens de referência (StainGAN) ou maior eficiência computacional (StainNet). Também foi empregada a equalização de canais de Hematoxilina e Eosina (H&E) para destacar componentes estruturais do tecido, contribuindo para uma análise mais consistente e robusta em diferentes cenários.

A Figura 25 apresenta exemplos dos métodos de pré-processamento aplicados neste estudo.

Figura 25 – Pré-processamentos aplicados no estudo



Fonte: elaborada pelo Autor

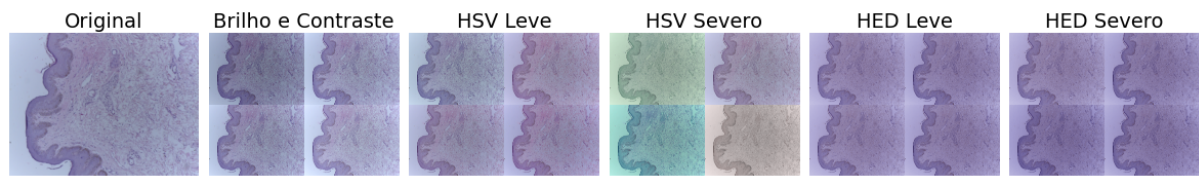
O Aumento de Representatividade difere do balanceamento de classes ao aplicar, durante o treinamento, perturbações aleatórias que simulam cenários reais de variação em imagens histopatológicas, como mudanças de brilho, contraste e cores. Essa estratégia busca ampliar a diversidade dos dados em tempo de treinamento, promovendo maior robustez e capacidade de generalização dos modelos sem gerar amostras irreais.

- Perturbações de Brilho e Contraste: $[-0,10, 0,10]$ para brilho e $[-0,15, 0,15]$ para contraste;
- Perturbações no Espaço de Cor HSV: $[-0,1, 0,1]$ para perturbações leves e $[-1,0, 1,0]$ para perturbações severas;
- Transformações no Espaço de Cor H&E: $[-0,05, 0,05]$ para perturbações leves e $[-0,2, 0,2]$ para perturbações severas.

Exemplos das transformações aplicadas são apresentados na Figura 26.

A partir das configurações definidas na etapa de balanceamento de dados, realizou-se uma busca em grade para selecionar, de forma conjunta, o melhor pré-processamento e aumento de representatividade, com foco em garantir maior diversidade das amostras. Diferentemente do balanceamento (voltado a equilibrar o conjunto), essa fase privilegiou técnicas que simulassem condições de aquisição variadas, tornando o modelo mais robusto.

Figura 26 – Transformações Aplicadas no Aumento de Representatividade



Fonte: elaborada pelo Autor.

A configuração escolhida ao fim desse processo passou a ser utilizada na otimização da arquitetura neural.

4.3.2 NAS: Pesquisa de Arquitetura Neural

Nesta etapa, foi aplicada a metodologia NAS (Neural Architecture Search) aos dados processados na etapa anterior. O processo é dividido em três subetapas principais: definição do Espaço de Busca, estratégia de busca e estimativa de desempenho.

4.3.2.1 Definição do Espaço de Busca

Neste estudo, foi utilizada a abordagem baseada em blocos (ou células) para reduzir a complexidade e facilitar a adaptação a diferentes conjuntos de dados (ELSKEN; METZEN; HUTTER, 2019; KANG et al., 2023). Também foram delimitados dois tipos de blocos principais — *stem*, para a extração inicial de características, e *backbone*, para a construção de representações progressivamente mais complexas.

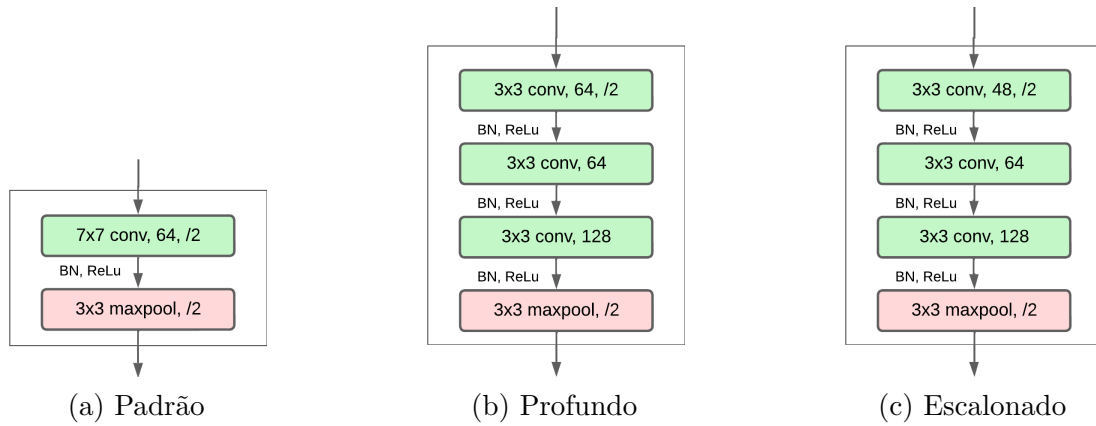
4.3.2.1.1 Blocos Stem

A camada de *Stem* tem dois objetivos principais: realizar a extração inicial das características de mais baixo nível e reduzir a resolução dos mapas de características na etapa inicial. As características de baixo nível, como bordas e texturas, desempenham um papel crucial na análise de imagens histopatológicas, pois fornecem as bases para o aprendizado hierárquico e a abstração de padrões mais complexos nas camadas subsequentes. A redução inicial da resolução dos mapas de características contribui para o consumo eficiente de recursos computacionais, reduzindo a quantidade de operações necessárias nas camadas posteriores da arquitetura, otimizando o desempenho geral do modelo.

Na Figura 27, são apresentados três tipos de blocos de *stem* disponíveis no espaço de busca:

- **Padrão (Figura 27a):** utiliza uma única camada convolucional com filtros 7×7 e *stride* de 2, característico da ResNet (HE et al., 2016).

Figura 27 – blocos do Stem



Fonte: elaborada pelo Autor.

- **Profundo (Figura 27b):** composto por três camadas convolucionais com filtros 3×3 e *stride* de 2 na primeira camada e o dobro de filtros na última (HE et al., 2018).
- **Escalonado (Figura 27c):** semelhante ao modelo profundo, mas com a quantidade de filtros escalonada entre as camadas (GUO et al., 2020). A quantidade de filtros em cada camada é ajustada progressivamente, começando com $\frac{3}{4}F$ na primeira camada, F na segunda e $2F$ na última, onde F representa a largura base da rede.

Cada configuração de *stem* tem uma abordagem distinta para extrair características de baixo nível. O *stem* padrão utiliza um campo receptivo maior logo no início, capturando padrões mais amplos. O *stem* profundo, com camadas adicionais, permite a correlação de características simples em padrões mais complexos. Já o *stem* escalonado oferece flexibilidade ao aumentar gradualmente o número de filtros, facilitando a transição entre características simples e complexas.

4.3.2.1.2 Blocos Backbone

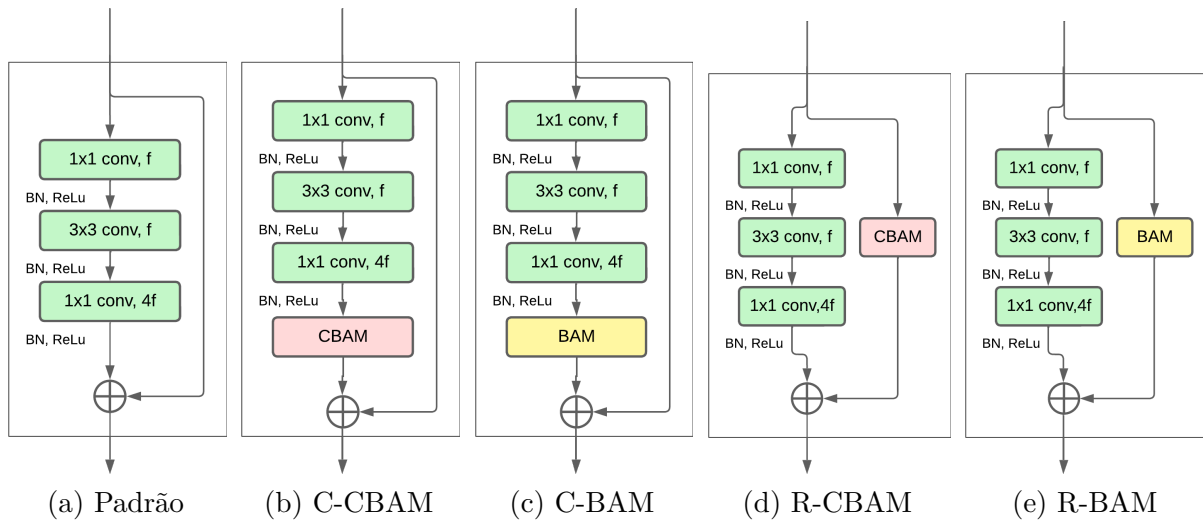
Os blocos convolucionais no *backbone* das redes CNNs são responsáveis por realizar o aprendizado hierárquico, correlacionando características de baixo nível para gerar abstrações progressivamente mais complexas ao longo da rede. A profundidade do *backbone* tem impacto direto no desempenho do modelo: redes mais profundas, em teoria, conseguem maior capacidade de abstração. No entanto, o aumento de camadas nem sempre resulta em ganhos proporcionais no desempenho, sendo acompanhado por custos computacionais e de memória elevados (ZHANG et al., 2017).

Os mecanismos de atenção oferecem uma alternativa eficiente para melhorar o desempenho das redes com um acréscimo relativamente baixo no custo computacional. Eles ajustam dinamicamente as características aprendidas, destacando informações relevantes e

suprimindo as irrelevantes. Na área de análise histopatológica, os mecanismos de atenção têm sido eficazes em tornar as características extraídas mais discriminativas, auxiliando diagnósticos precisos e fornecendo suporte confiável para patologistas (WOO et al., 2018b; PARK et al., 2018; LI et al., 2022b; CHEN et al., 2022b).

Neste estudo, dois tipos de blocos com mecanismos de atenção foram incluídos no espaço de busca: o BAM (*Bottleneck Attention Module* apresentado na subseção 2.4.3.1) e o CBAM (*Convolutional Block Attention Module* apresentado na subseção 2.4.3.2). Ambos combinam atenção espacial e por canal, mas apresentam diferenças na estrutura e no modo como as operações de atenção são aplicadas. Essas variações podem influenciar o desempenho, dependendo do tipo de dados e da tarefa.

Figura 28 – Células do Backbone



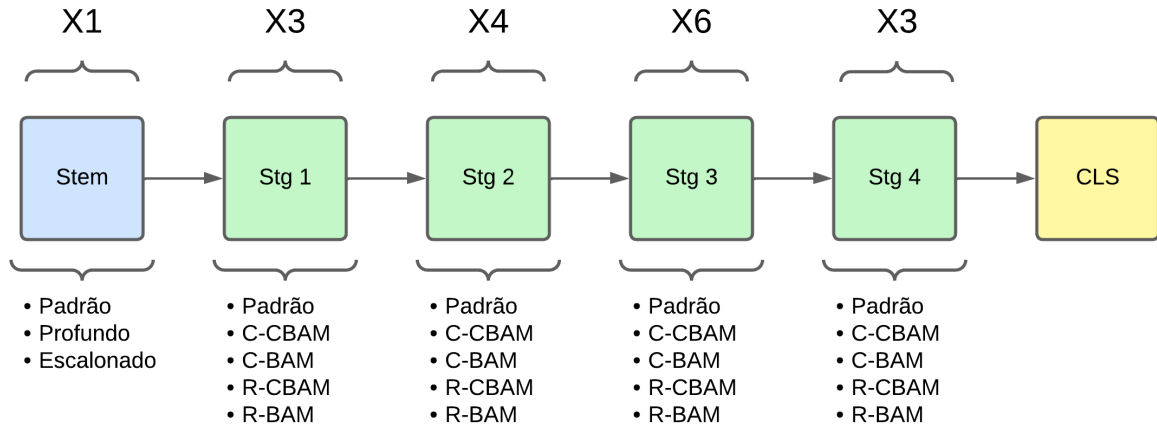
Fonte: elaborada pelo Autor.

A Figura 28 apresenta as células de *backbone* utilizadas no espaço de busca. Os blocos incluem o padrão ResNet (28a), o CBAM aplicado na saída convolucional (28b) e no ramo residual (28d), além do BAM implementado tanto na saída convolucional (28c) quanto no ramo residual (28e). Cada configuração foi incluída para explorar como diferentes combinações de mecanismos de atenção podem afetar o desempenho.

4.3.2.1.3 Definição da Arquitetura

As células definidas no espaço de busca foram integradas para formar arquiteturas completas, conforme ilustrado na Figura 29. A base da arquitetura é a ResNet-50, que mantém os hiper-parâmetros padrão, como a quantidade de filtros e blocos em cada estágio. A otimização ocorre pela seleção dos blocos ideais para cada estágio, permitindo flexibilidade no design da rede.

Figura 29 – Arquitetura e Espaço de Busca.



Fonte: elaborada pelo Autor.

O espaço de busca gerado resulta em 1.875 combinações possíveis, considerando as variações dos blocos *stem* e *backbone*. Devido ao custo computacional associado ao treinamento de cada combinação, foram adotados métodos de otimização heurística para explorar esse espaço de maneira eficiente, buscando arquiteturas que maximizem o desempenho com o menor custo possível.

4.3.2.2 Estratégia de Busca

Após a definição do espaço de busca, é necessário implementar uma solução eficiente para identificar a arquitetura neural ideal. A busca por arquiteturas ótimas pode ser conduzida utilizando técnicas tradicionais, como pesquisa em grade e aleatória, ou métodos mais avançados, como otimização bayesiana e baseada em gradientes (KANG et al., 2023). Devido ao alto custo computacional associado ao treinamento das 1.875 combinações possíveis no espaço de busca definido, a realização de uma busca exaustiva (em grade) torna-se inviável com os recursos disponíveis. Assim, a aplicação de algoritmos de otimização heurística é essencial para encontrar soluções de bom desempenho no espaço de busca.

Entre os métodos heurísticos disponíveis, foi adotado o algoritmo TPE (*Tree-structured Parzen Estimator*), apresentado na Seção 2.5, o algoritmo foi configurado para realizar 50 tentativas, sendo 15 aleatórias e 35 otimizadas. As tentativas aleatórias foram incluídas para explorar de maneira ampla o espaço de busca e fornecer uma base inicial para o algoritmo. A etapa de otimização subsequente utilizou as informações coletadas nas tentativas iniciais para ajustar progressivamente as recomendações, explorando regiões promissoras do espaço de busca.

4.3.2.3 Estimativa de Desempenho

No presente trabalho, a estimativa de desempenho foi conduzida utilizando validação cruzada, conforme detalhado na Sessão 2.7, a partir do treinamento completo do modelo. Essa abordagem foi escolhida para garantir uma avaliação mais robusta, minimizando a dependência de uma única divisão dos dados e permitindo que o modelo fosse avaliado em diferentes subconjuntos. A validação cruzada foi particularmente útil para evitar viés no processo de seleção de arquiteturas.

Devido ao alto custo computacional associado ao treinamento completo de todas as arquiteturas no espaço de busca, foi necessário mitigar esses custos, reduzindo o número de épocas para treinamento durante esta etapa. Essa redução foi fundamentada em observações feitas nas etapas anteriores, que demonstraram que a convergência dos modelos ocorria em um número menor de épocas do que inicialmente previsto. Essa estratégia permitiu que o processo de otimização fosse conduzido de maneira eficiente, sem comprometer a confiabilidade dos resultados obtidos.

4.4 Considerações Finais

Este capítulo descreveu as etapas da metodologia proposta, abordando desde o balanceamento de classes e pré-processamento de dados até a otimização de arquiteturas utilizando NAS. Também foram apresentados o conjunto de dados, as estratégias de avaliação e o ambiente experimental, com detalhes que asseguram a reprodutibilidade dos experimentos e embasam a aplicação das técnicas escolhidas para maximizar a generalização e o desempenho dos modelos treinados.

5 Resultados

Este capítulo apresenta e discute os resultados obtidos a partir da metodologia descrita na Seção 4.3. As análises estão organizadas por grau de ampliação das imagens, abrangendo as etapas de preparação de dados, com destaque para os achados nas subetapas de balanceamento de dados e aplicação de pré-processamento com aumento de representatividade. Além disso, são apresentados os resultados obtidos na etapa de pesquisa de arquitetura neural, incluindo a avaliação do desempenho das arquiteturas otimizadas.

5.1 Definição de Hiperparâmetros

Os hiper-parâmetros dos modelos foram mantidos consistentes em todas as etapas do estudo, com exceção do número de épocas. Na etapa de pesquisa de arquitetura neural, o número de épocas foi reduzido para mitigar o custo computacional dessa fase, considerando também que nas etapas anteriores os modelos apresentaram convergência em um número de épocas inferior ao inicialmente planejado. Os hiper-parâmetros adotados neste estudo estão detalhados na Tabela 3.

Tabela 3 – Hiper-parâmetros de Treinamento

Hiper-parâmetro	Preparação de Dados	Pesquisa de Arquitetura Neural.
Otimizador	Adam	Adam
Função de Perda	CrossEntropyLoss	CrossEntropyLoss
Taxa de Aprendizado	0.001	0.001
Tamanho de Lote	16	16
Resolução de Entrada	(384, 512)	(384, 512)
Quantidade de Épocas	120	50

Fonte: elaborada pelo Autor.

O otimizador utilizado foi o Adam, proposto por Kingma e Ba (2017), o qual é amplamente reconhecido como uma escolha eficaz para problemas de aprendizado de máquina. Foram utilizados os parâmetros padrão recomendados no estudo original: taxa de aprendizado 0.001, betas (0.9, 0.999) e $\epsilon = 1e - 8$. Esses valores são indicados como bons pontos de partida para a maioria dos problemas de aprendizado de máquina (KINGMA; BA, 2017).

O tamanho do lote foi fixado em 16, uma escolha que equilibra o uso eficiente de memória das aceleradoras gráficas e o desempenho computacional. Essa configuração não apenas garantiu a estabilidade do treinamento, mas também possibilitou a

execução simultânea de múltiplos experimentos na mesma GPU, otimizando os recursos computacionais disponíveis e reduzindo o tempo total de execução dos testes.

5.2 Preparação dos Dados

Nesta seção, são apresentados os resultados obtidos durante a etapa de preparação dos dados, com foco nas subetapas de balanceamento de classes e pré-processamento com aumento de representatividade. Inicialmente, foram avaliadas diferentes combinações de perfis de aumento geométrico para corrigir o desbalanceamento entre as classes. Em seguida, foi realizada a seleção do melhor pré-processamento aliado a perturbações, buscando melhorar a generalização dos modelos. Os detalhes e resultados específicos de cada etapa são apresentados nas subseções a seguir.

5.2.1 Balanceamento de Classes

O balanceamento de classes foi aplicado visando corrigir a desigualdade na distribuição das instâncias entre as classes no conjunto de dados, garantindo uma base mais equitativa para o treinamento do modelo. Para alcançar esse objetivo, foram definidos perfis de aumento geométrico que incluíram combinações de transformações como inversões e rotações. Cada perfil foi aplicado separadamente a ambas as classes, buscando identificar a proporção com melhor desempenho da arquitetura base na classificação. Na Tabela 4, são apresentadas as quantidades de imagens geradas para cada perfil de aumento em cada classe para o primeiro *fold*, os demais seguem as mesmas proporções de imagens geradas.

Tabela 4 – Quantidade de imagens geradas para o primeiro *fold*

Perfil de Aumento		Treino	Validação	Treino + Val	Teste	Total
Sem aumento	Câncer	40	5	45	10	55
	Normal	29	4	33	9	42
Inversões	Câncer	160	20	180	10	190
	Normal	116	16	132	9	141
Inversões + Rotação 45°	Câncer	480	60	540	10	550
	Normal	348	48	396	9	405
Inversões + Rotações 30° e 60°	Câncer	800	100	900	10	910
	Normal	580	80	660	9	669

Fonte: elaborada pelo Autor.

A combinação de perfis de aumento foi avaliada individualmente para cada classe, considerando ambas as ampliações do conjunto de dados. O estudo teve como objetivo identificar a proporção ideal para balancear as instâncias, minimizando a disparidade entre as classes e maximizando o desempenho do modelo base. Os resultados dessa análise, incluindo o desempenho dos modelos para cada perfil de aumento em cada ampliação, são

apresentados na Tabela 5 para as imagens com ampliação de 40x e na Tabela 6 para as ampliações de 100X.

Tabela 5 – Resultados do Balanceamento de Dados para imagens 40X

Perfis de aumento		F1-scr Médio	Câncer			Normal		
Câncer	Normal		F1-scr	Prec.	Recall	F1-scr	Prec.	Recall
Sem aumento	Sem aumento	0,701	0,695	0,825	0,600	0,707	0,614	0,833
Sem aumento	Inv+Rot 30° 60°	0,772	0,788	0,837	0,745	0,756	0,708	0,810
Sem aumento	Inv+Rot 45°	0,773	0,784	0,851	0,727	0,761	0,700	0,833
Sem aumento	Inversões	0,810	0,839	0,825	0,855	0,780	0,800	0,762
Inv+Rot 30° 60°	Sem aumento	0,760	0,785	0,808	0,764	0,736	0,711	0,762
Inv+Rot 30° 60°	Inv+Rot 30° 60°	0,783	0,788	0,886	0,709	0,779	0,698	0,881
Inv+Rot 30° 60°	Inv+Rot 45°	0,814	0,824	0,894	0,764	0,804	0,740	0,881
Inv+Rot 30° 60°	Inversões	0,866	0,869	0,977	0,782	0,863	0,774	0,976
Inv+Rot 45°	Sem aumento	0,835	0,843	0,915	0,782	0,826	0,760	0,905
Inv+Rot 45°	Inv+Rot 30° 60°	0,844	0,860	0,885	0,836	0,828	0,800	0,857
Inv+Rot 45°	Inv+Rot 45°	0,813	0,830	0,863	0,800	0,795	0,761	0,833
Inv+Rot 45°	Inversões	0,750	0,727	0,970	0,582	0,774	0,641	0,976
Inversões	Sem aumento	0,854	0,868	0,902	0,836	0,841	0,804	0,881
Inversões	Inv+Rot 30° 60°	0,791	0,815	0,830	0,800	0,767	0,750	0,786
Inversões	Inv+Rot 45°	0,802	0,822	0,846	0,800	0,782	0,756	0,810
Inversões	Inversões	0,835	0,843	0,915	0,782	0,826	0,760	0,905

Fonte: elaborada pelo Autor.

A avaliação dos perfis de aumento para a ampliação de 40X revelou que a combinação mais eficaz foi “Inversões + Rotações 30° e 60°” para a classe com diagnóstico de câncer e “Inversões” para a classe normal. Esses resultados são apresentados na Tabela 5. No entanto, um aspecto notável que contraria a hipótese inicial é que essa configuração acentuou ainda mais o desbalanceamento das classes, ampliando a proporção de imagens da classe com câncer, que já era a classe majoritária no dataset original. Apesar de o objetivo inicial ser equilibrar melhor as classes para favorecer o desempenho do modelo, a combinação que apresentou melhor desempenho reforça que, em alguns casos, proporções específicas entre as classes podem ser mais vantajosas, mesmo que não sejam equitativas.

Para as imagens com ampliação de 100X, os perfis de aumento “Inversões + Rotações 30° e 60°” para ambas as classes, e “Inversões + Rotação 45°” para a classe com diagnóstico de câncer em combinação com “Inversões + Rotações 30° e 60°” para a classe normal, apresentaram os melhores resultados de desempenho. Ambos os cenários demonstraram o mesmo desempenho final, como evidenciado na Tabela 6. Esses resultados reforçam a importância de considerar diferentes configurações de aumento para cada classe.

Os perfis de aumento “Sem aumento”, utilizados como *baseline* dos experimentos, não apresentaram desempenho competitivo em nenhuma das ampliações avaliadas. Para as imagens de 40X, essa configuração resultou no pior desempenho entre todas as combinações testadas. No caso das imagens de 100X, embora o desempenho tenha sido superior ao observado em 40X, ainda ficou significativamente abaixo das combinações que incluíam

Tabela 6 – Resultados do Balanceamento de Dados para imagens 100X

Perfis de aumento		F1-scr Médio	Câncer			Normal		
Câncer	Normal		F1-scr	Prec.	Recall	F1-scr	Prec.	Recall
Sem aumento	Sem aumento	0,833	0,852	0,868	0,836	0,814	0,795	0,833
Sem aumento	Inv+Rot 30° 60°	0,884	0,901	0,893	0,909	0,867	0,878	0,857
Sem aumento	Inv+Rot 45°	0,855	0,863	0,936	0,800	0,848	0,780	0,929
Sem aumento	Inversões	0,813	0,827	0,878	0,782	0,800	0,750	0,857
Inv+Rot 30° 60°	Sem aumento	0,814	0,816	0,930	0,727	0,813	0,722	0,929
Inv+Rot 30° 60°	Inv+Rot 30° 60°	0,906	0,917	0,926	0,909	0,894	0,884	0,905
Inv+Rot 30° 60°	Inv+Rot 45°	0,783	0,796	0,854	0,745	0,769	0,714	0,833
Inv+Rot 30° 60°	Inversões	0,801	0,826	0,833	0,818	0,776	0,767	0,786
Inv+Rot 45°	Sem aumento	0,896	0,906	0,941	0,873	0,886	0,848	0,929
Inv+Rot 45°	Inv+Rot 30° 60°	0,906	0,917	0,926	0,909	0,894	0,884	0,905
Inv+Rot 45°	Inv+Rot 45°	0,855	0,863	0,936	0,800	0,848	0,780	0,929
Inv+Rot 45°	Inversões	0,790	0,818	0,818	0,818	0,762	0,762	0,762
Inversões	Sem aumento	0,875	0,887	0,922	0,855	0,864	0,826	0,905
Inversões	Inv+Rot 30° 60°	0,854	0,868	0,902	0,836	0,841	0,804	0,881
Inversões	Inv+Rot 45°	0,824	0,838	0,880	0,800	0,809	0,766	0,857
Inversões	Inversões	0,852	0,875	0,860	0,891	0,829	0,850	0,810

Fonte: elaborada pelo Autor.

técnicas de aumento, como “Inversões + Rotações 30° e 60°”. Esses resultados destacam a relevância do aumento de dados no contexto de imagens histopatológicas, evidenciando que a ausência de manipulações para expandir e diversificar o conjunto de dados compromete a capacidade do modelo de generalizar e identificar padrões complexos.

Para a subetapa seguinte, os perfis de aumento “Inversões + Rotações 30° e 60°” e “Inversões” foram selecionados para as imagens com ampliação de 40x, enquanto para as imagens com ampliação de 100x a configuração “Inversões + Rotações 30° e 60°” foi escolhida para ambas as classes. A decisão pela última configuração em 100x, entre as duas que apresentaram desempenho equivalente, deve-se ao maior número de instâncias geradas, ampliando a diversidade do conjunto de dados e beneficiando potencialmente o treinamento do modelo.

Essas escolhas refletem as características distintas das classes analisadas. A classe “normal” apresenta um padrão mais consistente, facilitando a identificação pelo modelo. Por outro lado, a classe “câncer” é mais complexa, incluindo diferentes subtipos e graus histológicos, exigindo maior diversidade nas transformações aplicadas para representar suas variações adequadamente. Esses fatores evidenciam a complexidade intrínseca da classificação histopatológica e a importância de ajustes específicos na preparação dos dados para cada classe e ampliação.

5.2.2 Pré-processamento e Aumento de Representatividade

A análise da combinação entre pré-processamento e aumento de representatividade foi realizada para identificar as melhores configurações capazes de realçar características relevantes das imagens histopatológicas, ao mesmo tempo, em que minimizam os efeitos de artefatos que poderiam dificultar a classificação. O pré-processamento desempenha um papel essencial ao ajustar propriedades como contraste, iluminação e padrões de cor, destacando elementos fundamentais do tecido que auxiliam no diagnóstico. Por outro lado, o aumento de representatividade, particularmente voltado para variações de cor, busca simular alterações comuns em imagens histopatológicas, como diferenças nos padrões de coloração causadas por variações no corante ou nas condições de aquisição.

Estudos anteriores mostram que a combinação de técnicas de aumento de representatividade baseadas em cor com estratégias eficazes de pré-processamento tem o poder de melhorar o desempenho de redes neurais aplicadas à análise de imagens histopatológicas (TELLEZ et al., 2019; LAUANDE et al., 2022). Dessa forma, os experimentos realizados nesta etapa buscaram avaliar sistematicamente como essas duas estratégias, aplicadas em conjunto, podem impactar o desempenho do modelo. Os resultados obtidos são apresentados na Tabela 7 para imagens com ampliação de 40x e na Tabela 8 para imagens com ampliação de 100x.

Tabela 7 – F1-Score médio para Pré-processamento e Aumento em 40x.

Preprocessamento Aumento	Sem Aumento	HED		HSV		Brilho e Contraste
		Leve	Severo	Leve	Severo	
Sem Preprocessamento	0,787	0,822	0,797	0,445	0,522	0,894
stainGAN	0,809	0,769	0,840	0,463	0,478	0,815
stainNet	0,827	0,835	0,823	0,362	0,362	0,894
Vale et al. (2024) P49T1	0,786	0,747	0,819	0,362	0,449	0,738
Vale et al. (2024) P35T	0,772	0,735	0,747	0,445	0,362	0,825
Equal. H&E	0,854	0,660	0,757	0,484	0,362	0,362
Belfort et al. (2023)	0,822	0,835	0,860	0,362	0,362	0,844
Lauande. et al. (2022)	0,865	0,916	0,824	0,362	0,452	0,793

Fonte: elaborada pelo Autor.

A melhor combinação de pré-processamento e aumento de representatividade para as imagens com ampliação de 40x foi o método de pré-processamento proposto por Lauande et al. (2022), combinado com o aumento baseado no modelo HED leve. O pré-processamento estabelecido por Lauande et al. (2022) ajusta as propriedades de contraste e padrão de cor das imagens, realçando as características relevantes para o diagnóstico e melhorando a consistência visual das amostras. Por sua vez, o aumento HED simula variações nos canais de cor associados a esses corantes, gerando alterações que se assemelham às variações reais observadas em imagens histopatológicas. Os resultados, apresentados na Tabela 7, evidenciam o impacto positivo dessa configuração no desempenho do modelo.

Tabela 8 – F1-Score médio para Pré-processamento e Aumento em 100x

Pré-processamento Aumento	Sem Aumento	HED		HSV		Brilho e Contraste
		Leve	Severo	Leve	Severo	
Sem Pré-processamento	0,855	0,905	0,947	0,431	0,424	0,937
stainGAN	0,854	0,905	0,895	0,302	0,389	0,916
stainNet	0,845	0,916	0,916	0,410	0,483	0,844
Vale et al. (2024) P49T1	0,906	0,916	0,865	0,445	0,528	0,864
Vale et al. (2024) P35T	0,865	0,884	0,895	0,463	0,474	0,820
Equal. H&E	0,896	0,884	0,869	0,515	0,463	0,478
Belfort et al. (2023)	0,927	0,925	0,957	0,302	0,505	0,847
Lauande. et al. (2022)	0,844	0,916	0,895	0,389	0,477	0,885

Fonte: elaborada pelo Autor.

Para as imagens com ampliação de 100x, a melhor combinação de pré-processamento e aumento de representatividade foi o método de pré-processamento desenvolvido por Belfort et al. (2023), em conjunto com o aumento HED severo. O pré-processamento proposto pelos autores aplica equalização no canal cieLAB, realçando as características fundamentais das imagens histopatológicas, como bordas e texturas, cruciais para a diferenciação entre tecido saudável e canceroso. Essa abordagem foi particularmente eficaz em destacar elementos visuais relevantes para a classificação. Os resultados apresentados na Tabela 8 evidenciam que essa configuração foi a mais eficaz em termos de desempenho para as imagens com ampliação de 100x.

Nas Tabelas 7 e 8, é possível observar que as transformações baseadas no espaço de cor HSV apresentaram desempenho consistentemente inferior às outras abordagens de aumento para ambas as ampliações. Esse resultado pode ser atribuído à natureza das alterações geradas por essas transformações, que frequentemente resultam em imagens com padrões de cores irreais para o domínio das imagens histopatológicas. Alterações drásticas na tonalidade (hue) e saturação podem distorcer características importantes das imagens, dificultando o aprendizado do modelo e comprometendo sua capacidade de generalização.

Embora o objetivo do aumento aleatório seja simular variações realistas no padrão de cor, as perturbações geradas no espaço HSV podem criar cenários que não refletem condições observadas na prática clínica. Isso sugere que, para conjuntos de dados histopatológicos, é mais eficaz utilizar aumentos que mantenham maior fidelidade às características intrínsecas do tecido, como os baseados em transformações no espaço Hematoxilina e Eosina (HED).

Para ambas as ampliações, as transformações de brilho e contraste demonstraram um desempenho competitivo quando aplicadas sem pré-processamento, destacando-se como uma estratégia eficiente para introduzir variações de iluminação e contraste. No entanto, quando combinadas com os métodos de pré-processamento analisados, o desempenho dessas transformações sofreu uma redução significativa. Essa queda pode ser explicada

pelo fato de que os métodos de pré-processamento tendem a normalizar ou uniformizar os efeitos gerados pelas alterações de brilho e contraste, anulando, em parte, o impacto positivo dessas transformações.

A equalização de histograma aplicada sobre os canais Hematoxilina e Eosina (H&E), proposta para destacar as estruturas relevantes das imagens histopatológicas, apresentou resultados pouco competitivos em ambas as ampliações quando comparada a outros métodos de pré-processamento. Embora o conceito seja similar às abordagens de [Lauande et al. \(2022\)](#) e [Belfort et al. \(2023\)](#), que utilizam o CLAHE para destacar características importantes no domínio cieLAB e YUV, respectivamente, a aplicação da equalização diretamente nos canais H&E não resultou em ganhos de desempenho. Essa abordagem se baseia na premissa de que os canais H&E, amplamente utilizados em histopatologia digital, contêm informações estruturais essenciais para a classificação. No entanto, os resultados indicam que essa estratégia foi menos eficaz em realçar as características discriminativas das imagens.

O pré-processamento proposto por [Vale et al. \(2024\)](#), baseado na normalização de manchas utilizando o método Vahadane, não apresentou resultados competitivos em comparação aos demais métodos analisados neste estudo. [Vale et al. \(2024\)](#) destaca que a metodologia proposta inclui uma etapa essencial para a escolha da melhor imagem-alvo, realizada por meio de uma análise exaustiva que envolve a normalização de todo o conjunto de dados em relação a cada imagem-alvo, seguida do treinamento e avaliação de diferentes modelos. Essa abordagem é adaptada a duas arquiteturas distintas desenvolvidas no trabalho original, mas não foi implementada neste estudo devido às suas demandas computacionais intensivas. Ao utilizar apenas duas imagens específicas como alvos (P49T1 e P35T, destacadas por Vale em experimentos prévios), a flexibilidade da metodologia original foi limitada, o que pode ter impactado negativamente o desempenho do pré-processamento.

Dentre outras limitações da metodologia empregada neste estudo, destaca-se a ausência de otimização dos parâmetros internos dos métodos de pré-processamento baseados em CLAHE, utilizados por [Lauande et al. \(2022\)](#) e [Belfort et al. \(2023\)](#). Esses parâmetros poderiam ter sido ajustados de forma mais precisa para as características do dataset específico, o que possivelmente teria melhorado o desempenho observado. Da mesma forma, os métodos de pré-processamento baseados em aprendizado profundo, como os apresentados em [Shaban et al. \(2019\)](#) e [Kang et al. \(2021\)](#), foram utilizados com pesos pré-treinados em outros conjuntos de dados, sem serem ajustados ou re-treinados para o dataset deste estudo. Adicionalmente, as variações severas e leves dos aumentos e pré-processamentos aplicados não foram otimizadas. Uma exploração mais ampla desses parâmetros, aliada à sua adaptação ao conjunto de dados, poderia gerar melhores resultados.

Como resultado desta subetapa, foram selecionadas as melhores configurações de

pré-processamento e aumento de representatividade para cada ampliação. Para imagens de 40x, a configuração estabelecida por [Lauande et al. \(2022\)](#), combinada com o aumento HED leve, foi escolhida, obtendo o melhor desempenho para as imagens avaliadas. Para imagens de 100x, o pré-processamento desenvolvido por [Belfort et al. \(2023\)](#), aliado ao aumento HED severo, apresentou os melhores resultados. Os parâmetros dessas combinações foram definidos com base na análise sistemática realizada nesta etapa, e serão utilizados na próxima fase, destinada à aquisição e otimização da arquitetura neural. Ressalta-se, entretanto, que as limitações metodológicas, como a ausência de otimização dos parâmetros internos de pré-processamento e a utilização de pesos pré-treinados para os métodos baseados em aprendizado profundo, indicam oportunidades para melhorias futuras.

5.3 NAS: Pesquisa de Arquitetura Neural

Nesta sessão, serão apresentados os resultados obtidos na etapa de pesquisa de arquitetura neural (NAS), conduzida com base nas configurações de preparação de dados definidas nas sessões anteriores. Essa etapa teve como objetivo encontrar arquiteturas otimizadas para as imagens histopatológicas de câncer de pênis, considerando as ampliações de 40x e 100x sob as configurações de dados encontradas na etapa anterior.

As arquiteturas foram avaliadas conforme os critérios e metodologias descritos previamente, e as dez melhores configurações para cada ampliação são detalhadas nas Tabelas 9 e 10, respectivamente. Os resultados destacam as combinações de blocos de *stem* e *backbone* para cada estágio que apresentaram o melhor desempenho preditivo, reforçando o impacto da otimização baseada no espaço de busca definido. Além disso, será discutido o desempenho geral das arquiteturas e as principais diferenças observadas entre as ampliações.

5.3.1 Arquiteturas 40x

Para as imagens com ampliação de 40x, a melhor arquitetura encontrada apresentou uma configuração composta por um *stem* profundo, responsável pela extração inicial de características. Nos blocos de *backbone*, a arquitetura utilizou o módulo de atenção BAM adicionado ao ramo residual (R-BAM) no primeiro estágio, suprimindo e realçando características que serão somadas às saídas dos blocos. No segundo e terceiro estágios, foi implementado o módulo CBAM adicionado à saída do ramo residual (R-CBAM), também suprimindo e realçando características que serão somadas às saídas dos blocos convolucionais. Para o último estágio, a escolha do otimizador foi o bloco ResNet padrão sem adição de atenção. A melhor arquitetura para as imagens com ampliação de 40x é representada na Figura 30.

A frequência de aparecimento dos blocos *stem* nas dez melhores arquiteturas foi

Tabela 9 – Arquiteturas Otimizadas para Ampliação de 40x – Top 10 Configurações

Trial	Blc. Stem	Blc. Stg.1	Blc. Stg.2	Blc. Stg.3	Blc. Stg.4	F1-Score
20	Profundo	R-BAM	C-CBAM	Padrão	Padrão	0,894
13	Padrão	C-BAM	Padrão	C-CBAM	R-BAM	0,896
47	Profundo	R-BAM	R-BAM	Padrão	Padrão	0,896
8	Padrão	C-CBAM	C-CBAM	C-BAM	C-CBAM	0,896
44	Profundo	R-CBAM	R-CBAM	R-CBAM	C-CBAM	0,904
37	Prof. Esc.	R-BAM	C-CBAM	C-CBAM	Padrão	0,905
0	Profundo	R-BAM	C-CBAM	C-BAM	Padrão	0,905
6	Padrão	C-CBAM	C-BAM	Padrão	C-CBAM	0,905
19	Profundo	R-BAM	C-CBAM	R-CBAM	Padrão	0,906
41	Profundo	R-BAM	R-CBAM	R-CBAM	Padrão	0,927

Fonte: elaborada pelo Autor.

majoritariamente dominada pelo *stem* profundo, utilizado em 6 das configurações. Este bloco foi projetado para aprimorar a captura de características de baixo nível, ampliando a flexibilidade por meio de mais filtros e permitindo a correlação dessas características ainda nas resoluções mais altas dos mapas de características. A predominância do *stem* profundo reflete sua capacidade de adaptar-se melhor às necessidades de extração inicial em imagens histopatológicas, especialmente devido à complexidade das texturas presentes nesse tipo de imagem.

Para o primeiro estágio do *backbone*, 7 das 10 melhores arquiteturas, utilizam atenção aplicada ao ramo residual. Essa escolha visa atenuar características que poderiam dificultar o aprendizado nos blocos subsequentes após a soma, ou escalar características úteis presentes no ramo residual, garantindo que informações relevantes sejam destacadas e propagadas de maneira mais eficaz. Essa estratégia permite que o modelo priorize características mais discriminativas já nos estágios iniciais do aprendizado.

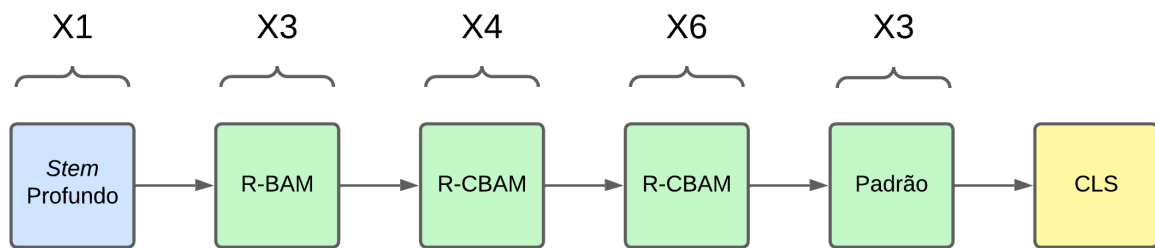
Para o segundo estágio do *backbone*, observou-se uma predominância de arquiteturas que aplicam a atenção na saída convolucional do bloco, com 6 das 10 melhores arquiteturas utilizando essa estratégia. Essa abordagem sugere que a atenção aplicada diretamente à saída convolucional ajuda a refinar as características geradas pelo bloco, suprimindo informações irrelevantes e amplificando padrões relevantes antes de serem combinados com o ramo residual. Entretanto, é importante destacar que a melhor arquitetura encontrada para as imagens com ampliação de 40x utiliza a atenção no ramo residual no segundo estágio. Quanto ao tipo de atenção, o módulo CBAM predominou, aparecendo em 7 das 10 melhores arquiteturas para o segundo estágio.

No terceiro estágio do *backbone*, houve uma redução na utilização de atenção aplicada na saída convolucional, presente em 4 das 10 melhores arquiteturas, enquanto os blocos padrão tiveram maior frequência de uso. Esse comportamento sugere que, nos estágios mais profundos, a necessidade de refinamento direto na saída convolucional diminui.

No entanto, a melhor arquitetura para imagens com ampliação de 40x manteve a atenção no ramo residual, destacando sua eficácia em preservar padrões relevantes e mitigar ruídos antes da soma com a saída convolucional. Quanto ao tipo de atenção, o módulo CBAM continuou predominante, aparecendo em 5 das 10 melhores configurações, reforçando sua consistência em diferentes estágios do aprendizado.

No último estágio do *backbone*, observou-se uma predominância dos blocos ResNet padrão, utilizados em 6 das 10 melhores arquiteturas, indicando que, nos estágios mais profundos, a necessidade de mecanismos de atenção é reduzida. Entre as arquiteturas que aplicaram atenção, o CBAM na saída convolucional apareceu em 3 configurações, evidenciando sua utilidade no refinamento final das características antes da saída do *backbone*. A melhor arquitetura para ampliação de 40x utilizou um bloco ResNet padrão nesse estágio.

Figura 30 – Melhor arquitetura para imagens com ampliação de 40X.



Fonte: elaborada pelo Autor.

5.3.2 Arquiteturas 100x

Para as imagens com ampliação de 100x, a melhor arquitetura utilizou um stem profundo para capturar características de baixo nível e ajustar a resolução das imagens. Nos estágios do *backbone*, foram aplicados módulos de atenção estrategicamente: BAM na saída convolucional no primeiro estágio (C-BAM), CBAM no ramo residual no segundo estágio (R-CBAM), BAM na saída convolucional no terceiro estágio (C-BAM) e BAM no ramo residual no quarto estágio (R-BAM). A melhor arquitetura para as imagens com ampliação de 100x é representada na Figura 31.

Nas camadas de stem, observou-se uma predominância do uso do stem profundo, mantendo o padrão identificado nas arquiteturas das imagens de 40x. Esse bloco foi utilizado em 6 das 10 melhores configurações, destacando sua importância para a extração inicial de características de baixo nível. O stem profundo se mostrou eficiente em capturar as texturas complexas presentes em imagens histopatológicas, correlacionando características básicas em resoluções mais altas e permitindo que o aprendizado hierárquico dos estágios subsequentes fosse mais robusto. A consistência no uso desse bloco nas duas ampliações reforça sua eficácia para o domínio de imagens histopatológicas.

Tabela 10 – Arquiteturas Otimizadas para Ampliação de 100x – Top 10 Configurações.

Trial	Blc. Stem	Blc. Stg.1	Blc. Stg.2	Blc. Stg.3	Blc. Stg.4	F1-Score
11	Profundo	R-CBAM	C-CBAM	R-CBAM	C-BAM	0,936
31	Profundo	C-BAM	C-BAM	C-BAM	R-BAM	0,936
0	Padrão	Padrão	C-BAM	R-BAM	C-CBAM	0,937
7	Prof. Esc.	C-CBAM	C-CBAM	R-BAM	R-CBAM	0,937
13	Profundo	Padrão	Padrão	R-BAM	C-CBAM	0,937
14	Prof. Esc.	Padrão	Padrão	Padrão	C-CBAM	0,937
28	Profundo	R-BAM	R-CBAM	C-BAM	Padrão	0,937
8	Profundo	C-CBAM	R-CBAM	Padrão	R-BAM	0,948
4	Prof. Esc.	C-CBAM	R-CBAM	C-CBAM	R-BAM	0,958
16	Profundo	C-BAM	R-CBAM	C-BAM	R-BAM	0,958

Fonte: elaborada pelo Autor.

No primeiro estágio, a atenção na saída convolucional predominou, aparecendo em 5 das 10 melhores arquiteturas, com o mecanismo CBAM utilizado em 4 delas. Entretanto, a melhor arquitetura empregou o BAM na saída convolucional. Além disso, blocos ResNet padrão foram mais frequentes do que nas arquiteturas de 40x, presentes em 3 configurações, indicando que, em algumas situações, o desempenho inicial pode ser alcançado sem atenção adicional, possivelmente devido às características das imagens de 100x.

O segundo estágio, o mecanismo CBAM predominou, aparecendo em 6 das 10 melhores arquiteturas. Observou-se um equilíbrio na aplicação da atenção, com 4 arquiteturas utilizando-a no ramo convolucional e 4 no ramo residual. Destaca-se que as 4 melhores arquiteturas optaram pela aplicação da atenção no ramo residual, evidenciando sua eficácia em refinar características relevantes antes da soma com a saída convolucional.

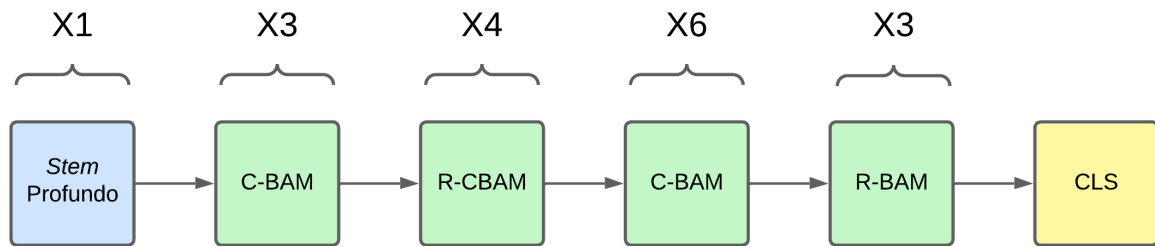
Para o terceiro estágio do *backbone*, o mecanismo BAM predominou, sendo utilizado em 6 das 10 melhores arquiteturas. Quanto à aplicação da atenção, observou-se um equilíbrio entre os ramos convolucional e residual, com ambos aparecendo em 4 das 10 configurações. Entretanto, as arquiteturas que aplicaram atenção no ramo convolucional apresentaram melhor desempenho geral.

No último estágio, o mecanismo BAM predominou, presente em 5 das 10 melhores arquiteturas. Com a aplicação no ramo residual sendo a mais frequente, sendo apresentada em 5 das 10 melhores arquiteturas.

5.3.3 Análise Comparativa das Arquiteturas

Entre os achados para as arquiteturas de 40x e 100x, algumas considerações relevantes podem ser feitas. Primeiramente, o uso consistente do stem profundo em 6 das 10 melhores arquiteturas para ambas as ampliações reforçam sua importância na extração inicial de características de baixo nível. Esse bloco demonstrou ser fundamental

Figura 31 – Melhor arquitetura para imagens com ampliação de 100X.



para captar as texturas complexas e os detalhes relevantes das imagens histopatológicas, estabelecendo uma base robusta para o aprendizado hierárquico nos estágios subsequentes.

Apesar dessa consistência, diferenças estratégicas nos estágios do *backbone* destacam a adaptação da arquitetura às necessidades específicas de cada ampliação. No primeiro estágio, as arquiteturas de 40x priorizaram a atenção no ramo residual, enquanto as de 100x optaram pela atenção na saída convolucional. Em ambas, o módulo CBAM foi amplamente utilizado, mas o BAM se destacou na melhor configuração de 100x. No segundo estágio, houve equilíbrio entre a aplicação de atenção no ramo residual e na saída convolucional, mas a atenção no ramo residual demonstrou maior eficácia, especialmente nas 4 melhores arquiteturas de 100x. Já no terceiro estágio, o BAM predominou nas arquiteturas de 100x, com um equilíbrio entre os ramos convolucional e residual, embora a atenção na saída convolucional tenha mostrado melhores resultados em termos de desempenho geral. No último estágio, houve uma clara predominância do bloco padrão nas imagens 40x, enquanto para as imagens 100x a atenção BAM foi mais frequente, com a atenção aplicada sobre o ramo residual sendo mais robusta.

Essas diferenças estratégicas nos estágios do backbone, aliadas à consistência no stem profundo, evidenciam a importância de uma seleção criteriosa e adaptativa dos blocos ao longo da arquitetura. Essa abordagem permite que as arquiteturas se ajustem às especificidades das ampliações, maximizando o desempenho preditivo e otimizando o processo de aprendizado em imagens histopatológicas.

5.4 Desempenho nas etapas do Método

Ao longo das etapas da metodologia, foi possível observar uma evolução consistente no desempenho dos modelos, com melhorias significativas ao final de cada etapa. Essa evolução reforça a importância de ajustes criteriosos na preparação dos dados, desde o balanceamento de classes e aplicação de aumentos de representatividade até a escolha adequada dos métodos de pré-processamento. Cada ajuste contribuiu para a construção de uma base de dados mais robusta, o que impactou diretamente no desempenho dos

modelos.

Tabela 11 – Comparação entre subetapas.

Etapa	F1-Score	
	40x	100x
Inicial	0,701	0,833
Balanceamento de Classes	0,866	0,905
Pré-processamento e Aumento de Representatividade	0,916	0,957
Pesquisa de Arquitetura Neural	0,927	0,958

Fonte: elaborada pelo Autor.

Na etapa inicial de balanceamento de classes, o impacto foi evidente. O F1-score para as imagens de 40x subiu de 0,701 para 0,866, enquanto para as imagens de 100x, o valor passou de 0,833 para 0,905. Esses ajustes foram fundamentais para reduzir os vieses iniciais no conjunto de dados e oferecer condições mais equilibradas para o treinamento do modelo.

Com a aplicação de pré-processamento e aumento de representatividade, houve um novo salto no desempenho. Para as imagens de 40x, o F1-score atingiu 0,916, e para as imagens de 100x, o valor subiu para 0,957. Essa etapa foi crucial para realçar características importantes das imagens histopatológicas e aumentar a generalização dos modelos, simulando variações realistas e minimizando os efeitos de artefatos.

Por fim, na etapa de pesquisa de arquitetura neural (NAS), as melhores arquiteturas atingiram F1-scores ainda mais elevados, consolidando o impacto de uma abordagem integrada e otimizada. Para as imagens de 40x, a melhor arquitetura alcançou 0,927, enquanto para as imagens de 100x, o desempenho foi ainda maior, chegando a 0,958.

Na Tabela 11, os resultados dessas etapas estão sintetizados, evidenciando que os ajustes sucessivos na metodologia foram essenciais para maximizar o potencial dos modelos.

5.5 Comparação com trabalhos relacionados

Nesta sessão, o desempenho do trabalho proposto será comparado com outros estudos que utilizaram a mesma base de dados, permitindo uma análise comparativa em relação. Os trabalhos selecionados para comparação incluem os estudos de [Lauande et al. \(2022\)](#), [Belfort et al. \(2023\)](#), [Vale et al. \(2024\)](#) e [Lauande et al. \(2024\)](#), que adotaram abordagens distintas para a análise de imagens histopatológicas.

Para as imagens com ampliação de 40x, o método proposto apresenta resultados superiores aos obtidos por [Belfort et al. \(2023\)](#), embora fique abaixo dos alcançados por [Lauande et al. \(2022\)](#) e [Lauande et al. \(2024\)](#). Já para as imagens com ampliação de 100x,

Tabela 12 – Comparação com trabalhos relacionados.

Trabalho	Método	Resultados (F1-Score)	
		40x	100x
(LAUANDE. et al., 2022)	CLAHE no canal Y do YUV + Densenet 201 + Transfer Learning	97,39(+/-2,13)	97,31(+/-3,62)
(BELFORT et al., 2023)	CLAHE no canal B do cieLAB + Densenet 121 + Modelo em Cascata	91,2(+/-2,3)	92,4(+/-1,5)
(VALE et al., 2024)	Pré-processamento Vahadane + Densenet 201 + Mecanismos de atenção	—	93,1(+/-4,6)
(LAUANDE et al., 2024)	CLAHE no canal Y do YUV + Transferência de domínio + Arquitetura Híbrida	95,78(+/-4,39)	94,99(+/-3,88)
Método desenvolvido	Balanceamento de Classes + Otimização de Pré-processamento e aumentos + Pesquisa de Arquitetura Neural (NAS)	92,25(+/-3,66)	95,99(+/-4,36)

Fonte: elaborada pelo Autor.

os resultados do método proposto superam os encontrados em [Belfort et al. \(2023\)](#), [Vale et al. \(2024\)](#) e [Lauande et al. \(2024\)](#), ficando abaixo apenas do desempenho apresentado por [Lauande. et al. \(2022\)](#).

No entanto, é importante observar que os trabalhos de [Lauande. et al. \(2022\)](#), [Belfort et al. \(2023\)](#) e [Vale et al. \(2024\)](#) foram conduzidos com uma versão anterior da base de dados, na qual alguns rótulos de instâncias foram posteriormente revisados e alterados. Isso limita a comparabilidade direta dos resultados desses estudos com o trabalho atual. Apenas o estudo [Lauande et al. \(2024\)](#) utilizou a mesma versão do conjunto de dados que o presente trabalho, reforçando a relevância da comparação específica com este estudo e destaca as contribuições do método proposto.

É importante destacar que todos os estudos apresentados empregam, em algum nível, técnicas de transferência de aprendizado, com o trabalho [Lauande et al. \(2024\)](#) apresentando um diferencial ao utilizar transferência de domínio de um dataset histopatológico com um número significativamente maior de imagens. Esse aspecto demonstra um caminho promissor para o aprimoramento do desempenho do modelo proposto, especialmente em cenários onde grandes volumes de dados externos podem ser aproveitados. No entanto, mesmo sem adotar técnicas de transferência de aprendizado, o presente trabalho conseguiu apresentar resultados competitivos, reforçando a eficiência da metodologia desenvolvida.

As contribuições do método proposto se destacam, especialmente pela aquisição de uma arquitetura otimizada para o dataset utilizado, por meio da pesquisa de arquitetura

neural (NAS). Além disso, a etapa de preparação de dados, que envolveu o balanceamento de classes, pré-processamento e aumento de representatividade, foi fundamental para a melhoria progressiva dos resultados. A evolução do desempenho ao longo das subetapas da metodologia demonstra a importância de ajustes criteriosos na preparação dos dados e evidência que, mesmo sem grandes bases externas, é possível alcançar um modelo robusto e competitivo. Essas contribuições tornam o método proposto uma solução relevante para a análise de imagens histopatológicas em cenários com recursos limitados.

5.6 Questões de Pesquisa

Dada as questões de pesquisa estabelecidas no início deste trabalho, esta seção apresenta as respostas fundamentadas nos resultados obtidos. Cada questão foi analisada com base nos experimentos conduzidos ao longo das etapas metodológicas. A seguir, são discutidas as principais conclusões extraídas dos achados experimentais.

- Qual a proporção ideal entre as classes no conjunto de dados que maximiza o desempenho do modelo, evitando viés de aprendizado e assegurando a generalização para dados desbalanceados, como observado no problema em questão?
 - Os experimentos indicaram que a proporção ideal não necessariamente exige um balanceamento perfeito entre as classes. Para as imagens de 40x, a melhor configuração envolveu um aumento seletivo na classe com diagnóstico de câncer, utilizando transformações de inversão e rotações (30° e 60°), enquanto a classe normal se beneficiou apenas de inversões. Já para as imagens de 100x, a melhor configuração consistiu na aplicação da transformação de inversão e rotações (30° e 60°) em ambas as classes. Apesar de a proporção resultante ainda ser desbalanceada, essas combinações maximizaram o desempenho do modelo, evidenciando que diferentes ampliações podem demandar estratégias distintas de balanceamento.
- Qual a combinação mais eficiente de estratégias de pré-processamento e aumento de representatividade que melhora a qualidade das imagens histopatológicas de câncer de pênis, potencializa a generalização do modelo e mitiga os efeitos de variações indesejadas nas características visuais?
 - Para as imagens de 40x, a melhor combinação foi o método de pré-processamento proposto por [Lauande et al. \(2022\)](#), aliado ao aumento de representatividade baseado no modelo HED leve. Esse conjunto de técnicas ajustou contrastes e padrões de cor, tornando as imagens mais homogêneas e realçando características relevantes.

- Para as imagens de 100x, o pré-processamento de [Belfort et al. \(2023\)](#), que emprega equalização no espaço cieLAB, combinado com o aumento HED severo, apresentou o melhor desempenho. Esse método destacou texturas e estruturas fundamentais para a classificação.
 - As transformações baseadas em HED mostraram um desempenho superior às demais, isso se justifica tendo em vista que esta transformação cria perturbações semelhantes às existentes no domínio histopatológico.
 - Transformações no espaço HSV demonstraram resultados inferiores, possivelmente por gerar padrões de coloração irreais, enquanto transformações de brilho e contraste mostraram eficácia isoladamente, mas seu impacto foi reduzido quando combinados com outros métodos de pré-processamento.
- Quais configurações de camadas de *stem* melhor extraem características de baixo nível, como texturas, nas imagens histopatológicas de câncer de pênis, e como essas configurações impactam o desempenho final do modelo?
 - O *stem* profundo foi a configuração mais eficaz para ambas as ampliações (40x e 100x), sendo utilizado em 6 das 10 melhores arquiteturas. Esse modelo se destacou por proporcionar uma extração inicial mais robusta das características de baixo nível, permitindo que padrões texturais fossem bem capturados desde os primeiros estágios do processamento. Sua presença nas melhores arquiteturas demonstra sua importância na construção de representações discriminativas para a classificação histopatológica.
 - Como a adoção de mecanismos de atenção influenciam a capacidade dos modelos de extrair representações discriminativas e relevantes nas imagens histopatológicas de câncer de pênis?
 - A adoção de mecanismos de atenção influenciou positivamente a capacidade dos modelos de extrair representações discriminativas e relevantes nas imagens histopatológicas de câncer de pênis. A aplicação de atenção no ramo residual permitiu suprimir características irrelevantes, como ruídos, que poderiam comprometer a extração de informações úteis nas camadas subsequentes. Já a adição de atenção na saída convolucional demonstrou eficácia na supressão de informações redundantes e no realce de padrões mais discriminativos, otimizando o aprendizado do modelo.
 - A evolução do desempenho entre a etapa de preparação de dados e os resultados obtidos na pesquisa de arquitetura neural (NAS) reforça a importância da alocação estratégica desses mecanismos ao longo da rede. As melhores arquiteturas identificadas apresentaram padrões específicos de posicionamento dos módulos de atenção, indicando que sua distribuição impacta diretamente

a capacidade do modelo de distinguir padrões histopatológicos relevantes, contribuindo para um aumento significativo no F1-score final.

- Como a pesquisa de arquitetura neural pode melhorar os resultados a partir de uma arquitetura otimizada para o domínio histopatológico de câncer de pênis?
 - A pesquisa de arquitetura neural (NAS) demonstrou ser uma abordagem eficaz para aprimorar os resultados da classificação histopatológica de câncer de pênis, permitindo a identificação de configurações otimizadas para o problema em questão. Apesar do incremento relativamente modesto entre os resultados da etapa de preparação de dados e aqueles obtidos com NAS, a melhoria observada reforça o potencial dessa técnica na adaptação da arquitetura ao domínio específico.
 - A presença de arquiteturas que não alcançaram bons resultados evidencia a necessidade de otimização, pois a escolha inadequada da arquitetura pode comprometer significativamente o desempenho do modelo. O processo de NAS possibilitou a seleção de blocos estruturais mais eficientes para diferentes estágios da rede, ajustando a extração e o refinamento das características de forma adaptativa.

5.7 Limitações

Embora os resultados obtidos neste estudo representem avanços na classificação de imagens histopatológicas de câncer de pênis, algumas limitações devem ser consideradas. Essas restrições podem ter influenciado o desempenho do modelo e devem ser levadas em conta para aprimoramentos futuros.

Uma das principais limitações está relacionada à base de dados utilizada. A maioria dos trabalhos correlatos foi realizada com uma versão anterior do dataset, na qual houve revisões e correções nos rótulos das instâncias. Essa diferença dificulta uma comparação direta entre os resultados, pois os modelos foram treinados e avaliados em conjuntos de dados potencialmente distintos. Além disso, a distribuição aleatória dos *folds* na validação cruzada pode introduzir variações nos desempenhos relatados, tornando as análises comparativas mais complexas.

Outra limitação refere-se à ausência de técnicas de transferência de aprendizado e adaptação de domínio. A maioria dos estudos relacionados empregou essas abordagens, o que pode ter contribuído para melhores resultados, especialmente em cenários com bases de dados reduzidas, como a utilizada neste trabalho. A implementação de transferência de aprendizado não foi viável neste estudo, pois exigiria o pré-treinamento de cada arquitetura gerada pela NAS, o que demandaria um alto custo computacional.

O espaço de busca definido para a NAS também impõe restrições. O conjunto de blocos arquiteturais explorado foi limitado a um número pré-definido de alternativas, o que pode ter restringido a descoberta de configurações ainda mais eficientes para a tarefa de classificação. Além disso, a necessidade de equilibrar o custo computacional com a busca por arquiteturas otimizadas levou a uma exploração incompleta do espaço de busca, impedindo possivelmente a identificação da melhor configuração global dentro desse espaço.

Além das limitações relacionadas à busca arquitetural, este estudo não explorou modelos baseados em *Transformers*. Embora essas arquiteturas tenham superado abordagens convolucionais em diversas tarefas de visão computacional (ZENG et al., 2021; TULI et al., 2021), sua aplicação na histopatologia digital ainda enfrenta desafios. *Transformers* requerem grandes quantidades de dados rotulados para treinamento eficaz, o que é frequentemente inviável nesse domínio (MA et al., 2023; JIANG et al., 2023). Estratégias como pré-treinamento em conjuntos de dados maiores e a combinação de blocos convolucionais com *Transformers* podem mitigar essa limitação, conforme explorado por Lauande et al. (2024). No entanto, no contexto deste estudo, a necessidade de treinar do zero cada arquitetura sugerida pela NAS tornou computacionalmente inviável a implementação dessas técnicas, justificando a exclusão de arquiteturas baseadas em *Transformers*.

Por fim, o ajuste fino de hiperparâmetros representa outra limitação do estudo. Parâmetros como as taxas de aprendizado do otimizador, os limiares de perturbação nos aumentos de dados e as configurações dos métodos de pré-processamento foram definidos empiricamente, sem uma busca exaustiva por ajustes mais precisos. A otimização desses fatores pode levar a melhorias adicionais no desempenho do modelo e deve ser considerada em trabalhos futuros.

5.8 Trabalhos Futuros

Com base nas limitações identificadas, algumas direções para trabalhos futuros podem ser exploradas para aprimorar a metodologia proposta e ampliar seu impacto na análise de imagens histopatológicas de câncer de pênis.

Uma das principais melhorias sugeridas é a adoção de técnicas de transferência de aprendizado e adaptação de domínio. Métodos que exploram o aprendizado a partir de bases histopatológicas mais extensas podem melhorar a capacidade de generalização do modelo, permitindo que ele aproveite características extraídas de grandes conjuntos de dados. Essa abordagem já foi empregada em estudos anteriores, como o de Lauande et al. (2024), e demonstrou impacto positivo no desempenho dos modelos.

Outro aspecto relevante é a otimização dos hiperparâmetros. O presente estudo

utilizou configurações predefinidas para componentes do pré-processamento, como os parâmetros do CLAHE e os limiares das perturbações aplicadas no aumento de cor e brilho. Além disso, hiperparâmetros internos da arquitetura neural, como taxas de aprendizado e regularizações, não foram ajustados de forma exaustiva. Trabalhos futuros podem explorar abordagens mais robustas para a otimização desses parâmetros, visando maximizar o desempenho da rede.

A expansão do espaço de busca na pesquisa de arquitetura neural (NAS) também representa uma oportunidade de aprimoramento. A metodologia atual restringiu a seleção dos blocos arquiteturais a um conjunto pré-definido de alternativas, o que pode ter limitado a descoberta de arquiteturas mais eficientes. Ampliar esse espaço para incluir uma gama maior de configurações, a abordagem *chain-structured neural networks* que introduz criação de uma rede baseada em um grafo cíclico direcionado onde os vértices são associados a operações (como convoluções padrão, convoluções dilatadas e operações de *pooling*) e as arestas o fluxo das características (ELSKEN; METZEN; HUTTER, 2019). Essa abordagem pode levar a modelos mais robustos e adaptáveis ao problema em questão, permitindo um aumento da complexidade do espaço de busca, possibilitando a descoberta de arquiteturas mais otimizadas para a classificação de imagens histopatológicas.

Além disso, é fundamental reduzir o custo computacional da pesquisa NAS. A alta demanda computacional restringiu a quantidade de experimentos viáveis, tornando inviável uma busca mais ampla por soluções alternativas. Estratégias baseadas em métodos proxies podem ser adotadas para mitigar esse problema, incluindo estimativas de baixa fidelidade, que reduzem o tempo de busca ao utilizar interrupção antecipada do treinamento, imagens de menor resolução e subconjuntos reduzidos dos dados de treinamento (LIU; ZHANG; JIN, 2022).

6 Conclusão

O câncer de pênis, embora raro, causa impactos severos na saúde física, psicológica e social dos pacientes, sendo mais prevalente em regiões onde fatores socioeconômicos e culturais dificultam o diagnóstico precoce, resultando em prognósticos desfavoráveis. A análise histopatológica é o método padrão para a detecção da doença, porém sua subjetividade pode levar a variações nos diagnósticos e atrasos no tratamento. Nesse contexto, técnicas baseadas em aprendizado profundo podem auxiliar na padronização das avaliações, reduzindo o tempo necessário para diagnósticos mais precisos e melhorando as taxas de sobrevida.

Para enfrentar esses desafios, este trabalho desenvolveu uma metodologia baseada em aprendizado profundo, estruturada em múltiplas etapas, para otimizar as imagens histopatológicas de câncer de pênis e a arquitetura neural responsável pelo diagnóstico. O processo metodológico foi delineado de forma que a melhor configuração encontrada em uma etapa foi utilizada como ponto de partida para a seguinte, incrementando o desempenho da metodologia gradativamente.

A metodologia proposta neste trabalho trouxe contribuições significativas para a área de classificação histopatológica baseada em aprendizado profundo. Primeiramente, destaca-se a proposição de uma abordagem estruturada para a preparação de dados, incluindo técnicas de balanceamento de classes, pré-processamento e aumento de representatividade. Essa estrutura demonstrou um impacto positivo na robustez do modelo, evidenciando a importância de uma preparação criteriosa dos dados para aumentar a capacidade de generalização das redes neurais. Além disso, essa metodologia serve como uma base sólida que pode ser utilizada e aprimorada em futuros trabalhos voltados para o diagnóstico automatizado de câncer de pênis.

Outra contribuição relevante está na investigação estratégica de mecanismos de atenção e na otimização da extração de características por meio da configuração das camadas *stem*. A análise conduzida permitiu avaliar como diferentes configurações dessas camadas impactam na captura de características de baixo nível, como texturas, fundamentais para a classificação histopatológica. Além disso, a exploração do posicionamento ideal dos módulos de atenção ao longo da arquitetura revelou configurações mais eficazes para realçar padrões morfológicos discriminativos, fornecendo diretrizes para futuras implementações de redes neurais nesse domínio.

Por fim, o desenvolvimento de uma arquitetura otimizada para o problema de classificação histopatológica de câncer de pênis representa uma das principais contribuições deste estudo. Utilizando uma abordagem de pesquisa de arquitetura neural baseada em

células (NAS), foi possível identificar configurações de rede mais adequadas ao conjunto de dados. Além da obtenção de uma arquitetura competitiva, a aplicação da metodologia NAS evidenciou seu potencial como ferramenta para descoberta de modelos otimizados, podendo ser explorada em outros contextos da patologia digital.

Entretanto, a aplicação da metodologia proposta a outros cenários da histopatologia digital deve ser realizada integralmente, desde as etapas de preparação de dados e otimização da arquitetura, sendo devidamente ajustadas ao novo conjunto de imagens. A aplicação direta dos achados deste trabalho em diferentes domínios histopatológicos pode não garantir um desempenho ideal, uma vez que características como coloração, padrões morfológicos e variações intrínsecas aos tecidos diferem entre conjuntos de dados. Isso é evidenciado pelas diferenças observadas entre as ampliações de 40x e 100x do mesmo tecido, que demandaram estratégias distintas para balanceamento de classes, pré-processamento e configuração arquitetural. Dessa forma, para adaptar a metodologia a outros contextos, é essencial aplicar a metodologia proposta, não apenas os achados para as imagens utilizadas neste trabalho.

Parte do presente trabalho — especificamente a etapa referente à Pesquisa de Arquitetura Neural (NAS) — foi apresentada na HCIST 2024 - *International Conference on Health and Social Care Information Systems and Technologies*. Nessa ocasião, evidenciou-se o potencial da automação na otimização de redes neurais para a análise histopatológica, bem como o papel dos mecanismos de atenção na extração de padrões discriminativos. O processo de preparação dos dados, que incluiu estratégias adicionais de balanceamento e normalização, foi posteriormente incorporado como evolução desse trabalho, dando continuidade ao aperfeiçoamento da metodologia.

Referências

ACANFORA, G.; CARILLO, A. M.; DELLO IACOVO, F.; SALATIELLO, M.; PISAPIA, P.; BELLEVICINE, C.; TRONCONE, G.; VIGLIAR, E. Interobserver variability in cytopathology: How much do we agree? *Cytopathology*, v. 35, n. 4, p. 444–453, 2024. Disponível em: <<https://doi.org/10.1111/cyt.13378>>. Citado na página 23.

ARESTA, G.; ARAÚJO, T.; KWOK, S.; CHENNAMSETTY, S. S.; SAFWAN, M.; ALEX, V.; MARAMI, B.; PRASTAWA, M.; CHAN, M.; DONOVAN, M.; FERNANDEZ, G.; ZEINEH, J.; KOHL, M.; WALZ, C.; LUDWIG, F.; BRAUNEWELL, S.; BAUST, M.; VU, Q. D.; TO, M. N. N.; KIM, E.; KWAK, J. T.; GALAL, S.; SANCHEZ-FREIRE, V.; BRANCATI, N.; FRUCCI, M.; RICCIO, D.; WANG, Y.; SUN, L.; MA, K.; FANG, J.; KONE, I.; BOULMANE, L.; CAMPILHO, A.; ELOY, C.; POLÓNIA, A.; AGUIAR, P. Bach: Grand challenge on breast cancer histology images. *Medical Image Analysis*, v. 56, p. 122–139, 2019. ISSN 1361-8415. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1361841518307941>>. Citado na página 24.

BELFORT, F.; SILVA, I.; SILVA, A.; PAIVA, A. Detecção de câncer peniano em imagens histopatológicas usando redes neurais convolucionais em cascata. In: *Anais do XXIII Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2023. p. 328–339. ISSN 2763-8952. Disponível em: <<https://doi.org/10.5753/sbcas.2023.229942>>. Citado 15 vezes nas páginas 36, 37, 38, 50, 51, 52, 53, 58, 69, 70, 71, 72, 77, 78 e 80.

BERGSTRA, J.; BARDENET, R.; BENGIO, Y.; KÉGL, B. Algorithms for hyperparameter optimization. In: SHAW-TAYLOR, J.; ZEMEL, R.; BARTLETT, P.; PEREIRA, F.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2011. v. 24. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf>. Citado na página 34.

BROUWER, O.; TAGAWA, S.; ALBERSEN, M. et al. Eau guidelines. In: *EAU Annual Congress Paris 2024*. Paris, France: European Association of Urology, 2024. Citado 2 vezes nas páginas 21 e 22.

CHEN, H.; LI, C.; LI, X.; RAHAMAN, M. M.; HU, W.; LI, Y.; LIU, W.; SUN, C.; SUN, H.; HUANG, X.; GRZEGORZEK, M. Il-mcam: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. *Computers in Biology and Medicine*, v. 143, p. 105265, 2022. ISSN 0010-4825. Disponível em: <<https://doi.org/10.1016/j.compbmed.2022.105265>>. Citado na página 15.

CHEN, H.; LI, C.; LI, X.; RAHAMAN, M. M.; HU, W.; LI, Y.; LIU, W.; SUN, C.; SUN, H.; HUANG, X.; GRZEGORZEK, M. Il-mcam: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. *Computers in Biology and Medicine*, v. 143, p. 105265, 2022. ISSN 0010-4825. Disponível em: <<https://doi.org/10.1016/j.compbmed.2022.105265>>. Citado 4 vezes nas páginas 49, 52, 53 e 62.

- CHEN, H.; LI, C.; WANG, G.; LI, X.; Mamunur Rahaman, M.; SUN, H.; HU, W.; LI, Y.; LIU, W.; SUN, C.; AI, S.; GRZEGORZEK, M. Gashis-transformer: A multi-scale visual transformer approach for gastric histopathological image detection. *Pattern Recognition*, v. 130, p. 108827, 2022. ISSN 0031-3203. Disponível em: <<https://doi.org/10.1016/j.patcog.2022.108827>>. Citado na página 15.
- CHEN, H.; LI, Y.; ZHU, H.; ZHANG, Y. Learn basic skills and reuse: Modularized adaptive neural architecture search (manas). In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. New York, NY, USA: Association for Computing Machinery, 2022. (CIKM '22), p. 169–179. ISBN 9781450392365. Disponível em: <<https://doi.org/10.1145/3511808.3557385>>. Citado na página 16.
- CHITTY-VENKATA, K. T.; EMANI, M.; VISHWANATH, V.; SOMANI, A. K. Neural architecture search for transformers: A survey. *IEEE Access*, v. 10, p. 108374–108412, 2022. Citado na página 33.
- COELHO, R. W. P.; PINHO, J. D.; MORENO, J. S.; GARBIS, D. V. e. O.; NASCIMENTO, A. M. T. do; LARGES, J. S.; CALIXTO, J. R. R.; RAMALHO, L. N. Z.; SILVA, A. A. M. da; NOGUEIRA, L. R.; FEITOZA, L. de M.; SILVA, G. E. B. Penile cancer in maranhão, northeast brazil: the highest incidence globally? *BMC Urology*, v. 18, n. 1, p. 50, May 2018. ISSN 1471-2490. Disponível em: <<https://doi.org/10.1186/s12894-018-0365-0>>. Citado na página 14.
- DABLAIN, D. A.; BELLINGER, C.; KRAWCZYK, B.; CHAWLA, N. V. Efficient augmentation for imbalanced deep learning. In: *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. Anaheim, California, USA: IEEE Institute of Electrical and Electronics Engineers, 2023. p. 1433–1446. Citado na página 36.
- DANO, H.; ALTINAY, S.; ARNOULD, L.; BLETARD, N.; COLPAERT, C.; DEDEURWAERDERE, F.; DESSAUVAGIE, B.; DUWEL, V.; FLORIS, G.; FOX, S.; GEROSA, C.; JAFFER, S.; KURPERSHOEK, E.; LACROIX-TRIKI, M.; LAKA, A.; LAMBEIN, K.; MACGROGAN, G. M.; MARCHIÓ, C.; MARTINEZ, D. M.; NOFECH-MOZES, S.; PEETERS, D.; RAVARINO, A.; REISENBICHLER, E.; RESETKOVA, E.; SANATI, S.; SCHELFHOUT, A.-M.; SCHELFHOUT, V.; SHAABAN, A. M.; SINKE, R.; STANCIU-POP, C. M.; STOBBE, C.; van Deurzen, C. H.; Van de Vijver, K.; Van Rompuy, A.-S.; VERSCHUERE, S.; VINCENT-SALOMON, A.; WEN, H.; BOUZIN, C.; GALANT, C.; Van Bockstal, M. R. Interobserver variability in upfront dichotomous histopathological assessment of ductal carcinoma in situ of the breast: the decision study. *Modern Pathology*, v. 33, n. 3, p. 354–366, 2020. ISSN 0893-3952. Disponível em: <<https://doi.org/10.1038/s41379-019-0367-9>>. Citado na página 23.
- DAWSON, H. Digital pathology – rising to the challenge. *Frontiers in Medicine*, v. 9, 2022. ISSN 2296-858X. Disponível em: <<https://doi.org/10.3389/fmed.2022.888896>>. Citado na página 24.
- DEWI, A. K.; PURWANTO, B.; WIDJIATI. Introduction of histopathology. In: KARA, A.; GELEN, V.; KARA, H. (Ed.). *Molecular Histopathology and Cytopathology*. Rijeka: IntechOpen, 2023. cap. 1. Disponível em: <<https://doi.org/10.5772/intechopen.110225>>. Citado 2 vezes nas páginas 22 e 23.
- DING, R.; ZHOU, X.; TAN, D.; SU, Y.; JIANG, C.; YU, G.; ZHENG, C. A deep multi-branch attention model for histopathological breast cancer image classification.

Complex & Intelligent Systems, v. 10, n. 3, p. 4571–4587, Jun 2024. ISSN 2198-6053. Disponível em: <<https://doi.org/10.1007/s40747-024-01398-z>>. Citado 3 vezes nas páginas 49, 52 e 53.

ELSKEN, T.; METZEN, J. H.; HUTTER, F. Neural architecture search: A survey. *The Journal of Machine Learning Research*, JMLR. org, v. 20, n. 1, p. 1997–2017, 2019. Citado 4 vezes nas páginas 33, 35, 60 e 83.

GHANATE, A.; KOTHIWALE, S.; SINGH, S.; BERTRAND, D.; KRISHNA, C. M. Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario. *Journal of Biomedical Optics*, Society of Photo-Optical Instrumentation Engineers, v. 16, n. 2, p. 025003–025003, 2011. Citado na página 14.

GHANATE, A. D.; KOTHIWALE, S.; SINGH, S. P.; KRISHNA, C. M.; BERTRAND, D. Comparative evaluation of spectroscopic models using different multivariate statistical tools in a multicancer scenario. *Journal of Biomedical Optics*, SPIE, v. 16, n. 2, p. 025003, 2011. Disponível em: <<https://doi.org/10.1117/1.3548303>>. Citado na página 23.

GIONA, S. The epidemiology of penile cancer. *Urologic Cancers [Internet]*, Exon Publications, 2022. Citado na página 21.

GUO, J.; HE, H.; HE, T.; LAUSEN, L.; LI, M.; LIN, H.; SHI, X.; WANG, C.; XIE, J.; ZHA, S.; ZHANG, A.; ZHANG, H.; ZHANG, Z.; ZHANG, Z.; ZHENG, S.; ZHU, Y. Gluoncv and gluonnlp: Deep learning in computer vision and natural language processing. *Journal of Machine Learning Research*, v. 21, n. 23, p. 1–7, 2020. Disponível em: <<http://jmlr.org/papers/v21/19-429.html>>. Citado na página 61.

GUO, M.-H.; XU, T.-X.; LIU, J.-J.; LIU, Z.-N.; JIANG, P.-T.; MU, T.-J.; ZHANG, S.-H.; MARTIN, R. R.; CHENG, M.-M.; HU, S.-M. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, v. 8, n. 3, p. 331–368, Sep 2022. Disponível em: <<https://doi.org/10.1007/s41095-022-0271-y>>. Citado na página 31.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE Institute of Electrical and Electronics Engineers, 2016. p. 770–778. Citado 2 vezes nas páginas 29 e 60.

HE, T.; ZHANG, Z.; ZHANG, H.; ZHANG, Z.; XIE, J.; LI, M. *Bag of Tricks for Image Classification with Convolutional Neural Networks*. 2018. Citado na página 61.

HERNANDEZ, B. Y.; BARNHOLTZ-SLOAN, J.; GERMAN, R. R.; GIULIANO, A.; GOODMAN, M. T.; KING, J. B.; NEGOITA, S.; VILLALON-GOMEZ, J. M. Burden of invasive squamous cell carcinoma of the penis in the united states, 1998–2003. *Cancer*, v. 113, n. S10, p. 2883–2891, 2008. Disponível em: <<https://doi.org/10.1002/cncr.23743>>. Citado na página 14.

HIREN, M.; F., A.-A. J.; AMIT, P.; JITENDRA, C.; KEYUR, M.; AND, V. A. Multi-channel local binary pattern guided convolutional neural network for breast cancer classification. *The Open Biomedical Engineering Journal*, v. 15, 2021. Disponível em: <<https://doi.org/10.2174/1874120702115010132>>. Citado na página 15.

- HOQUE, M. Z.; KESKINARKAUS, A.; NYBERG, P.; SEPPÄNEN, T. Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. *Information Fusion*, v. 102, p. 101997, 2024. ISSN 1566-2535. Disponível em: <<https://doi.org/10.1016/j.inffus.2023.101997>>. Citado na página 36.
- 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.; PLATANIOTIS, K. N. Computational pathology: A survey review and the way forward. *Journal of Pathology Informatics*, v. 15, p. 100357, 2024. ISSN 2153-3539. Disponível em: <<https://doi.org/10.1016/j.jpi.2023.100357>>. Citado 2 vezes nas páginas 24 e 25.
- HU, J.; SHEN, L.; ALBANIE, S.; SUN, G.; WU, E. *Squeeze-and-Excitation Networks*. 2019. Citado na página 32.
- INCA, I. N. d. C. *Câncer de pênis*. 2022. <<https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/penis>>. [Accessed 12-12-2024]. Citado na página 22.
- INCA, I. N. d. C. *SUS-ONCO ano VI n. 63 - Julho de 2022 Biópsia em bloco cirúrgico*. Esplanada dos Ministérios - Bloco G - Edifício Sede - CEP: 70058-900 - Brasília/DF, 2022. Citado na página 22.
- JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. *Electronic Markets*, v. 31, n. 3, p. 685–695, Sep 2021. Disponível em: <<https://doi.org/10.1007/s12525-021-00475-2>>. Citado na página 25.
- JIANG, Y.; SUI, X.; WU, X.; ZHENG, Y.; ZHANG, Y. A semi-supervised learning approach with consistency regularization for tumor histopathological images analysis. *Frontiers in Oncology*, v. 12, 2023. Citado na página 82.
- JIN, H.; SONG, Q.; HU, X. Auto-keras: An efficient neural architecture search system. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: Association for Computing Machinery, 2019. (KDD '19), p. 1946–1956. ISBN 9781450362016. Disponível em: <<https://doi.org/10.1145/3292500.3330648>>. Citado na página 34.
- JÚNIOR, A. A. L. T.; MELO, S. P. da C.; PINHO, J. D.; SOBRINHO, T. B. M.; ROCHA, T. M. S.; DUARTE, D. R. D.; BARBOSA, L. de O.; DUARTE, W. E.; BELFORT, M. R. de C.; DUARTE, K. G.; NETO, A. L. da S.; CALIXTO, J. de R. R.; PAIVA, L. C. P.; NASCIMENTO, F. S. M. S. do; JUNIOR, A. M. A.; KHAYAT, A. S.; CORRÊA, R. da G. C. F.; LAGES, J. S.; REIS, R. B. dos; ARAÚJO, W. S.; SILVA, G. E. B. A comprehensive analysis of penile cancer in the region with the highest worldwide incidence reveals new insights into the disease. *BMC Cancer*, v. 22, n. 1, p. 1063, Oct 2022. ISSN 1471-2407. Disponível em: <<https://doi.org/10.1186/s12885-022-10127-z>>. Citado 3 vezes nas páginas 14, 21 e 22.
- KANG, H.; LUO, D.; FENG, W.; ZENG, S.; QUAN, T.; HU, J.; LIU, X. Stainnet: A fast and robust stain normalization network. *Frontiers in Medicine*, v. 8, 2021. ISSN 2296-858X. Disponível em: <<https://doi.org/10.3389/fmed.2021.746307>>. Citado 3 vezes nas páginas 40, 58 e 71.

- KANG, J.-S.; KANG, J.; KIM, J.-J.; JEON, K.-W.; CHUNG, H.-J.; PARK, B.-H. Neural architecture search survey: A computer vision perspective. *Sensors*, v. 23, n. 3, 2023. ISSN 1424-8220. Disponível em: <<https://doi.org/10.3390/s23031713>>. Citado 3 vezes nas páginas 34, 60 e 63.
- KINGMA, D. P.; BA, J. *Adam: A Method for Stochastic Optimization*. 2017. Disponível em: <<https://arxiv.org/abs/1412.6980>>. Citado na página 65.
- KIRAN, N.; SAPNA, F.; KIRAN, F.; KUMAR, D.; RAJA, F.; SHIWLANI, S.; PALADINI, A.; SONAM, F.; BENDARI, A.; PERKASH, R. S. et al. Digital pathology: transforming diagnosis in the digital age. *Cureus*, Cureus Inc., v. 15, n. 9, 2023. Citado 3 vezes nas páginas 23, 24 e 25.
- KORESH, H. J. D. Impact of the preprocessing steps in deep learning-based image classifications. *National Academy Science Letters*, v. 47, n. 6, p. 645–647, Dec 2024. ISSN 2250-1754. Disponível em: <<https://doi.org/10.1007/s40009-023-01372-2>>. Citado na página 36.
- KORKES, F.; RODRIGUES, A. F. S.; BACCAGLINI, W.; CUNHA, F. T. S.; SLONGO, J.; SPIESS, P.; GLINA, S. Penile cancer trends and economic burden in the brazilian public health system. *einstein (São Paulo)*, Instituto Israelita de Ensino e Pesquisa Albert Einstein, v. 18, p. eAO5577, 2020. ISSN 1679-4508. Disponível em: <https://doi.org/10.31744/einstein_journal/2020AO5577>. Citado na página 14.
- KRICHEN, M. Convolutional neural networks: A survey. *Computers*, v. 12, n. 8, 2023. ISSN 2073-431X. Disponível em: <<https://doi.org/10.3390/computers12080151>>. Citado na página 27.
- KUMAR, T.; BRENNAN, R.; MILEO, A.; BENDECHACHE, M. Image data augmentation approaches: A comprehensive survey and future directions. *IEEE Access*, v. 12, p. 187536–187571, 2024. Citado 2 vezes nas páginas 41 e 43.
- LAAK, J. van der; LITJENS, G.; CIOMPI, F. Deep learning in histopathology: the path to the clinic. *Nature Medicine*, v. 27, n. 5, p. 775–784, May 2021. ISSN 1546-170X. Disponível em: <<https://doi.org/10.1038/s41591-021-01343-4>>. Citado 2 vezes nas páginas 14 e 15.
- LAUANDE, M. G. M.; JUNIOR, G. B.; ALMEIDA, J. D. S. de; SILVA, A. C.; COSTA, R. M. Gil da; TELES, A. M.; SILVA, L. L. da; BRITO, H. O.; VIDAL, F. C. B.; VALE, J. G. A. do; JUNIOR, J. R. D. R.; CUNHA, A. Building a densenet-based neural network with transformer and mbconv blocks for penile cancer classification. *Applied Sciences*, v. 14, n. 22, 2024. ISSN 2076-3417. Disponível em: <<https://doi.org/10.3390/app142210536>>. Citado 6 vezes nas páginas 50, 51, 53, 77, 78 e 82.
- LAUANDE., M. G. M.; TELES., A. M.; Lima da Silva., L.; MATOS., C. E. F.; Braz Júnior., G.; Cardoso de Paiva., A.; Sousa de Almeida., J. D.; OLIVEIRA., R. M. G. da C.; BRITO., H. O.; NASCIMENTO., A. P. S. A. dos; PESTANA., A. G.; PESTANA., A. P. S. A. dos; SANTOS., A. G.; LOPES., F. F. Classification of histopathological images of penile cancer using densenet and transfer learning. In: INSTICC. *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 4: VISAPP*. Online: SciTePress, 2022. p. 976–983. ISBN 978-989-758-555-5. ISSN 2184-4321. Citado 15 vezes nas páginas 36, 37, 50, 51, 52, 53, 55, 58, 69, 70, 71, 72, 77, 78 e 79.

- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436–444, May 2015. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/nature14539>>. Citado na página 27.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Citado na página 26.
- LI, H.; WU, P.; WANG, Z.; MAO, J.; ALSAADI, F. E.; ZENG, N. A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis. *Computers in Biology and Medicine*, v. 151, p. 106265, 2022. ISSN 0010-4825. Disponível em: <<https://doi.org/10.1016/j.compbimed.2022.106265>>. Citado na página 15.
- LI, H.; WU, P.; WANG, Z.; MAO, J.; ALSAADI, F. E.; ZENG, N. A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis. *Computers in Biology and Medicine*, v. 151, p. 106265, 2022. ISSN 0010-4825. Disponível em: <<https://doi.org/10.1016/j.compbimed.2022.106265>>. Citado 5 vezes nas páginas 27, 50, 52, 53 e 62.
- LIU, S.; ZHANG, H.; JIN, Y. A survey on computationally efficient neural architecture search. *Journal of Automation and Intelligence*, v. 1, n. 1, p. 100002, 2022. ISSN 2949-8554. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2949855422000028>>. Citado na página 83.
- LIU, Y.; SUN, Y.; XUE, B.; ZHANG, M.; YEN, G. G.; TAN, K. C. A survey on evolutionary neural architecture search. *IEEE Transactions on Neural Networks and Learning Systems*, v. 34, n. 2, p. 550–570, 2023. Citado na página 33.
- MA, D.; FANG, H.; WANG, N.; Lü, H.; MATTHEWS, J.; CHAO, Z. Transformer-optimized generation, detection, and tracking network for images with drainage pipeline defects. *Computer-Aided Civil and Infrastructure Engineering*, v. 38, p. 2109–2127, 2023. Citado na página 82.
- MACENKO, M.; NIETHAMMER, M.; MARRON, J. S.; BORLAND, D.; WOOSLEY, J. T.; GUAN, X.; SCHMITT, C.; THOMAS, N. E. A method for normalizing histology slides for quantitative analysis. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. Boston, Massachusetts, U.S.A.: IEEE Institute of Electrical and Electronics Engineers, 2009. p. 1107–1110. Citado na página 40.
- N, T. R.; GUPTA, R. A survey on machine learning approaches and its techniques:. In: *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. Bhopal, India: IEEE Institute of Electrical and Electronics Engineers, 2020. p. 1–6. Citado na página 25.
- NOH, S.-H. Performance comparison of cnn models using gradient flow analysis. *Informatics*, v. 8, n. 3, 2021. ISSN 2227-9709. Disponível em: <<https://doi.org/10.3390/informatics8030053>>. Citado na página 29.
- OLIVEIRA, R. T. V. d.; BERNARDES, G. d. O.; ALMEIDA, I. P.; RIBEIRO, M. d. F.; ALMEIDA, N. M. d.; MACHADO, L. C. d. S. Prevenção do câncer de pênis e a valorização da saúde do homem / prevention of penile cancer and enhancement of men's

health. *Brazilian Journal of Health Review*, v. 3, n. 2, p. 1527–1530, Mar. 2020. Disponível em: <<https://doi.org/10.34119/bjhrv3n2-016>>. Citado na página 14.

OYELADE, O. N.; EZUGWU, A. E. A bioinspired neural architecture search based convolutional neural network for breast cancer detection using histopathology images. *Scientific Reports*, v. 11, n. 1, p. 19940, Oct 2021. Disponível em: <<https://doi.org/10.1038/s41598-021-98978-7>>. Citado na página 16.

PARK, J.; WOO, S.; LEE, J.-Y.; KWEON, I. S. *BAM: Bottleneck Attention Module*. 2018. Disponível em: <<https://arxiv.org/abs/1807.06514>>. Citado 3 vezes nas páginas 15, 32 e 62.

REN, P.; XIAO, Y.; CHANG, X.; HUANG, P.-y.; LI, Z.; CHEN, X.; WANG, X. A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 54, n. 4, maio 2021. ISSN 0360-0300. Disponível em: <<https://doi.org/10.1145/3447582>>. Citado 3 vezes nas páginas 16, 33 e 34.

RICHTER, S.; RUETHER, J. D.; WOOD, L.; CANIL, C.; MORETTO, P.; VENNER, P.; GINGERICH, J.; EMMENEGGER, U.; EISEN, A.; ZALEWSKI, P.; JOSHUA, A.; MUKHERJEE, S. D.; HENG, D.; CZAYKOWSKI, P.; SOULIERES, D.; BLAIS, N.; RENDON, R.; FLESHNER, N.; CROOK, J. M.; SRIDHAR, S. S. Management of carcinoma of the penis: Consensus statement from the canadian association of genitourinary medical oncologists (cagmo). *Canadian Urological Association Journal*, v. 7, n. 11-12, p. e797–811, Dec. 2013. Disponível em: <<https://doi.org/10.5489/cuaj.1794>>. Citado na página 14.

RUIFROK, A.; JOHNSTON, D. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*, v. 23, 01 2001. Citado 2 vezes nas páginas 38 e 45.

RUSSELL, S. J.; NORVIG, P. *Artificial intelligence: a modern approach*. Hoboken, NJ, U.S.A.: Pearson, 2021. Citado na página 25.

SANZ-MOTILVA, V.; MARTORELL, A.; MANRIQUE-SILVA, E.; TERRADEZ-MAS, L.; REQUENA, C.; TRAVES, V.; SANMARTÍN, O.; RODRÍGUEZ-PERALTO, J.; NAGORE, E. Interobserver variability in the histopathological evaluation of melanoma: Analysis of 60 cases. *Actas Dermo-Sifiliográficas*, 2024. ISSN 0001-7310. Disponível em: <<https://doi.org/10.1016/j.ad.2024.05.023>>. Citado na página 23.

SHABAN, M. T.; BAUR, C.; NAVAB, N.; ALBARQOUNI, S. Staingan: Stain style transfer for digital histological images. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice, Italy: IEEE Institute of Electrical and Electronics Engineers, 2019. p. 953–956. Citado 3 vezes nas páginas 39, 58 e 71.

SHARMA, N.; SHARMA, R.; JINDAL, N. Machine learning and deep learning applications-a vision. *Global Transitions Proceedings*, v. 2, n. 1, p. 24–28, 2021. ISSN 2666-285X. 1st International Conference on Advances in Information, Computing and Trends in Data Engineering (AICDE - 2020). Disponível em: <<https://doi.org/10.1016/j.gltp.2021.01.004>>. Citado na página 25.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data*, v. 6, n. 1, p. 60, Jul 2019. ISSN 2196-1115. Disponível em: <<https://doi.org/10.1186/s40537-019-0197-0>>. Citado na página 36.

SRINIDHI, C. L.; CIGA, O.; MARTEL, A. L. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, v. 67, p. 101813, 2021. ISSN 1361-8415. Disponível em: <<https://doi.org/10.1016/j.media.2020.101813>>. Citado na página 15.

TELLEZ, D.; LITJENS, G.; BÁNDI, P.; BULTEN, W.; BOKHORST, J.-M.; CIOMPI, F.; van der Laak, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*, v. 58, p. 101544, 2019. ISSN 1361-8415. Disponível em: <<https://doi.org/10.1016/j.media.2019.101544>>. Citado 3 vezes nas páginas 36, 58 e 69.

THOMAS, A.; NECCHI, A.; MUNEER, A.; TOBIAS-MACHADO, M.; TRAN, A. T. H.; ROMPUY, A.-S. V.; SPIESS, P. E.; ALBERSEN, M. Penile cancer. *Nature Reviews Disease Primers*, v. 7, n. 1, p. 11, Feb 2021. ISSN 2056-676X. Disponível em: <<https://doi.org/10.1038/s41572-021-00246-5>>. Citado na página 21.

TSAKU, N. Z.; KOSARAJU, S. C.; AQILA, T.; MASUM, M.; SONG, D. H.; MONDAL, A. M.; KOH, H. M.; KANG, M. Texture-based deep learning for effective histopathological cancer image classification. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. San Diego, CA, USA: IEEE Institute of Electrical and Electronics Engineers, 2019. p. 973–977. Citado na página 15.

TULI, S.; DASGUPTA, I.; GRANT, E.; GRIFFITHS, T. L. Are convolutional neural networks or transformers more like human vision? *arXiv preprint arXiv:2105.07197*, 2021. Citado na página 82.

VAHADANE, A.; PENG, T.; SETHI, A.; ALBARQOUNI, S.; WANG, L.; BAUST, M.; STEIGER, K.; SCHLITTER, A. M.; ESPOSITO, I.; NAVAB, N. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging*, v. 35, n. 8, p. 1962–1971, 2016. Citado 3 vezes nas páginas 39, 40 e 58.

VALE, J.; SILVA, I.; MATOS, C.; JÚNIOR, G. B.; LAUANDE, M. Redes densenet com mecanismos de atenção múltipla aplicadas à classificação automática de câncer peniano em imagens histopatológicas. In: *Anais do XXIV Simpósio Brasileiro de Computação Aplicada à Saúde*. Porto Alegre, RS, Brasil: SBC, 2024. p. 495–506. ISSN 2763-8952. Disponível em: <<https://doi.org/10.5753/sbcas.2024.2755>>. Citado 11 vezes nas páginas 39, 50, 51, 52, 53, 58, 69, 70, 71, 77 e 78.

VIEIRA, C. B.; TEIXEIRA-JÚNIOR, A.; FEITOZA, L.; PINHO, J.; CALIXTO, J.; NASCIMENTO, F. S. M. S. do; CAMPOS, M. A. G.; LAGES, J.; JUNIOR, A. M. A.; SOARES, F.; CUNHA, I.; SILVA, G. E. B. A cohort study among 402 patients with penile cancer in maranhão, northeast brazil with the highest worldwide incidence. *BMC Research Notes*, v. 13, n. 1, p. 442, Sep 2020. ISSN 1756-0500. Disponível em: <<https://doi.org/10.1186/s13104-020-05283-z>>. Citado na página 21.

WANG, N.; ZHANG, C.; WEI, X.; YAN, T.; ZHOU, W.; ZHANG, J.; KANG, H.; YUAN, Z.; CHEN, X. Harnessing the power of optical microscopy for visualization and analysis of

- histopathological images. *Biomed. Opt. Express*, Optica Publishing Group, v. 14, n. 10, p. 5451–5465, Oct 2023. Disponível em: <<https://doi.org/10.1364/BOE.501893>>. Citado na página 23.
- WILSON, M. L.; FLEMING, K. Global cancer care: The role of pathology. *American journal of clinical pathology*, v. 145 1, p. 6–7, 2016. Citado na página 23.
- WOO, S.; PARK, J.; LEE, J.-Y.; KWEON, I. S. Cbam: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer Cham, 2018. Citado na página 15.
- WOO, S.; PARK, J.; LEE, J.-Y.; KWEON, I. S. *CBAM: Convolutional Block Attention Module*. 2018. Citado 2 vezes nas páginas 32 e 62.
- XU, W.; XU, L.; LIU, N.; LÜ, N. Histological image diagnosis of breast cancer based on multi-attention convolution neural network. *Journal of Shanghai Jiaotong University (Science)*, Feb 2024. ISSN 1995-8188. Disponível em: <<https://doi.org/10.1007/s12204-024-2705-4>>. Citado 3 vezes nas páginas 50, 52 e 53.
- ZENG, Y.; SUN, H.; KATTO, J.; FAN, Y. Accelerating convolutional neural network inference based on a reconfigurable sliced systolic array. *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021. Citado na página 82.
- ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. *Dive into Deep Learning*. Cambridge, England: Cambridge University Press, 2023. <<https://D2L.ai>>. Citado 3 vezes nas páginas 26, 28 e 29.
- ZHANG, C.; BENGIO, S.; HARDT, M.; RECHT, B.; VINYALS, O. *Understanding deep learning requires rethinking generalization*. 2017. Citado na página 61.
- ZUIDERVELD, K. Viii.5. - contrast limited adaptive histogram equalization. In: HECKBERT, P. S. (Ed.). *Graphics Gems*. Academic Press, 1994. p. 474–485. ISBN 978-0-12-336156-1. Disponível em: <<https://doi.org/10.1016/B978-0-12-336156-1.50061-6>>. Citado na página 37.