



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
CURSO DE GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

Tainá da Cruz

**Desenvolvimento de um Protótipo Computacional para Apoio à Análise
Citopatológica: Uma Abordagem com Describe Anything Model para Descrição de
Regiões Celulares**

Florianópolis
2026

Tainá da Cruz

**Desenvolvimento de um Protótipo Computacional para Apoio à Análise
Citopatológica: Uma Abordagem com Describe Anything Model para Descrição de
Regiões Celulares**

Trabalho de Conclusão de Curso do Curso de
Graduação em Ciências da Computação do Centro
Tecnológico da Universidade Federal de Santa
Catarina para a obtenção do título de bacharel em
Ciências da Computação.
Orientador: Prof. Aldo von Wangenheim, Dr. rer.nat.

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Cruz, Tainá da

Desenvolvimento de um Protótipo Computacional para Apoio
à Análise Citopatológica : Uma Abordagem com Describe
Anything Model para Descrição de Regiões Celulares / Tainá
da Cruz ; orientador, Aldo von Wangenheim, 2026.

86 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Ciências da Computação, Florianópolis, 2026.

Inclui referências.

1. Ciências da Computação. 2. Imagens Citopatológicas. 3.
Classificação Anotacional. 4. Segmentação Interativa. 5.
Descrição Visual Localizada. I. Wangenheim, Aldo von. II.
Universidade Federal de Santa Catarina. Graduação em
Ciências da Computação. III. Título.

Tainá da Cruz

**Desenvolvimento de um Protótipo Computacional para Apoio à Análise
Citopatológica:** Uma Abordagem com Describe Anything Model para Descrição de
Regiões Celulares

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de
“bacharel em Ciências da Computação” e aprovado em sua forma final pelo Curso de
Graduação em Ciências da Computação.

Florianópolis, 10 de julho de 2026.

Profª. Lúcia Helena Martins Pacheco, Dra.
Coordenadora do Curso

Banca Examinadora:

Prof. Aldo von Wangenheim, Dr. rer.nat.
Orientador

Prof. Luis Otávio Santos, Dr.
Avaliador
Universidade Federal de Santa Catarina

Prof. Alexandre Sherlley Casimiro Onofre,
Dr.
Avaliador
Universidade Federal de Santa Catarina

“It’s dangerous to go alone! Take this.”
(The Legend Of Zelda, 1986)

RESUMO

O presente trabalho desenvolve e avalia tecnicamente um protótipo computacional para apoio à anotação e à análise visual de imagens médicas derivadas de exames citológicos, com ênfase na integração do *Describe Anything Model* (DAM) para descrição visual localizada e classificação anotacional auxiliar. A pesquisa toma como motivação a importância da detecção precoce de alterações celulares, a complexidade do processo manual de análise citopatológica, a limitação de bases de dados que associem regiões celulares anotadas a descrições textuais especializadas e a necessidade de ferramentas capazes de organizar, processar e documentar regiões de interesse em ambiente computacional. Nesse contexto, o trabalho integra recursos de recebimento e processamento de imagens, suporte ao fluxo de conversão de arquivos *Carl Zeiss Image* (CZI) para *Digital Imaging and Communications in Medicine* (DICOM) aplicado a *Whole Slide Imaging* (WSI), segmentação interativa orientada por pontos e uma abordagem híbrida de análise regional baseada em regiões segmentadas. A implementação adota backend em FastAPI, interface web em React com Vite e organização modular entre camadas de domínio, serviços, infraestrutura e rotas de *Application Programming Interface* (API). A avaliação assume caráter técnico, exploratório e descritivo, com verificação funcional dos módulos implementados e experimentos sobre o *dataset* UFSC Multi-Cyto. O protótipo permite registrar imagens, consultar metadados, extrair regiões ou ladrilhos, executar segmentação interativa com *Segment Anything Model 2* (SAM2), preparar regiões segmentadas, gerar descrições visuais localizadas com DAM e disponibilizar classificação anotacional auxiliar baseada em *embeddings* visuais. Na etapa experimental, a adaptação textual com *Low-Rank Adaptation* (LoRA) demonstra funcionamento exploratório, enquanto o uso do DAM congelado como extrator visual permite treinar classificadores para classes anotacionais. A configuração principal apresentou acurácia de 0,7245, macro-F1 de 0,6743 e weighted-F1 de 0,7242 no conjunto de validação, indicando viabilidade exploratória do uso de *embeddings* visuais do DAM como base para classificação anotacional auxiliar. O conjunto de resultados não caracteriza validação diagnóstica nem substituição de especialistas. O protótipo constitui prova de conceito para apoio computacional à análise regional, à documentação de regiões e à interoperabilidade em imagens citopatológicas.

Palavras-chave: Imagens Citopatológicas. Describe Anything Model. Segmentação Interativa. Descrição Visual Localizada. Classificação Anotacional.

LISTA DE FIGURAS

Figura 1 – Lâmina de microscópio do colo uterino usando esfregaço (Papanicolaou).	15
Figura 2 – Tecido cutâneo em exame histológico com diferentes visualizações. . . .	16
Figura 3 – Diferentes técnicas de coloração celular.	19
Figura 4 – Exemplo de imagem citopatológica da cavidade oral corada por Papanicolaou, pertencente à classe <code>pap_em_divisao</code>	20
Figura 5 – Exemplo de imagem citopatológica do colo do útero corada por Feulgen, pertencente à classe <code>feulgen_velhas</code>	20
Figura 6 – Exemplo de imagem citopatológica do colo do útero corada por AgNOR, pertencente à classe <code>agnor_non_viable_nucleus</code>	21
Figura 7 – Ampliação de imagem WSI.	23
Figura 8 – Segmentação Interativa com SAM2.	25
Figura 9 – Arquitetura do SAM2.	26
Figura 10 – Arquitetura do DAM.	27
Figura 11 – Interface Web do Protótipo.	34
Figura 12 – Inserção de uma imagem PNG no protótipo.	36
Figura 13 – Inserção de uma imagem CZI no protótipo.	37
Figura 14 – Inserção de pontos em uma imagem citopatológica da classe <code>pap_em_divisao</code>	39
Figura 15 – Máscara do SAM2 aplicada em região de imagem citopatológica da classe <code>feulgen_sujeira</code>	40
Figura 16 – Descrição localizada de região selecionada para a classe <code>feulgen_sujeira</code> .	43
Figura 17 – Pipeline integrado de análise regional.	45
Figura 18 – Componente visual para verificação do resultado JSON obtido.	46
Figura 19 – Análise de região anotada presente no conjunto de validação do <i>dataset</i> .	47
Figura 20 – Matriz de confusão normalizada do classificador com até 500 exemplos por classe.	56
Figura 21 – Matriz de confusão normalizada do classificador com até 1.000 exemplos por classe.	57
Figura 22 – Matriz de confusão normalizada do classificador com até 2.000 exemplos por classe.	58
Figura 23 – Matriz de confusão normalizada do classificador com aumento de dados para classes raras.	59
Figura 24 – Principais confusões do classificador com até 500 exemplos por classe. .	60
Figura 25 – Principais confusões do classificador com até 1.000 exemplos por classe.	60
Figura 26 – Principais confusões do classificador com até 2.000 exemplos por classe.	61
Figura 27 – Principais confusões do classificador com aumento de dados para classes raras.	61

LISTA DE QUADROS

Quadro 1 – Principais Pontos da Resolução da Telepatologia	17
Quadro 2 – Categorias do Sistema Bethesda para Citopatologia Cervical	22
Quadro 3 – Classes anotacionais consideradas nos experimentos com o UFSC Multi- Cyto	33
Quadro 4 – Organização lógica da arquitetura do protótipo	35
Quadro 5 – Endpoints do módulo de arquivos e interoperabilidade	38
Quadro 6 – Estrutura funcional do módulo de segmentação interativa	41
Quadro 7 – Estrutura funcional do módulo de descrição visual localizada	42
Quadro 8 – Frentes exploratórias relacionadas ao DAM	44
Quadro 9 – Fluxo do pipeline integrado de análise regional	46
Quadro 10 – Resumo do ambiente experimental	49
Quadro 11 – Validação funcional dos módulos da API	50
Quadro 12 – Manifestos experimentais utilizados	53

LISTA DE TABELAS

Tabela 1 – Resultados de validação dos classificadores treinados sobre embeddings visuais do DAM	55
Tabela 2 – Comparação entre validação e teste do checkpoint principal	55
Tabela 3 – Métricas por classe no checkpoint principal	62

LISTA DE ABREVIATURAS E SIGLAS

AgNOR	<i>Argyrophilic Nucleolar Organizer Regions</i>
API	<i>Application Programming Interface</i>
CFM	<i>Conselho Federal de Medicina</i>
CZI	<i>Carl Zeiss Image</i>
DAM	<i>Describe Anything Model</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DNA	<i>Ácido Desoxirribonucleico</i>
DP	<i>Deep Learning</i>
GPU	<i>Graphics Processing Unit</i>
HIV	<i>Vírus da Imunodeficiência Humana</i>
HPV	<i>Papilomavírus Humano</i>
IA	<i>Inteligência Artificial</i>
LoRA	<i>Low-Rank Adaptation</i>
ML	<i>Machine Learning</i>
MLLMs	<i>Multimodal Large Language Models</i>
MLP	<i>Multilayer Perceptron</i>
MRXS	<i>Formato de lâmina virtual MIRAX</i>
NDPI	<i>Formato de imagem de patologia digital NanoZoomer</i>
PEFT	<i>Parameter-Efficient Fine-Tuning</i>
RQE	<i>Registro de Qualificação de Especialista</i>
SAM2	<i>Segment Anything Model 2</i>
SCN	<i>Formato de lâmina digital da Leica</i>
SUS	<i>Sistema Único de Saúde</i>
SVS	<i>ScanScope Virtual Slide</i>
TBS	<i>Sistema Bethesda</i>
TIFF	<i>Tagged Image File Format</i>
Web	<i>World Wide Web</i>
WSI	<i>Whole Slide Imaging</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	13
1.1.1	Objetivo Geral	13
1.1.2	Objetivos Específicos	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	CONCEITOS BÁSICOS	15
2.1.1	Citologia	15
2.1.2	Histologia	16
2.1.3	Telepatologia	16
2.2	PROCESSO DIAGNÓSTICO EM CITOPATOLOGIA	17
2.2.1	Câncer Oral	17
2.2.2	Câncer Cervical	18
2.2.3	Preparação de Lâminas e Técnicas de Coloração	18
2.2.3.1	Papanicolaou	19
2.2.3.2	Feulgen	19
2.2.3.3	AgNOR	20
2.2.4	Análise das Lâminas Citopatológicas	21
2.2.4.1	Sistema Bethesda	21
2.2.4.2	Whole Slide Imaging	22
2.2.5	Formatos de Imagem	23
2.2.5.1	Formato Carl Zeiss Image	24
2.2.5.2	Digital Imaging and Communications in Medicine para Whole Slide Imaging	24
2.3	CONCEITOS TÉCNICOS DE INTELIGÊNCIA ARTIFICIAL	24
2.3.1	Inteligência Artificial em Imagens Citopatológicas	24
2.3.2	Segmentação Interativa	25
2.3.2.1	Segment Anything Model 2	25
2.3.3	Modelos Multimodais de Visão e Linguagem	26
2.3.3.1	Describe Anything Model	27
2.3.4	Adaptação de Modelos e Aprendizado por Transferência	28
2.3.4.1	Low-Rank Adaptation e Parameter-Efficient Fine-Tuning	28
2.3.4.2	Embeddings Visuais e Classificadores Multilayer Perceptron	28
2.4	TRABALHOS RELACIONADOS	28
2.4.1	<i>Collaborative Telepathology in a Statewide Telemedicine Environment</i>	29
2.4.2	<i>Systematic Literature Review on Multimodal Large Language Models applied to Medical Image Data</i>	29
3	DESENVOLVIMENTO	31
3.1	MÉTODO DE PESQUISA	31

3.2	CONJUNTO DE DADOS	32
3.3	ARQUITETURA DO SISTEMA	34
3.3.1	Módulo de Arquivos e Interoperabilidade	35
3.3.2	Módulo de Segmentação Interativa	38
3.3.3	Módulo de Descrição Visual Localizada	40
3.3.4	Adaptação Exploratória e Classificação Anotacional Auxiliar	42
3.3.5	Pipeline Integrado de Análise Regional	44
3.3.6	Interface Web	47
3.4	LIMITAÇÕES DE IMPLEMENTAÇÃO	48
4	RESULTADOS E EXPERIMENTOS	49
4.1	AMBIENTE EXPERIMENTAL	49
4.2	VALIDAÇÃO FUNCIONAL DA API	49
4.3	AVALIAÇÃO DOS MÓDULOS IMPLEMENTADOS	51
4.3.1	Arquivos, Tiles e Metadados	51
4.3.2	Segmentação Interativa	51
4.3.3	Descrição Localizada com DAM	51
4.3.4	Validação Funcional da Conversão CZI-DICOM	52
4.4	DATASET EXPERIMENTAL	52
4.5	EXPERIMENTOS COM DESCRIBE ANYTHING MODEL	53
4.5.1	Adaptação Textual com LoRA	54
4.5.2	DAM Congelado como Extrator Visual	54
4.6	EXPERIMENTOS COM CLASSES ANOTACIONAIS	62
4.6.1	Aumento de Dados para Classes Raras	62
4.7	ANÁLISE QUALITATIVA DOS RESULTADOS	63
4.8	DISCUSSÃO DAS LIMITAÇÕES	63
5	CONCLUSÃO	65
5.1	TRABALHOS FUTUROS	65
	REFERÊNCIAS	67
	APÊNDICE A – REPOSITÓRIO	72
	APÊNDICE B – ARTIGO EM FORMATO SBC	73

1 INTRODUÇÃO

O câncer, comumente referido como tumor maligno, consiste em um grupo de mais de 100 doenças caracterizadas pelo crescimento desordenado de células. Essas células proliferam de modo invasivo e descontrolado, formando tumores que podem se disseminar para outras partes do corpo, acometendo tecidos e órgãos adjacentes, processo conhecido como metástase (INCA, 2021). A capacidade de invasão e proliferação é o que distingue os tumores malignos, associados ao câncer, dos tumores benignos, que não apresentam esse comportamento (Boaretto et al., 2024).

A detecção precoce de um tumor é fundamental para a remissão dos sintomas e a evolução do estado de saúde do indivíduo acometido, podendo ser tratada com sucesso se diagnosticado precocemente. Os estágios iniciais de desenvolvimento do câncer podem ser identificados por alterações no núcleo das células afetadas; todavia, o processo manual dessa análise pode ser demorado e estar sujeito a variações nas percepções de diferentes profissionais (Victória Matias et al., 2021a).

Nesse contexto, a citopatologia é utilizada para a análise de células por meio do exame de amostras obtidas de tecidos ou fluidos corporais (Johns Hopkins Medicine, 2024). Para o câncer cervical, exames citopatológicos são empregados em estratégias de rastreamento e permitem a observação de alterações em células coletadas do colo uterino (NCBI, 2024). Para o câncer oral, a investigação de alterações na cavidade oral pode envolver exame clínico, biópsia e citologia esfoliativa, na qual células coletadas por raspagem ou escovação são examinadas microscopicamente para identificação de anormalidades (PDQ Adult Treatment Editorial Board, 2025).

A digitalização de lâminas e o uso de imagens citopatológicas possibilitam que regiões celulares sejam visualizadas e analisadas em ambiente computacional (Kumar; Gupta; Gupta, 2020). Entretanto, a manipulação dessas imagens pode envolver formatos específicos de microscopia, como CZI, e padrões de interoperabilidade em imagem médica, como DICOM aplicado a WSI (Carl Zeiss Microscopy, 2026; DICOM Standards Committee, 2026; Brundage et al., 2022). Esse cenário motiva a construção de ferramentas capazes de apoiar o recebimento, o processamento e a documentação de regiões de interesse em imagens citopatológicas.

Em trabalho relacionado, Evaristo et al. (2025) conduziram uma revisão sistemática sobre *Multimodal Large Language Models* (MLLMs) aplicados a dados de imagem médica, analisando 30 artigos publicados entre 2023 e março de 2025. A revisão identificou avanços em aplicações de apoio diagnóstico especializado e *grounding* visual em nível de pixel, mas também apontou desafios relacionados à escassez e qualidade de dados anotados, custos computacionais, confiabilidade dos modelos, alucinações, generalização e integração clínica. Esses resultados reforçam a necessidade de investigar, de forma computacional e exploratória, o uso de modelos multimodais em imagens médicas.

Diante disso, técnicas de *Inteligência Artificial* (IA) têm sido investigadas como apoio a tarefas de prognóstico, predição e seleção de tratamento em câncer (Zhang; Shi; Wang, 2023). Este trabalho investiga uma alternativa computacional baseada em descrição visual localizada e segmentação interativa orientada por pontos. Modelos como o DAM foram propostos para gerar descrições detalhadas de regiões específicas em imagens e vídeos (Lian et al., 2025). Para delimitar regiões de interesse, destacam-se modelos como o SAM2, um modelo fundacional de segmentação promptável aplicado a imagens e vídeos (Ravi et al., 2024).

Assim, o trabalho propõe o desenvolvimento e a avaliação de um protótipo de API para apoio à anotação e à exploração visual de imagens citopatológicas. O protótipo integra recebimento e processamento de imagens, suporte a arquivos CZI com conversão para DICOM, segmentação interativa orientada por pontos com SAM2, descrição visual localizada com DAM e uma interface *World Wide Web* (Web) para demonstração do fluxo. A proposta é apresentada como protótipo computacional e utiliza mecanismos de IA para apoio à avaliação visual de imagens citológicas. A avaliação considera testes funcionais, exemplos de entrada e saída, análise qualitativa e experimentos exploratórios com classes anotacionais do *dataset* UFSC Multi-Cyto (Santos et al., 2025).

1.1 OBJETIVOS

Nas seções abaixo estão descritos o objetivo geral e objetivos específicos do trabalho.

1.1.1 Objetivo Geral

Desenvolver e avaliar um protótipo computacional para apoio à análise de imagens médicas derivadas de exames citológicos, integrando processamento de imagens, segmentação interativa com SAM2, descrição visual localizada com DAM, classificação anotacional baseada em *embeddings* visuais e suporte exploratório à interoperabilidade CZI-DICOM.

1.1.2 Objetivos Específicos

- a) Implementar uma API para recebimento, registro e processamento de imagens citopatológicas;
- b) Implementar suporte a arquivos CZI com conversão para DICOM;
- c) Integrar segmentação interativa orientada por pontos, com retorno estruturado de máscara, polígono, caixa delimitadora e metadados, através do SAM2;
- d) Integrar descrição visual localizada de regiões segmentadas com DAM;
- e) Investigar o uso do DAM como extrator visual para apoio à classificação anotacional;

- f) Avaliar o protótipo por validação funcional, análise qualitativa e experimentos com o *dataset* UFSC Multi-Cyto.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta as bases teóricas que fundamentam o desenvolvimento desse trabalho, com ênfase na citologia e nos aspectos relacionados aos cânceres oral e cervical. São apresentados os métodos de análise citológica e a aplicação de modelos de IA neste contexto. Por fim, discutem-se os trabalhos correlatos.

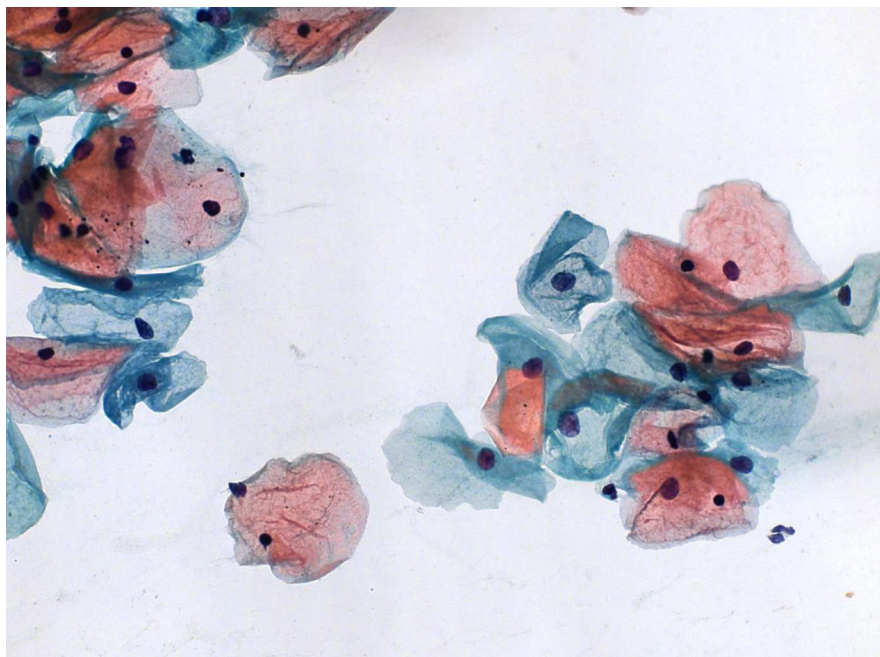
2.1 CONCEITOS BÁSICOS

Nessa seção são descritos conceitos básicos relacionados ao presente trabalho.

2.1.1 Citologia

A citologia é o ramo da biologia dedicado ao estudo da morfologia, estrutura, ciclo de vida e alterações das células. Tradicionalmente, esse campo está associado à análise microscópica das estruturas celulares, com ênfase no diagnóstico de anomalias e doenças malignas (NCBI, 2003). O termo citopatologia é utilizado de forma equivalente quando o foco está no processo diagnóstico, em que células obtidas por esfoliação, raspagem ou aspiração por agulha fina são processadas em lâminas e examinadas microscopicamente para identificar tumores, doenças pré-cancerosas e algumas doenças infecciosas (Stanford Pathology, 2024). A Figura 1 ilustra um exemplo típico das estruturas celulares analisadas nesse contexto.

Figura 1 – Lâmina de microscópio do colo uterino usando esfregaço (Papanicolaou).



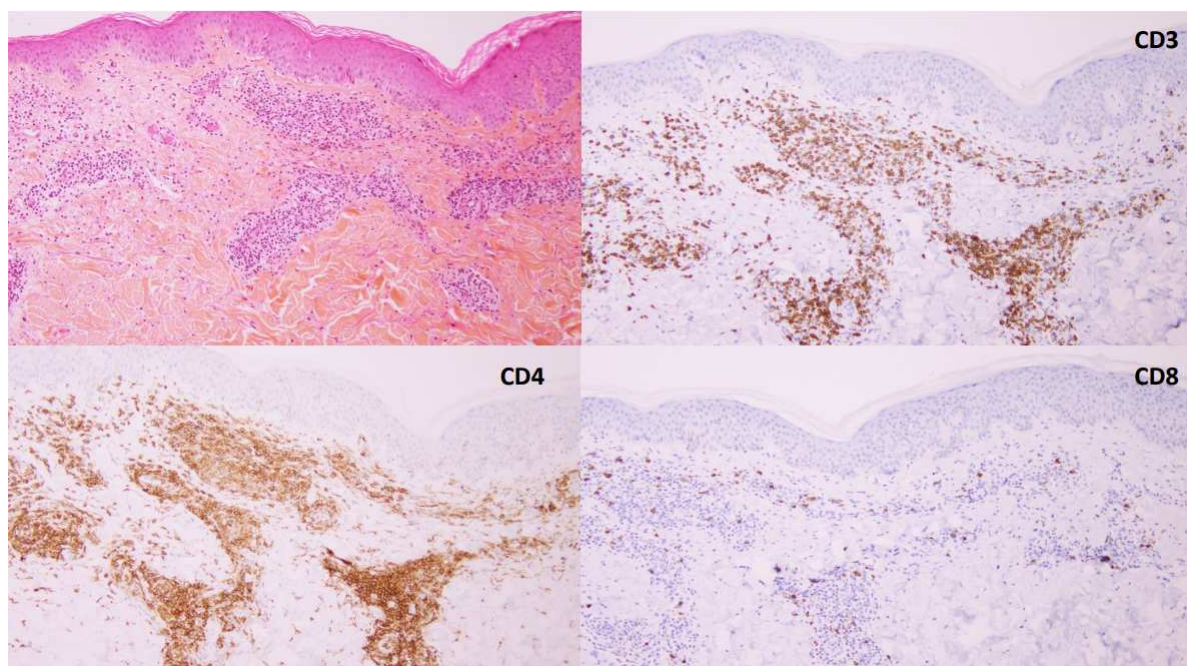
Fonte: Trevisan Rezende et al. (2020).

Por ser um procedimento diagnóstico de baixo custo e não invasivo, é utilizado para apoiar o diagnóstico de diversas doenças. No entanto, a realização manual do processo, mesmo sendo realizado por profissionais com extenso treinamento, torna-o demorado e repetitivo, apresentando critérios diagnósticos subjetivos que permanecem vulneráveis à interpretação humana (Victória Matias et al., 2021b).

2.1.2 Histologia

Enquanto a citologia foca na análise de células isoladas, a histologia, outro campo relacionado, dedica-se ao estudo de tecidos inteiros, analisando sua estrutura e organização como um todo (Johns Hopkins Medicine, 2024). A Figura 2 demonstra a abordagem histológica através de quatro visualizações do mesmo tecido cutâneo.

Figura 2 – Tecido cutâneo em exame histológico com diferentes visualizações.



Fonte: Masson (2022).

O escopo deste trabalho concentra-se na análise citológica, não abrangendo exames histológicos no desenvolvimento e testes da ferramenta. A seção 2.2.3 abordará os métodos de coloração empregados em citologia.

2.1.3 Telepatologia

A telepatologia se caracteriza pela obtenção e transmissão de imagens estáticas ou dinâmicas de amostras, tanto macroscópicas quanto microscópicas, com o objetivo de diagnóstico, consulta e educação médica (Sartori; Hilleshein; Ribeiro, 2014). Em 2019, o *Conselho Federal de Medicina* (CFM) regulamentou a prática da telepatologia no Brasil,

definindo-a como uma especialidade médica mediada por tecnologias para envio de dados e imagens. A resolução estabelece que todos os serviços devem obedecer a normas técnicas que garantam a integridade, veracidade e confidencialidade das informações transmitidas (FEHOESP, 2019; CFM..., 2019). Em vista disso, ferramentas para esse fim devem garantir o cumprimento de todos os critérios estabelecidos. O Quadro 1 apresenta os principais pontos da resolução.

Quadro 1 – Principais Pontos da Resolução da Telepatologia

A transmissão das imagens só pode acontecer se houver um médico em cada uma das pontas.
As imagens devem obedecer a critérios mínimos em relação à qualidade das imagens e à segurança no envio das informações.
O envio das informações deve garantir a integridade, veracidade, confidencialidade, privacidade e sigilo das informações.
A transmissão das imagens das lâminas é acompanhada dos dados pessoais e clínicos do paciente, que deve autorizar o processo.
As lâminas virtuais devem ser guardadas por, no mínimo, três meses. A imagem digital da lâmina física deve ser arquivada por cinco anos, que é o mesmo período de armazenamento da lâmina física.
A responsabilidade da transmissão deve ser assumida por um médico que tenha o <i>Registro de Qualificação de Especialista</i> (RQE) em patologia.

Fonte: CFM... (2019)

Embora os critérios do Quadro 1 sejam relevantes para sistemas de telepatologia em uso clínico, esse trabalho não busca conformidade regulatória integral. Em caráter acadêmico e experimental, esses critérios são utilizados como referência para contextualizar a importância de integridade, confidencialidade, rastreabilidade e confiabilidade no tratamento de imagens médicas digitais (FEHOESP, 2019; CFM..., 2019).

2.2 PROCESSO DIAGNÓSTICO EM CITOPATOLOGIA

Esta seção apresenta aspectos do processo diagnóstico em citopatologia relacionados aos cânceres cervical e oral.

2.2.1 Câncer Oral

O câncer oral compreende neoplasias malignas que acometem os lábios e diferentes regiões da cavidade oral, como língua, gengiva, mucosa jugal, assoalho da boca e palato. A maioria dessas neoplasias tem origem em células escamosas, sendo o tabagismo e o consumo elevado de álcool importantes fatores de risco (PDQ Adult Treatment Editorial Board,

2025). A investigação de alterações na cavidade oral pode envolver exame clínico, biópsia e citologia esfoliativa, na qual células são coletadas por raspagem ou escovação e examinadas microscopicamente para identificação de anormalidades (PDQ Adult Treatment Editorial Board, 2025). No contexto deste trabalho, essa fundamentação é relevante porque as imagens coradas por Papanicolaou do *dataset* UFSC Multi-Cyto são provenientes da cavidade oral (Santos et al., 2025).

2.2.2 Câncer Cervical

O câncer cervical, ou câncer do colo do útero, é o quarto tipo de câncer mais comum entre mulheres no mundo. A infecção persistente por tipos oncogênicos do *Papilomavírus Humano* (HPV) constitui seu principal fator etiológico, sendo os tipos 16 e 18 os mais frequentemente associados ao câncer cervical invasivo. A progressão da infecção para lesões intraepiteliais de alto grau e, posteriormente, para o câncer invasivo está associada à presença de outros fatores de risco, como tabagismo, infecção pelo *Vírus da Imunodeficiência Humana* (HIV), coinfeções genitais e uso de contraceptivos orais (NCBI, 2024).

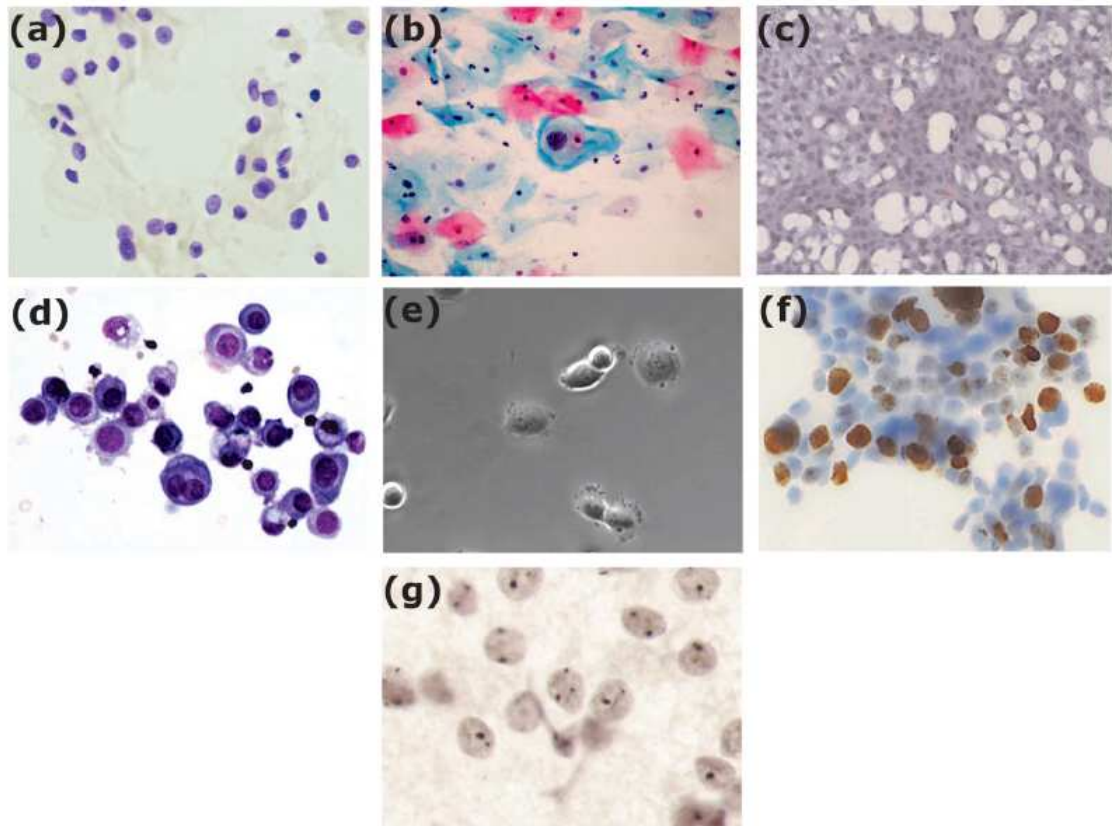
Nos Estados Unidos e em outros países desenvolvidos, os principais métodos de rastreamento e detecção precoce incluem os testes de HPV e o exame citopatológico de Papanicolaou, que permitem identificar, respectivamente, a exposição a tipos virais de risco e alterações citológicas anormais (NCBI, 2024). Nesse contexto, imagens citopatológicas multicoloração do colo do útero, associadas às categorias AgNOR e Feulgen do *dataset* UFSC Multi-Cyto, foram utilizadas como base para o desenvolvimento e a avaliação técnica do protótipo (Santos et al., 2025).

2.2.3 Preparação de Lâminas e Técnicas de Coloração

Na citologia, a análise do estado patológico celular ocorre por diversas técnicas, incluindo a quantificação de *Ácido Desoxirribonucleico* (DNA), a detecção de proteínas-chave na superfície celular, a identificação de patógenos e a avaliação da morfologia celular. Para cada método, aplicam-se técnicas específicas de coloração ou processamento que destacam diferentes estruturas celulares de interesse. As células são coletadas por meio de aspiração ou raspagem, e são fixadas em lâminas para preservação de sua estrutura (Victória Matias et al., 2021b). Na Figura 3, Victória Matias et al. (2021b) ilustram diferentes técnicas de coloração celular: Feulgen (a), Papanicolaou (b), H&E (c), Giemsa (d), microscopia de contraste de fase sem coloração (e), Hematoxilina+DAB (f) e AgNOR (g).

O *dataset* UFSC Multi-Cyto, utilizado nos experimentos deste trabalho, é disponibilizado como uma base de citopatologia multicoloração e contém classes anotacionais associadas a imagens de Papanicolaou, Feulgen e *Argyrophilic Nucleolar Organizer Regions* (AgNOR) (Santos et al., 2025). As subseções seguintes apresentam uma contextualização das três técnicas.

Figura 3 – Diferentes técnicas de coloração celular.



Fonte: Victória Matias et al. (2021b).

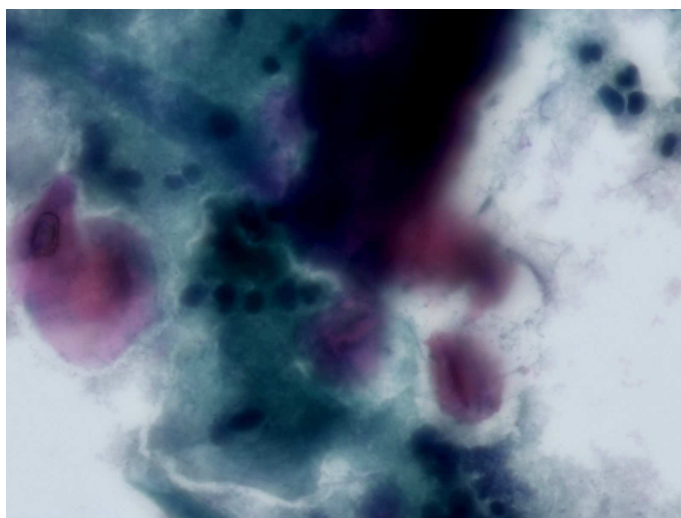
2.2.3.1 Papanicolaou

Embora a coloração de Papanicolaou seja geralmente aplicada à identificação de alterações citológicas do colo do útero, essa técnica também pode ser utilizada na análise citopatológica da cavidade oral. Na literatura de visão computacional, imagens de citologia oral coradas por Papanicolaou são utilizadas em tarefas de segmentação, detecção e classificação de núcleos celulares (Victória Matias et al., 2021a). No *dataset* UFSC Multi-Cyto, as imagens associadas às classes com prefixo **pap_** são provenientes da cavidade oral (Santos et al., 2025). A Figura 4 apresenta um exemplo da classe Papanicolaou com aspecto de divisão celular.

2.2.3.2 Feulgen

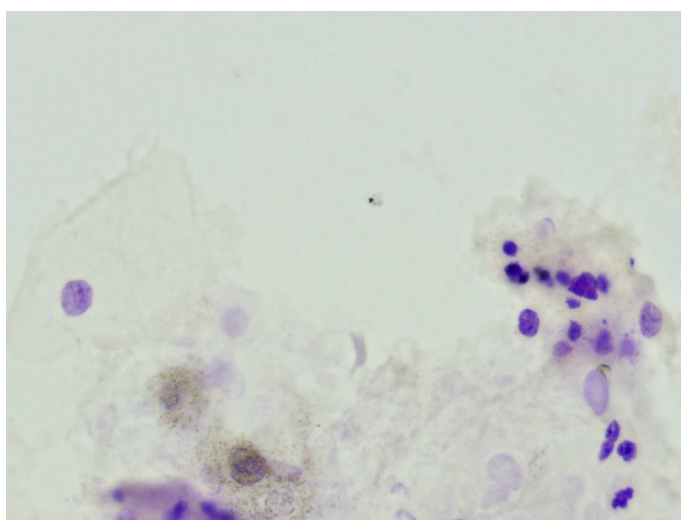
A coloração de Feulgen é apresentada na literatura entre as técnicas empregadas na análise citológica assistida por visão computacional (Victória Matias et al., 2021b). No *dataset* UFSC Multi-Cyto, as imagens obtidas com essa coloração são provenientes do colo do útero e suas categorias anotacionais são identificadas pelo prefixo **feulgen_** (Santos et al., 2025). A Figura 5 apresenta um exemplo da classe Feulgen para células ou regiões

Figura 4 – Exemplo de imagem citopatológica da cavidade oral corada por Papanicolaou, pertencente à classe `pap_em_divisao`.



Fonte: Santos et al. (2025).

Figura 5 – Exemplo de imagem citopatológica do colo do útero corada por Feulgen, pertencente à classe `feulgen_velhas`.



Fonte: Santos et al. (2025).

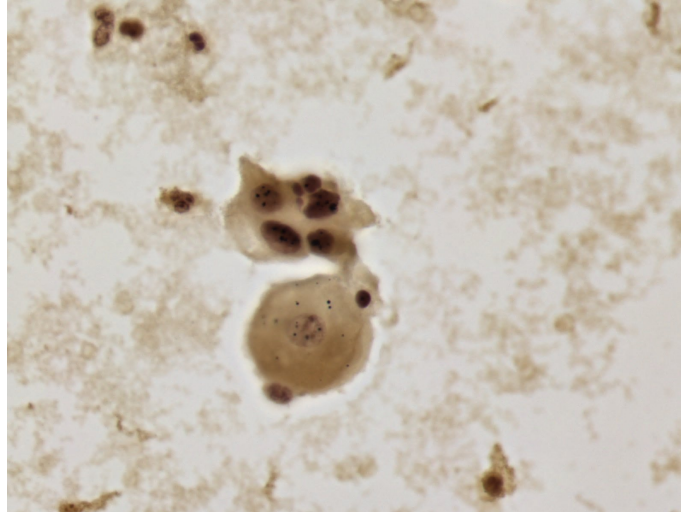
com aspecto envelhecido.

2.2.3.3 AgNOR

A coloração AgNOR também é descrita entre as técnicas utilizadas na análise citológica assistida por visão computacional (Victória Matias et al., 2021b). No *dataset* UFSC Multi-Cyto, as imagens obtidas com essa coloração são provenientes do colo do

útero e suas categorias anotacionais são identificadas pelo prefixo `agnor_` (Santos et al., 2025). A Figura 6 apresenta um exemplo da classe AgNOR para núcleo com aspecto não viável.

Figura 6 – Exemplo de imagem citopatológica do colo do útero corada por AgNOR, pertencente à classe `agnor_non_viable_nucleus`.



Fonte: Santos et al. (2025).

2.2.4 Análise das Lâminas Citopatológicas

Esta seção apresenta os conceitos, métodos e tecnologias empregados na análise citológica de lâminas microscópicas.

2.2.4.1 Sistema Bethesda

O *Sistema Bethesda* (TBS) é um método padronizado de nomenclatura para a elaboração de relatórios citopatológicos cervicais, estabelecendo critérios de classificação para resultados de exames de Papanicolaou, por meio de categorias que refletem a gravidade das lesões cervicais. A versão mais recente do sistema, publicada em 2014, inclui as seguintes categorias principais (Pangarkar, 2022):

Essas categorias promovem uma comunicação clara entre patologistas e médicos na avaliação dos exames citopatológicos (Pangarkar, 2022). No contexto do presente trabalho, a estrutura padronizada do TBS permite a compreensão de como as lâminas são classificadas manualmente, fornecendo uma referência para a análise e automatização do processo, a partir de uma perspectiva alinhada às práticas laboratoriais.

Quadro 2 – Categorias do Sistema Bethesda para Citopatologia Cervical

Categoria	Descrição
N/D	Não diagnóstico por amostras inadequadas para avaliação.
NILM	Negativo para lesão intraepitelial ou malignidade.
ASC	Células escamosas atípicas, incluindo alterações de significado indeterminado e aquelas em que não se pode excluir lesão intraepitelial escamosa de alto grau.
LSIL	Lesão intraepitelial escamosa de baixo grau.
HSIL	Lesão intraepitelial escamosa de alto grau.
Carcinoma de células escamosas	Neoplasia maligna de origem escamosa.
Células glandulares atípicas	Alterações em células endocervicais, endometriais ou glandulares, incluindo casos sem outra especificação e aqueles que favorecem processo neoplásico.
Adenocarcinoma endocervical <i>in situ</i>	Neoplasia glandular endocervical não invasiva.
Adenocarcinoma	Neoplasia glandular classificada como endocervical, endometrial, extrauterina ou sem outra especificação.
Outras neoplasias malignas	Outras neoplasias malignas identificadas no exame, que devem ser especificadas no relatório.

Fonte: Adaptado de Pangarkar (2022).

2.2.4.2 Whole Slide Imaging

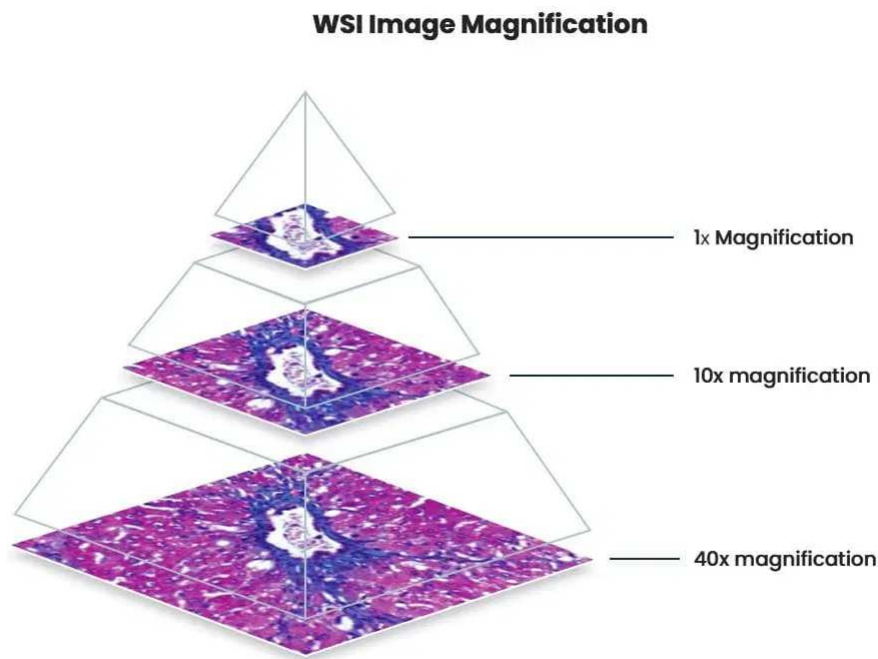
WSI é uma tecnologia que permite a digitalização completa de lâminas de tecido em alta resolução, transformando-as em imagens virtuais. Desde sua introdução em 1999, o WSI tem sido utilizado em diversas aplicações na patologia, incluindo diagnóstico, educação e pesquisa (Kumar; Gupta; Gupta, 2020). A lâmina virtual é gerada pela digitalização sequencial de diferentes áreas da amostra, que são integradas em uma matriz única. Este processo resulta em uma imagem virtual que permite ampliação significativa com preservação dos detalhes morfológicos (Kantseml, 2024).

A tecnologia envolve quatro processos principais: aquisição de imagem, armazenamento, processamento e visualização (Kumar; Gupta; Gupta, 2020). Essas características são descritas abaixo:

- **Aquisição de Imagem:** Realizada por *scanners* digitais que capturam sequencialmente imagens da lâmina, as quais são posteriormente montadas em uma única imagem virtual em alta resolução.

- **Armazenamento e Processamento:** As imagens digitais são armazenadas em sistemas que permitem fácil acesso e gerenciamento.
- **Visualização:** As imagens podem ser visualizadas em *softwares* especializados, permitindo anotações e exportações para outros formatos.

Figura 7 – Ampliação de imagem WSI.



Fonte: Kantseml (2024).

Kumar, Gupta e Gupta (2020) constataram que as vantagens do WSI incluem a eliminação da necessidade de armazenamento físico de lâminas, fácil acesso às imagens via internet e a possibilidade de integração com algoritmos de IA. No entanto, desafios como o alto custo dos equipamentos e a resistência à adoção dessa nova tecnologia ainda precisam ser superados (Kumar; Gupta; Gupta, 2020).

2.2.5 Formatos de Imagem

Em aplicações de patologia digital e análise computacional de lâminas, os formatos de imagem influenciam a forma como os dados visuais, metadados e estruturas de alta resolução são armazenados, acessados e compartilhados (Kumar; Gupta; Gupta, 2020). Nesse domínio, existem diferentes formatos associados a fabricantes de scanners e sistemas de lâminas digitais, como *ScanScope Virtual Slide* (SVS), da Aperio, *Formato de imagem de patologia digital NanoZoomer* (NDPI), da Hamamatsu, *Formato de lâmina digital da*

Leica (SCN), da Leica, *Formato de lâmina virtual MIRAX* (MRXS), da MIRAX, além de variações baseadas em *Tagged Image File Format* (TIFF) piramidal (Goode et al., 2013). No contexto deste trabalho, destacam-se o formato CZI, associado à aquisição de imagens de microscopia, e o padrão DICOM aplicado a WSI, relacionado à interoperabilidade de imagens médicas digitais (Carl Zeiss Microscopy, 2026; DICOM Standards Committee, 2026; Brundage et al., 2022).

2.2.5.1 Formato Carl Zeiss Image

CZI é um formato de arquivo desenvolvido pela ZEISS para atender requisitos de aquisição e armazenamento de imagens de microscopia. O formato permite armazenar imagens multidimensionais, como séries temporais, pilhas em profundidade, experimentos multiposição e lâminas virtuais, juntamente com metadados de aquisição, incluindo data, configurações do microscópio, dimensões, escala, contrastes, objetivos e parâmetros de câmera ou detector (Carl Zeiss Microscopy, 2026).

2.2.5.2 Digital Imaging and Communications in Medicine para Whole Slide Imaging

DICOM é um padrão mantido pela *National Electrical Manufacturers Association* para representação e comunicação de dados em imagem médica (DICOM Standards Committee, 2026). No contexto de patologia digital, sua aplicação a WSI busca favorecer a padronização e a interoperabilidade de imagens de lâmina inteira, especialmente quando há necessidade de converter formatos proprietários para estruturas compatíveis com sistemas de arquivamento, visualização e processamento médico (Brundage et al., 2022).

2.3 CONCEITOS TÉCNICOS DE INTELIGÊNCIA ARTIFICIAL

Esta seção apresenta conceitos técnicos de IA relacionados ao desenvolvimento do protótipo, com ênfase em análise computacional de imagens citopatológicas, segmentação interativa e descrição visual localizada.

2.3.1 Inteligência Artificial em Imagens Citopatológicas

Avanços recentes em IA têm alterado o processo diagnóstico patológico ao automatizar tarefas tradicionalmente manuais, como classificação celular e análise morfológica. Segundo Shafi e Parwani (2023), a integração dessas ferramentas às práticas patológicas tem otimizado o processamento e interpretação de grandes volumes de dados, permitindo diagnósticos mais confiáveis (Shafi; Parwani, 2023).

Os MLLMs, fundamentados em técnicas de *Machine Learning* (ML), têm aplicações significativas na medicina diagnóstica. Asif, Wenhui, ur-Rehman et al. (2024) evidencia a eficácia desses modelos na detecção precoce de doenças e no diagnóstico preciso de condições oncológicas e crônicas. No contexto da citopatologia, Giarnieri e Scardapane

(2023) demonstra que algoritmos de *Deep Learning* (DP) possibilitam a automatização de processos complexos, como detecção e segmentação celular, aprimorando a precisão diagnóstica (Giarnieri; Scardapane, 2023).

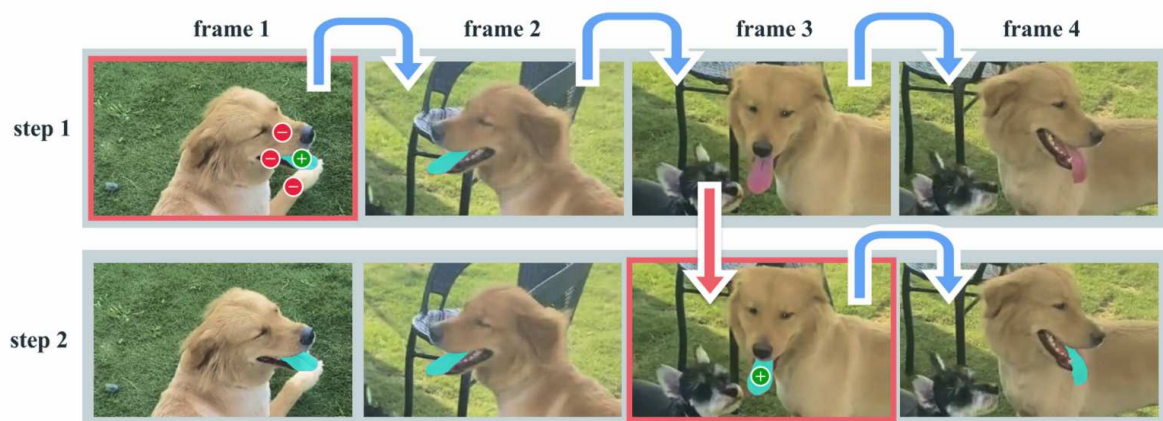
Em aplicações médicas, a literatura também aponta o uso de modelos multimodais para tarefas como apoio diagnóstico especializado, geração de relatórios e localização visual, embora ainda existam limitações relacionadas à disponibilidade de dados anotados, custo computacional, confiabilidade, alucinações e integração clínica (Evaristo et al., 2025). Por esse motivo, no presente trabalho, modelos de segmentação e descrição são utilizados em escopo exploratório e computacional, voltado à análise visual de regiões.

2.3.2 Segmentação Interativa

A segmentação é uma tarefa distinta da classificação na medida em que a classificação atribui uma categoria a uma imagem ou região, enquanto a segmentação busca delimitar espacialmente os elementos de interesse, frequentemente em nível de pixel (Giarnieri; Scardapane, 2023). Na segmentação interativa, a delimitação é orientada por entradas fornecidas pelo usuário, como cliques, pontos, caixas ou máscaras, que funcionam como instruções para indicar a região de interesse (Ravi et al., 2024). Essa abordagem é adequada a fluxos em que o usuário seleciona uma área específica da imagem para posterior análise ou anotação.

2.3.2.1 Segment Anything Model 2

Figura 8 – Segmentação Interativa com SAM2.



Fonte: Ravi et al. (2024).

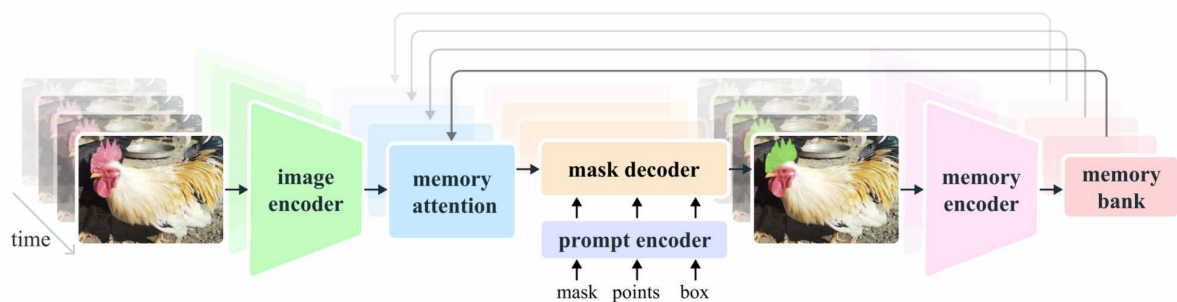
SAM2¹ é descrito por Ravi et al. (2024) como um modelo fundacional para seg-

¹ Página oficial do Meta Segment Anything Model 2. Disponível em: <https://ai.meta.com/research/sam2/>. Acesso em: 15 jun. 2026.

mentação visual promptável em imagens e vídeos, tratando a imagem como um caso particular de vídeo com um único quadro. Na formulação proposta pelos autores, a tarefa de *Promptable Visual Segmentation* permite fornecer pontos, caixas ou máscaras para definir a região de interesse e refiná-la iterativamente, inclusive em diferentes quadros no caso de vídeos, como apresenta a Figura 8.

Para isso, o modelo emprega uma arquitetura baseada em transformadores com *streaming memory*, capaz de armazenar informações sobre o objeto segmentado, os *prompts* já fornecidos e as previsões anteriores, o que favorece a propagação e a correção das máscaras ao longo do processo de segmentação (Ravi et al., 2024). Além do modelo, os autores apresentam uma *data engine* interativa utilizada na construção do conjunto SA-V, reforçando a proposta do SAM2 como uma extensão do SAM original para cenários de segmentação mais gerais e iterativos (Ravi et al., 2024). No presente trabalho, o SAM2 fundamenta a etapa de segmentação interativa a partir de pontos indicados pelo usuário, produzindo uma máscara da região selecionada.

Figura 9 – Arquitetura do SAM2.



Fonte: Ravi et al. (2024).

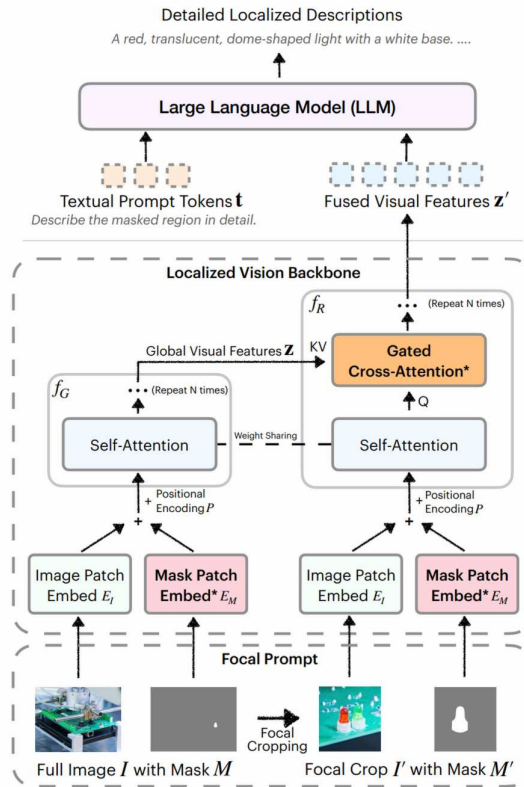
2.3.3 Modelos Multimodais de Visão e Linguagem

MLLMs são modelos que articulam informação visual e textual para executar tarefas que envolvem interpretação de imagens e geração de respostas em linguagem natural. Em imagem médica, revisões recentes indicam aplicações em apoio especializado, sistemas interativos, geração de relatórios e localização visual, mas também destacam limitações metodológicas e técnicas que impedem assumir desempenho clínico sem avaliação específica (Evaristo et al., 2025). Nesse trabalho, essa classe de modelos é considerada principalmente para a geração de descrições visuais localizadas de regiões previamente selecionadas.

2.3.3.1 Describe Anything Model

DAM² é um modelo proposto para *detailed localized captioning*, isto é, para a geração de descrições textuais detalhadas de regiões específicas em imagens e vídeos (Lian et al., 2025). Segundo Lian et al. (2025), a tarefa de descrição localizada permanece desafiadora porque modelos de visão e linguagem frequentemente operam sobre representações globais da imagem, o que pode reduzir detalhes finos da região de interesse, enquanto o recorte isolado da região pode remover informações contextuais relevantes.

Figura 10 – Arquitetura do DAM.



Fonte: Lian et al. (2025).

Para lidar com essa limitação, o DAM combina dois componentes principais: o *focal prompt*, que representa a região especificada com maior densidade de informação visual, e uma arquitetura visual localizada, que incorpora a máscara da região de forma espacialmente alinhada e integra o contexto global da imagem (Lian et al., 2025). A Figura 10 apresenta a arquitetura do modelo.

Na formulação do DAM, a região de interesse é representada por uma máscara binária associada à imagem, permitindo que a descrição gerada seja orientada à área selecionada (Lian et al., 2025). No escopo do trabalho, o DAM fundamenta a etapa de

² Página oficial do projeto Describe Anything. Disponível em: <https://describe-anything.github.io/>. Acesso em: 14 jun. 2026.

descrição textual da região segmentada, utilizando como entrada uma região preparada a partir da imagem e da máscara produzida no *pipeline* de análise regional, com restrição da saída a características visuais observáveis.

2.3.4 Adaptação de Modelos e Aprendizado por Transferência

O aprendizado por transferência consiste em reaproveitar conhecimento obtido em um domínio ou tarefa de origem para melhorar o aprendizado em um domínio ou tarefa relacionada, reduzindo a dependência de grandes volumes de dados no domínio-alvo (Zhuang et al., 2020). Em modelos de grande porte, essa ideia pode aparecer tanto no uso de modelos pré-treinados como extratores de características quanto em estratégias de adaptação parcial, nas quais apenas uma parte dos parâmetros é ajustada para uma nova tarefa.

2.3.4.1 Low-Rank Adaptation e Parameter-Efficient Fine-Tuning

Parameter-Efficient Fine-Tuning (PEFT) reúne métodos de ajuste de modelos que buscam adaptar modelos pré-treinados, treinando apenas um subconjunto reduzido de parâmetros, em vez de atualizar todos os pesos do modelo (Lialin et al., 2023). Entre essas técnicas, LoRA propõe congelar os pesos do modelo pré-treinado e inserir matrizes treináveis de baixo posto em camadas do modelo, reduzindo o número de parâmetros treináveis e o custo de memória em comparação com o ajuste completo (Hu et al., 2021). No contexto deste trabalho, essa fundamentação é relevante para distinguir adaptações leves e exploratórias de um treinamento completo de modelos fundacionais.

2.3.4.2 Embeddings Visuais e Classificadores Multilayer Perceptron

Outra estratégia de adaptação consiste em manter um modelo pré-treinado congelado e utilizá-lo como extrator de características. Nesse caso, a imagem ou região de interesse é transformada em uma representação vetorial, frequentemente chamada de *embedding*, que pode ser usada por um classificador supervisionado em uma tarefa específica (Zhuang et al., 2020). Um *Multilayer Perceptron* (MLP) é uma rede neural *feedforward* composta por camadas sucessivas, capaz de aprender mapeamentos entre vetores de entrada e saídas previstas durante o treinamento supervisionado (Goodfellow; Bengio; Courville, 2016). Essa abordagem permite avaliar se representações visuais extraídas de modelos pré-treinados contêm informação útil para distinguir classes anotacionais, sem modificar os pesos internos do modelo utilizado como extrator.

2.4 TRABALHOS RELACIONADOS

Nesta seção são apresentados os estudos relacionados ao escopo deste trabalho.

2.4.1 Collaborative Telepathology in a Statewide Telemedicine Environment

Andreóli Petrolini et al. (2019) desenvolveram um sistema de telepatologia colaborativa para o *Sistema Único de Saúde* (SUS) brasileiro, integrando funcionalidades de diagnóstico remoto e ambiente virtual para colaboração entre patologistas. O sistema permite o *upload* de espécimes digitalizados (WSI), oferece ferramentas de comunicação entre especialistas e possui um ambiente colaborativo que simula um microscópio virtual, possibilitando a análise conjunta e em tempo real das amostras.

Os autores implementaram um fluxo de trabalho que inclui solicitação de exames, digitalização e *upload* de espécimes, análise colaborativa e geração padronizada de relatórios seguindo o padrão DICOM *Structured Reporting*. A avaliação do sistema, realizada com 21 participantes entre estudantes de pós-graduação e professores de patologia do Brasil e México, utilizou o modelo ADEQUATE³ e demonstrou resultados positivos em aspectos como efetividade, capacidade de aprendizado e conforto na utilização.

O trabalho de Andreóli Petrolini et al. (2019) apresenta uma solução integrada para teleconsultas patológicas no contexto do sistema público de saúde, com foco em colaboração remota entre especialistas, análise conjunta de espécimes digitalizados e geração padronizada de relatórios. O presente trabalho se relaciona a essa proposta por também atuar no contexto de imagens patológicas digitais e interoperabilidade em imagem médica. Entretanto, diferencia-se por não propor um ambiente de telepatologia colaborativa nem um sistema de diagnóstico remoto, mas um protótipo de API independente para apoio à anotação e à análise visual localizada de imagens citopatológicas, integrando suporte a arquivos CZI, conversão para DICOM, segmentação interativa com SAM2 e descrição visual localizada com DAM.

2.4.2 Systematic Literature Review on Multimodal Large Language Models applied to Medical Image Data

Evaristo et al. (2025) apresentam uma revisão sistemática sobre MLLMs aplicados a dados de imagem médica. O estudo analisou publicações de 2023 a março de 2025, com o objetivo de identificar avanços na aplicação desses modelos, analisar mudanças em arquiteturas e estratégias de treinamento e discutir desafios técnicos de implementação. Os resultados da revisão indicam que, nesse período, as aplicações de MLLMs em imagem médica se diversificaram, abrangendo apoio diagnóstico especializado, sistemas interativos, geração de relatórios e tarefas de localização visual mais precisa. O estudo também observa a adoção de componentes específicos de domínio, técnicas de ajuste eficiente em parâmetros, *pipelines* de treinamento em múltiplas etapas e métodos de alinhamento.

³ Universidade Federal de Santa Catarina. *The adequate model: software quality evaluation model for telemedicine and telehealth systems*. Disponível em: <https://repositorio.ufsc.br/handle/123456789/186763>. Acesso em: 01 dez. 2024.

Como conclusão, os autores destacam que a área ainda enfrenta limitações relevantes, incluindo escassez de dados médicos anotados em larga escala, custo da anotação especializada, custos computacionais, dificuldades de interpretação, risco de alucinações, generalização limitada e desafios de integração clínica. Esses resultados se relacionam diretamente ao presente trabalho, pois reforçam a dificuldade de obter bases estruturadas com pares de imagem médica e descrição textual especializada. Nesse contexto, o protótipo desenvolvido investiga, de forma exploratória, o uso de segmentação interativa e descrição visual localizada como apoio computacional à anotação de regiões citopatológicas, sem apresentar as descrições geradas como laudos, diagnósticos ou validação clínica.

3 DESENVOLVIMENTO

Nas seções subsequentes, são descritos os itens específicos da proposta, incluindo método de pesquisa, dados utilizados, arquitetura do sistema, módulos implementados, adaptação exploratória, *pipeline* integrado, interface web de demonstração e limitações de implementação.

3.1 MÉTODO DE PESQUISA

O presente trabalho caracteriza-se como uma pesquisa aplicada, de caráter exploratório e descritivo, conduzida por meio do desenvolvimento e da avaliação técnica de um protótipo computacional (Prodanov; Freitas, 2013). A pesquisa é aplicada porque tem como finalidade construir uma solução computacional voltada a um problema prático de apoio à anotação e à exploração visual de imagens citopatológicas. Seu caráter exploratório decorre da integração de componentes recentes de segmentação e descrição visual localizada, enquanto o caráter descritivo está relacionado ao registro da arquitetura implementada, dos módulos desenvolvidos, das entradas e saídas do sistema e das limitações observadas.

A pesquisa foi conduzida por meio de atualizações incrementais do *software*, considerando um ambiente com capacidade de processamento de imagens médicas e execução de modelos computacionais de visão. O *backend* do protótipo foi desenvolvido em Python com o *framework* FastAPI, enquanto a interface web de demonstração foi implementada em TypeScript com React e Vite. Além disso, foram utilizadas bibliotecas e ferramentas específicas para armazenamento local, manipulação de imagens, processamento de máscaras, leitura de arquivos CZI, conversão para DICOM, segmentação interativa com SAM2 e descrição visual localizada com DAM (Ravi et al., 2024; Lian et al., 2025).

O objetivo metodológico não foi validar clinicamente um sistema de diagnóstico, mas desenvolver e avaliar uma API para apoio à anotação de imagens citopatológicas. A avaliação do protótipo considera o funcionamento dos módulos implementados, a estrutura das respostas produzidas, exemplos de entrada e saída, análise qualitativa dos resultados e experimentos exploratórios com classes anotacionais do *dataset* UFSC Multi-Cyto (Santos et al., 2025).

O desenvolvimento foi organizado em etapas incrementais. Inicialmente, foi definida uma arquitetura independente para o protótipo, separando responsabilidades entre domínio, serviços, infraestrutura, rotas de API e interface web. Em seguida, foram implementados módulos para recebimento e registro de imagens, suporte a arquivos CZI, conversão para DICOM, segmentação interativa orientada por pontos, descrição visual localizada com DAM e integração do fluxo por meio de um *endpoint* de análise regional. A Figura 11 apresenta a interface Web do protótipo desenvolvido.

Por fim, foram conduzidos experimentos relacionados ao uso do DAM no contexto

do *dataset* UFSC Multi-Cyto. Esses experimentos envolveram adaptação textual leve com LoRA e uso do DAM congelado como extrator de *embeddings* visuais para treinamento de um classificador MLP. Essa etapa é tratada como avaliação computacional sobre classes anotacionais, sem caracterização como *fine-tuning* completo do DAM (Hu et al., 2021; Lialin et al., 2023; Zhuang et al., 2020).

3.2 CONJUNTO DE DADOS

Esta seção descreve os dados utilizados no desenvolvimento e na avaliação técnica do protótipo. No escopo do trabalho, os dados são divididos em dois grupos: arquivos empregados nos fluxos de conversão CZI para DICOM e regiões anotadas do *dataset* UFSC Multi-Cyto utilizadas nos experimentos exploratórios com DAM.

O conjunto de dados utilizado nos experimentos exploratórios foi o *UFSC Multi-Cyto: Multi-Stain Cytopathology Image Dataset*, disponibilizado no Mendeley Data (Santos et al., 2025). Esse conjunto contém 17.422 fragmentos de imagens citopatológicas com anotações por instância e classes anotacionais associadas às colorações Papanicolaou (cavidade oral), AgNOR (colo do útero) e Feulgen (colo do útero). Os dados são fornecidos nos formatos COCO e YOLO e estão divididos em conjuntos de treinamento (12.502), validação (2.448) e teste (2.472). Além disso, Santos et al. (2025) informam que a coleta de dados foi realizada de acordo com protocolos aprovados pelo Comitê de Ética em Pesquisa da Universidade Federal de Santa Catarina (Santos et al., 2025). Durante o desenvolvimento, foram identificadas 22 classes anotacionais no *dataset*. O Quadro 3 apresenta as classes identificadas nos experimentos, agrupadas conforme o prefixo anotacional utilizado no *dataset*.

Para os experimentos com DAM, as anotações do *dataset* foram usadas para gerar exemplos derivados de regiões. O processo implementado no protótipo recorta a região anotada da imagem original, aplica a máscara da anotação como canal alfa e registra os exemplos em um manifesto estruturado. Esses exemplos foram utilizados para uma adaptação textual leve com LoRA e extração de *embeddings* visuais do DAM para treinamento de um classificador MLP.

A escolha por explorar descrições localizadas e classes anotacionais se relaciona à lacuna observada na revisão sistemática desenvolvida no contexto do trabalho, que identificou limitações na disponibilidade de bases multimodais contendo imagens médicas acompanhadas de textos diagnósticos especializados (Evaristo et al., 2025). Assim, o uso do DAM investiga a geração de descrições visuais localizadas como apoio computacional, sem caracterizar diagnóstico clínico.

Para os testes de interoperabilidade, foi utilizado um arquivo CZI do conjunto de imagens associado ao trabalho de Ramirez Puebla, Mark Welch e Borisy (2024). Esse arquivo foi empregado para validação funcional de leitura, extração de *tile* e conversão para DICOM, não compondo os experimentos citopatológicos realizados com o *dataset*.

Quadro 3 – Classes anotacionais consideradas nos experimentos com o UFSC Multi-Cyto

Grupo	Classe anotacional	Descrição operacional utilizada no protótipo
Papanicolaou	pap_anormal	Região marcada como anormal no <i>dataset</i> , sem interpretação como diagnóstico clínico.
Papanicolaou	pap_saudavel	Região marcada como saudável ou normal no <i>dataset</i> .
Papanicolaou	pap_fora_de_foco	Região com nitidez insuficiente ou foco visual limitado.
Papanicolaou	pap_sanguinea	Região com presença visual de sangue ou componente hemático.
Papanicolaou	pap_em_divisao	Célula ou região com aspecto de divisão celular no padrão visual do <i>dataset</i> .
Papanicolaou	pap_nao_definida	Região sem classe visual definida no vocabulário anotacional adotado.
Papanicolaou	pap_reacional	Região com alterações reacionais ou inflamatórias no padrão do <i>dataset</i> , sem interpretação diagnóstica.
AgNOR	agnor_Nucleus	Núcleo isolado ou região nuclear principal.
AgNOR	agnor_Cluster	Agrupamento visual de estruturas nucleares ou regiões nucleares próximas.
AgNOR	agnor_Satellite	Pequena estrutura satélite próxima à região nuclear.
AgNOR	agnor_Nucleus_out_of_focus	Núcleo ou região nuclear com foco visual limitado.
AgNOR	agnor_Overlapped_Nuclei	Núcleos sobrepostos ou parcialmente fundidos.
AgNOR	agnor_non_viable_nucleus	Núcleo com aspecto não viável, degradado ou fortemente alterado no padrão do <i>dataset</i> .
AgNOR	agnor_Leukocyte_Nucleus	Núcleo compatível com leucócito conforme o padrão visual do <i>dataset</i> .
Feulgen	feulgen_alterada	Região com alteração visual relevante de morfologia ou coloração.
Feulgen	feulgen_intermediaria	Região com padrão intermediário entre classes visuais do <i>dataset</i> .
Feulgen	feulgen_velhas	Células ou regiões com aspecto envelhecido ou degenerado no padrão do <i>dataset</i> .
Feulgen	feulgen_sujeira	Sujeira, artefato ou detrito visual.
Feulgen	feulgen_sobreposicao	Sobreposição visual de células, núcleos ou estruturas.
Feulgen	feulgen_nao_identificado	Região que não se encaixa visualmente nas demais categorias definidas.
Feulgen	feulgen_neutrofilo	Estrutura compatível com neutrófilo no padrão visual do <i>dataset</i> .
Feulgen	feulgen_mancha	Mancha ou marca de coloração sem estrutura celular claramente definida.

Fonte: Elaborado pela autora com base em Santos et al. (2025).

3.3 ARQUITETURA DO SISTEMA

O protótipo foi organizado como uma aplicação composta por *backend* em Python, uma interface Web de demonstração em TypeScript e *scripts* de apoio para execução, configuração e experimentos. A API recebe requisições, registra arquivos, aciona os serviços de processamento e retorna respostas estruturadas. A interface Web possui funcionalidades para carregamento de imagens, seleção de pontos, visualização da imagem carregada e de *tiles* extraídos de arquivos CZI, execução do *pipeline* e acionamento do fluxo de conversão para DICOM. Os experimentos com DAM foram desenvolvidos e executados por *scripts* específicos.

Figura 11 – Interface Web do Protótipo.

Análise de Regiões Citológicas SAM2 + DAM

Protótipo para segmentação interativa de regiões em imagens citopatológicas.

1. Upload

No file chosen

Imagem e pontos

Clique na imagem para adicionar um ponto positivo. Use Shift, Ctrl ou Alt + clique para adicionar ponto negativo.

☒ máscara ☒ polígono ☒ bounding box ☒ pontos

Imagem: não carregada Máscara: idle

Envie uma imagem JPG ou PNG para iniciar a análise.

Pontos selecionados: nenhum ponto selecionado.

2. Ações

Fonte: Elaborado pela autora.

A organização interna do *backend* separa responsabilidades em camadas. Essa separação organiza as entidades do domínio, a orquestração dos casos de uso, os roteadores HTTP e os adaptadores concretos para armazenamento, leitura de imagens, conversão, segmentação e descrição. O Quadro 4 resume essa organização.

No *backend*, a aplicação FastAPI registra os roteadores de saúde, arquivos, *jobs*, segmentação, descrição e *pipeline* sob o prefixo `/api/v1`. As requisições são estruturadas

Quadro 4 – Organização lógica da arquitetura do protótipo

Camada	Responsabilidade no protótipo
api	Define os roteadores HTTP, os esquemas de entrada e saída e o prefixo versionado dos <i>endpoints</i> sob <i>/api/v1</i> .
domain	Define entidades, portas de comunicação entre camadas e erros de domínio usados pelos serviços.
services	Orquestra os casos de uso da aplicação, incluindo arquivos, <i>tiles</i> , conversão, segmentação, descrição, classificação anotacional auxiliar e <i>pipeline</i> integrado.
infrastructure	Implementa adaptadores concretos para armazenamento local, leitura de arquivos CZI, conversão para DICOM, execução de SAM2, comunicação com DAM e classificação anotacional auxiliar.
frontend	Implementa a <i>interface web</i> de demonstração para <i>upload</i> de imagens, seleção de pontos, visualização de <i>tiles</i> exportados, execução do <i>pipeline</i> , acionamento da conversão de CZI para DICOM e visualização dos resultados.
scripts	Reúne rotinas de configuração, execução da API, execução da interface, servidor DAM e experimentos exploratórios.

Fonte: Elaborado pela autora.

e encaminhadas aos serviços correspondentes. Os serviços utilizam portas para acessar armazenamento local, manipular arquivos, exportar *tiles*, executar segmentação, produzir descrições e registrar *jobs* de conversão. O armazenamento local é organizado em diretórios para *uploads*, *tiles* exportados, máscaras, *jobs* e saídas DICOM.

3.3.1 Módulo de Arquivos e Interoperabilidade

O módulo de arquivos e interoperabilidade implementa a entrada de dados do protótipo. Sua função é receber imagens citopatológicas, registrar os arquivos enviados, permitir o acesso a imagens comuns ou regiões extraídas de arquivos CZI e oferecer suporte à conversão para DICOM. Essa etapa se relaciona diretamente ao objetivo de construir uma API para processar imagens citopatológicas e lidar com formatos associados à microscopia digital e à interoperabilidade em imagem médica (Carl Zeiss Microscopy, 2026; DICOM Standards Committee, 2026; Brundage et al., 2022).

O recebimento de arquivos é realizado por meio de *upload* HTTP. Ao salvar um arquivo, o sistema gera um identificador interno, preserva o nome original, identifica o tipo de arquivo e registra essas informações em armazenamento local. Imagens comuns, como PNG e JPG, podem ser acessadas como uma única imagem completa, representada

pelo *tile* full.

Figura 12 – Inserção de uma imagem PNG no protótipo.

1. Upload

Choose File | 05_2019_12_...m25513.png | Enviar imagem

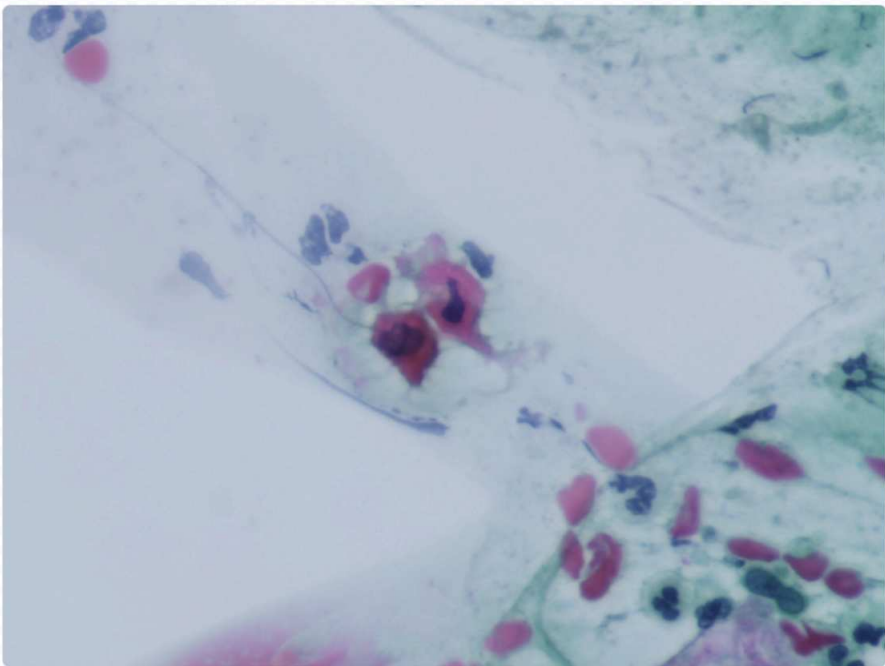
Arquivo enviado: 05_2019_12_13_10_45_0038_b0s0c0x52800-1600y109200-1200m25513.png

Imagem e pontos

Clique na imagem para adicionar um ponto positivo. Use Shift, Ctrl ou Alt + clique para adicionar ponto negativo.

☐ máscara ☒ polígono ☒ bounding box ☒ pontos

Imagem: carregada Máscara: idle



Pontos selecionados: nenhum ponto selecionado.

2. Ações

Analisar região | Remover último ponto | Limpar pontos

Resposta do upload

```
{  "file_id": "8d9ecd41-d3f8-4bba-8d12-fee3944dbc4",  "original_name": "05_2019_12_13_10_45_0038_b0s0c0x52800-1600y109200-1200m25513.png",  "kind": "image",  "content_type": "image/png"}
```

Resultado da análise

Nenhuma análise executada.

Fonte: Elaborado pela autora.

O serviço para a inserção de arquivos CZI, por sua vez, permite consultar metadados, listar regiões disponíveis conforme o tamanho padrão de *tile* e exportar uma região

selecionada para imagem PNG, que pode ser consumida pelos módulos de segmentação e descrição.

Figura 13 – Inserção de uma imagem CZI no protótipo.

The screenshot displays a web application interface with three main sections:

- 1. Upload:** Contains a file selection button labeled "Choose File" with the filename "Fig2A_raw.czi" next to it, and a blue button labeled "Enviar imagem". Below this, it shows "Arquivo enviado: Fig2A_raw.czi" and a dropdown menu for "Tile visualizado" currently set to "scene_0_tile_0".
- Conversão DICOM-WSI:** Features two blue buttons: "Converter CZI para DICOM-WSI" and "Atualizar job".
- Imagem e pontos:** Includes instructions: "Clique na imagem para adicionar um ponto positivo. Use Shift, Ctrl ou Alt + clique para adicionar ponto negativo." Below the instructions are four checkboxes: "máscara" (unchecked), "polígono" (checked), "bounding box" (checked), and "pontos" (checked). At the bottom of this section are two status indicators: "Imagem: carregada" and "Máscara: idle".

The central part of the interface is a large image viewer showing a dark, textured image. Below the viewer, a status message reads: "Pontos selecionados: nenhum ponto selecionado."

Fonte: Elaborado pela autora.

No fluxo de conversão para DICOM, a API recebe o identificador de um arquivo CZI, cria um registro de *job*, executa o adaptador de conversão e armazena a saída em diretório

local. O estado do *job* pode ser consultado posteriormente, e os arquivos gerados podem ser disponibilizados para download em formato compactado. A validação funcional do fluxo é apresentada no capítulo de resultados, indicando o arquivo CZI utilizado e a saída DICOM gerada. O Quadro 5 resume os *endpoints* deste módulo e suas responsabilidades.

Quadro 5 – Endpoints do módulo de arquivos e interoperabilidade

Endpoint	Responsabilidade
POST /api/v1/files	Recebe um arquivo por <i>upload</i> , registra o arquivo localmente e retorna identificador, nome original, tipo e <i>content type</i> .
GET /api/v1/files/{file_id}/metadata	Retorna metadados quando o arquivo é CZI; para imagens comuns, informa que metadados CZI não são aplicáveis.
GET /api/v1/files/{file_id}/tiles	Lista <i>tiles</i> disponíveis para arquivos CZI ou retorna o <i>tile full</i> para imagens comuns.
GET /api/v1/files/{file_id}/tiles/{tile_id}	Exporta e retorna a imagem correspondente ao <i>tile</i> solicitado.
POST /api/v1/files/{file_id}/convert/dicom	Inicia o fluxo de conversão de um arquivo CZI para DICOM e retorna o identificador do <i>job</i> .
GET /api/v1/jobs/{job_id}	Consulta o estado, a saída e eventuais erros associados a um <i>job</i> de conversão.
GET /api/v1/jobs/{job_id}/download	Disponibiliza os arquivos gerados pelo <i>job</i> em formato compactado.

Fonte: Elaborado pela autora.

Internamente, o módulo é composto por serviços e adaptadores. O serviço de arquivos delega o armazenamento ao repositório local, enquanto o serviço de CZI coordena a leitura de metadados, a listagem de *tiles* e a exportação de regiões. O adaptador de CZI utiliza bibliotecas para abrir o arquivo, interpretar metadados e ler uma região de interesse. O serviço de conversão utiliza um adaptador específico para transformar o arquivo de entrada em saída DICOM, registrando o resultado no armazenamento de *jobs*. Essa divisão mantém o fluxo de arquivos separado dos módulos de SAM2 e DAM, permitindo que a interoperabilidade seja avaliada de modo independente.

3.3.2 Módulo de Segmentação Interativa

O módulo de segmentação interativa é responsável por delimitar uma região de interesse a partir de pontos informados pelo usuário. Essa etapa utiliza o SAM2 como

modelo de segmentação promptável, conforme a fundamentação apresentada no Capítulo 2. No protótipo, o modelo é empregado para produzir uma máscara espacial da região indicada.

Figura 14 – Inserção de pontos em uma imagem citopatológica da classe `pap_em_divisao`.



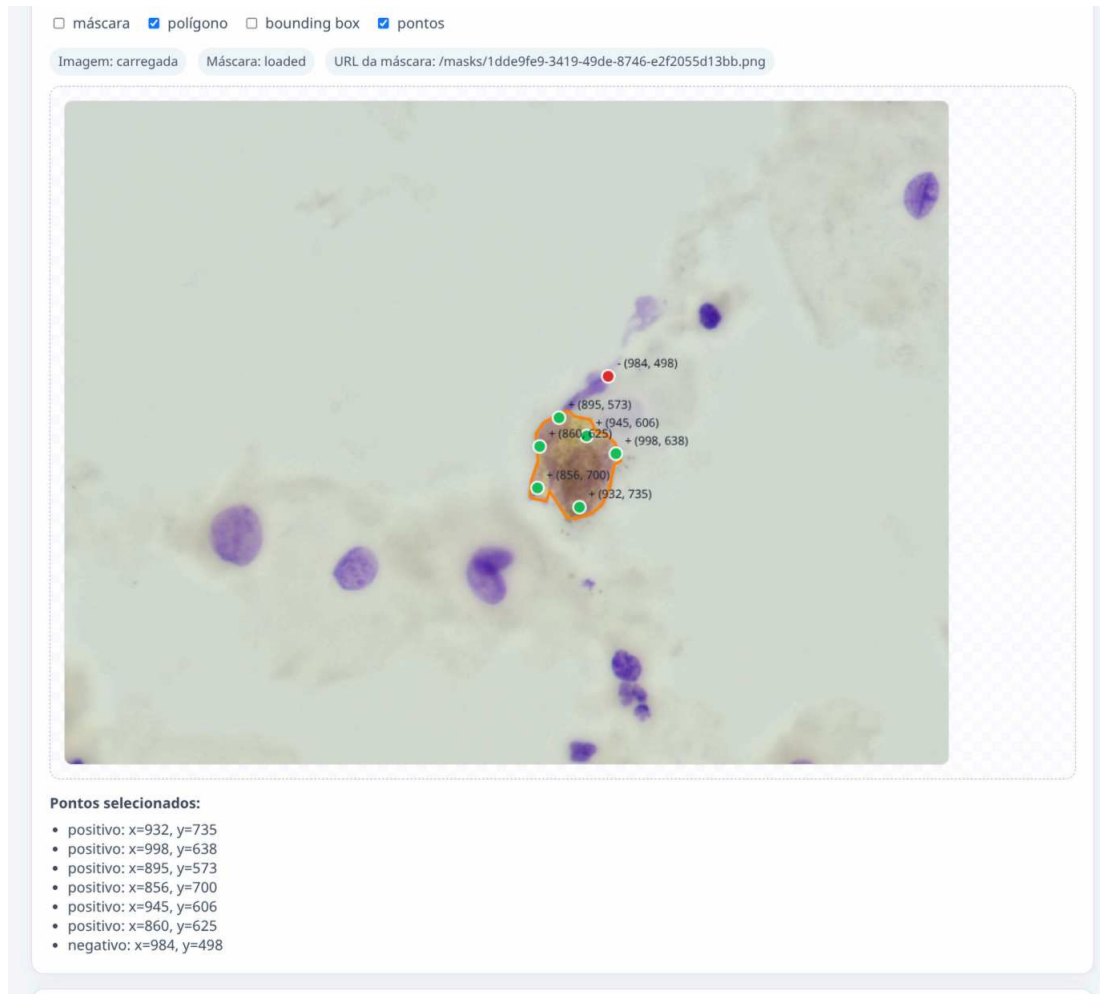
Fonte: Elaborado pela autora.

A requisição de segmentação recebe o identificador do arquivo, um identificador opcional de *tile* e uma lista de pontos. Cada ponto contém coordenadas *x* e *y* e um rótulo numérico, em que 1 representa ponto positivo e 0 representa ponto negativo. A partir dessas informações, o serviço exporta a imagem ou *tile* correspondente, encaminha a imagem ao adaptador de segmentação e retorna uma resposta estruturada.

O adaptador do SAM2 valida os pontos de entrada, carrega a imagem em formato RGB, executa a inferência do modelo e seleciona a máscara com maior pontuação entre as alternativas retornadas. Quando há múltiplos pontos, o adaptador utiliza uma etapa de refinamento: primeiro obtém uma hipótese inicial a partir de um ponto positivo e, em seguida, refina a segmentação usando todos os pontos informados. Após a inferência, a

máscara é convertida em caixa delimitadora, polígono aproximado e área em pixels. A máscara também é salva como imagem PNG e disponibilizada por meio de uma URL local. O Quadro 6 resume a estrutura do módulo.

Figura 15 – Máscara do SAM2 aplicada em região de imagem citopatológica da classe feulgen_sujeira.



Fonte: Elaborado pela autora.

Essa organização permite que a segmentação seja usada tanto pelo *endpoint* específico de SAM2 quanto pelo *pipeline* integrado de análise regional. No *pipeline*, a máscara produzida por este módulo passa a servir como entrada para a etapa de descrição visual localizada, descrita na seção seguinte.

3.3.3 Módulo de Descrição Visual Localizada

O módulo de descrição visual localizada é responsável por transformar a região segmentada em uma descrição textual objetiva. Essa etapa utiliza o DAM, modelo voltado à descrição localizada de regiões específicas em imagens e vídeos (Lian et al., 2025). No

Quadro 6 – Estrutura funcional do módulo de segmentação interativa

Etapa	Descrição
Entrada	Identificador do arquivo, <i>tile</i> opcional e lista de pontos com coordenadas e rótulo positivo ou negativo.
Processamento	Exportação da imagem ou <i>tile</i> , validação dos pontos, inferência com SAM2, seleção da máscara e conversão para representações estruturadas.
Saída	Identificador da máscara, caixa delimitadora, polígono, área em pixels, caminho ou URL da máscara e metadados técnicos da execução.
Endpoint	POST /api/v1/segmentation/sam2.

Fonte: Elaborado pela autora.

contexto do trabalho, o modelo é empregado para descrever características visuais observáveis da região marcada, como forma, contorno, coloração, contraste, textura, sobreposição e qualidade da visualização.

A inclusão do módulo se relaciona à lacuna identificada na revisão sistemática desenvolvida no contexto do trabalho, que apontou limitações na disponibilidade de bases de imagens médicas acompanhadas de textos diagnósticos especializados (Evaristo et al., 2025). Como o *dataset* UFSC Multi-Cyto contém classes anotacionais e máscaras, mas não descrições textuais especializadas por região, o DAM é explorado como alternativa para gerar descrições visuais localizadas a partir de regiões previamente delimitadas (Santos et al., 2025; Lian et al., 2025).

No protótipo, a descrição pode ser acionada pelo *endpoint* específico POST /api/v1/description/dam ou pelo *pipeline* integrado. A entrada contém o identificador do arquivo, um *tile* opcional, pontos de interesse e uma instrução textual. No fluxo específico de descrição, o sistema exporta a imagem ou *tile*, executa a segmentação interativa e encaminha a região resultante ao serviço de descrição. A Figura 16 demonstra esse resultado.

A preparação da entrada para o DAM combina a imagem RGB com a máscara produzida pela segmentação. O sistema calcula um recorte com contexto ao redor da caixa delimitadora, redimensiona a região quando necessário e gera uma imagem RGBA, na qual o canal alfa representa a máscara da região segmentada. Essa imagem é codificada como PNG em base64 e enviada ao servidor DAM por uma interface compatível com requisições de chat multimodal. A instrução textual enviada ao modelo restringe a resposta a características visuais observáveis em português brasileiro e explicita que a saída não deve conter diagnóstico clínico, afirmação de presença ou ausência de câncer, classificação de malignidade ou conclusão diagnóstica.

A resposta do módulo contém a descrição textual, o nome do modelo utilizado e

metadados técnicos da execução, incluindo parâmetros de geração e informações sobre o preparo da imagem. O Quadro 7 resume a estrutura funcional deste módulo.

Quadro 7 – Estrutura funcional do módulo de descrição visual localizada

Etapa	Descrição
Entrada	Imagem ou <i>tile</i> , pontos de interesse, máscara segmentada e instrução textual.
Preparação	Recorte contextual da região, combinação da imagem RGB com a máscara no canal alfa e codificação em PNG base64.
Processamento	Envio da região preparada ao servidor DAM com instrução textual.
Saída	Descrição visual localizada, nome do modelo e metadados técnicos da execução.
Endpoint	POST <code>/api/v1/description/dam</code> .

Fonte: Elaborado pela autora.

3.3.4 Adaptação Exploratória e Classificação Anotacional Auxiliar

Além do uso do DAM para descrição localizada, foram implementados experimentos exploratórios de adaptação ao domínio citopatológico. Essa etapa foi motivada pela escassez de bases com pares imagem e texto especializado, identificada na revisão sistemática, e a disponibilidade, no UFSC Multi-Cyto, de regiões anotadas com máscaras e classes associadas (Evaristo et al., 2025; Santos et al., 2025).

O primeiro passo consistiu na preparação de exemplos derivados do *dataset*. A partir das anotações em formato COCO, o projeto gera recortes das regiões anotadas, aplica a máscara da anotação como canal alfa e registra os exemplos em um manifesto JSONL. Cada exemplo contém a imagem RGBA, a classe anotacional, uma instrução textual e uma resposta esperada em formato controlado. Essa organização permite utilizar as regiões tanto em experimentos de adaptação textual quanto na extração de representações visuais.

A primeira rota experimental implementada foi a adaptação textual leve com LoRA. Essa técnica pertence ao conjunto de métodos de ajuste eficiente em parâmetros, nos quais o modelo pré-treinado permanece majoritariamente congelado e apenas um subconjunto reduzido de parâmetros é treinado (Hu et al., 2021; Lialin et al., 2023). No projeto, essa etapa foi aplicada ao componente linguístico do DAM, com o objetivo de avaliar se o modelo poderia aprender o formato de resposta esperado para as classes anotacionais. Essa rota não deve ser interpretada como *fine-tuning* completo do DAM, pois os componentes visuais do modelo não são ajustados de ponta a ponta.

Figura 16 – Descrição localizada de região selecionada para a classe `feulgen_sujeira`.

2. Ações

Analisar região
Remover último ponto
Limpar pontos

3. Descrição retornada

Com base no classificador por embeddings visuais do DAM, a região selecionada corresponde a categoria anotacional Feulgen para sujeira, artefato ou detrito visual (confiança: 0.8460). A região visível é uma mancha de coloração purple, com uma textura indistinta e um contorno irregular. A coloração é concentrada em uma área central, com uma graduação de intensidade de purple, sendo mais intensa na parte superior e mais fina na parte inferior. A mancha não apresenta nenhuma estrutura celular definida ou nenhuma outra característica visível. Observação: esta descrição é visual e localizada, gerada a partir da região segmentada, e não corresponde a diagnóstico clínico.

Classificação anotacional

Classe sugerida: `feulgen_sujeira` (85%)

```
[
  {
    "label": "feulgen_sujeira",
    "score": 0.8460327982902527,
    "definition": "categoria anotacional Feulgen para sujeira, artefato ou detrito visual."
  },
  {
    "label": "feulgen_mancha",
    "score": 0.13725939393043518,
    "definition": "categoria anotacional Feulgen para mancha ou marca de coloração sem estrutura celular claramente definida."
  },
  {
    "label": "pap_nao_definida",
    "score": 0.006610050331801176,
    "definition": "categoria anotacional Papanicolaou para região sem classe visual definida no vocabulário do dataset."
  }
]
```

Fonte: Elaborado pela autora.

A segunda rota experimental utilizou o DAM congelado como extrator de *embeddings* visuais. Nessa abordagem, as regiões anotadas são transformadas em vetores de características extraídos do *pipeline* visual do modelo, e esses vetores são utilizados para treinar uma cabeça classificadora MLP. Essa estratégia se fundamenta no aprendizado por transferência, pois reaproveita representações aprendidas por um modelo pré-treinado e treina um classificador supervisionado para uma tarefa específica (Zhuang et al., 2020).

Para integrar essa etapa ao protótipo, o servidor DAM foi estendido com um *endpoint* local de classificação anotacional auxiliar. A API consome esse *endpoint* por meio de um adaptador específico, que prepara a região com o mesmo mecanismo RGBA usado na descrição visual e retorna a classe mais provável, a confiança, uma lista *top-k* de candidatas e metadados da execução. Esse resultado é usado como contexto auxiliar para a descrição textual.

O Quadro 8 resume as frentes implementadas. Os resultados quantitativos, métricas, *checkpoints* e comparação entre configurações são apresentados no capítulo de resultados.

Quadro 8 – Frentes exploratórias relacionadas ao DAM

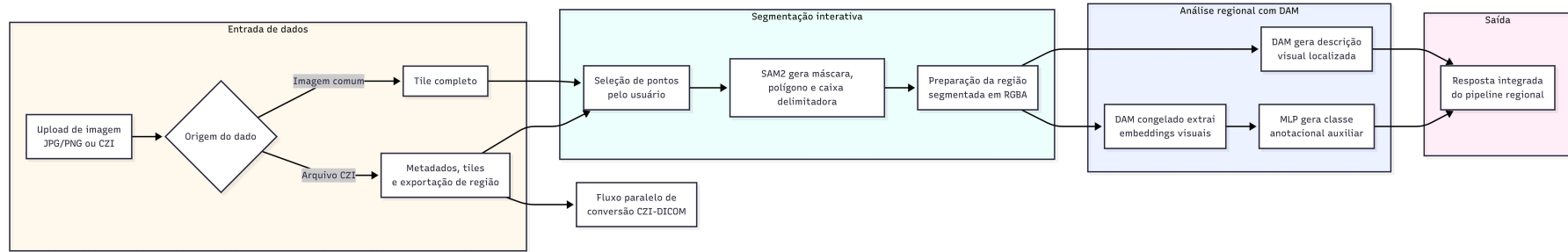
Frente	Papel no projeto
Preparação de exemplos	Geração de recortes RGBA a partir das anotações do UFSC Multi-Cyto e construção de manifesto estruturado.
LoRA textual	Adaptação leve do componente linguístico para testar aderência ao formato de resposta esperado.
Embeddings visuais	Uso do DAM congelado para extrair representações visuais das regiões anotadas.
Classificador MLP	Treinamento supervisionado de uma cabeça classificadora sobre <i>embeddings</i> visuais para classes anotacionais.
Classificação auxiliar na API	Disponibilização do resultado anotacional como contexto para o <i>pipeline</i> de descrição.

Fonte: Elaborado pela autora.

3.3.5 Pipeline Integrado de Análise Regional

O *pipeline* integrado de análise regional reúne os módulos de arquivos, segmentação, classificação anotacional auxiliar e descrição visual localizada em um único fluxo. Ele é exposto pelo *endpoint* `POST /api/v1/pipeline/analyze-region` e representa o caminho principal do protótipo para análise de uma região selecionada pelo usuário. A requisição do *pipeline* contém o identificador do arquivo, um *tile* opcional, a lista de pontos e uma instrução textual. A partir desses dados, o serviço exporta a imagem ou *tile* correspondente, executa a segmentação interativa, encaminha a região para a classificação anotacional auxiliar e, por fim, solicita a descrição visual localizada ao DAM.

Figura 17 – Pipeline integrado de análise regional.



Fonte: Elaborado pela autora.

A resposta do *pipeline* é um objeto JSON composto por três blocos: **segmentation**, **description** e **classification**. Essa estrutura pode ser verificada no componente visual de resultado da análise no protótipo.

Figura 18 – Componente visual para verificação do resultado JSON obtido.



Fonte: Elaborado pela autora.

O fluxo integrado é apresentado no Quadro 9. A conversão CZI para DICOM permanece fora do *pipeline* de análise regional, e a visualização direta de arquivos DICOM na interface configura-se como possibilidade de evolução futura do protótipo.

Quadro 9 – Fluxo do pipeline integrado de análise regional

Ordem	Etapa
1	Recuperação da imagem completa ou exportação do <i>tile</i> associado ao arquivo registrado.
2	Execução da segmentação interativa a partir dos pontos informados pelo usuário.
3	Preparação da região segmentada para consumo pelos módulos baseados no DAM.
4	Execução da classificação anotacional auxiliar sobre <i>embeddings</i> visuais.
5	Geração da descrição visual localizada.
6	Retorno de resposta JSON com segmentação, descrição e classificação auxiliar.

Fonte: Elaborado pela autora.

3.3.6 Interface Web

A interface web foi implementada como recurso de demonstração do fluxo principal do protótipo. Sua função é permitir que o usuário envie uma imagem, visualize a imagem ou um *tile* exportado no navegador, selecione pontos de interesse e acione o *pipeline* integrado de análise regional. A tela principal organiza o fluxo em etapas, como *upload* da imagem, seleção de pontos, execução da análise e exibição dos resultados. O componente de visualização permite adicionar pontos positivos por clique e pontos negativos por combinação da tecla **Ctrl** e clique. Após a execução do *pipeline*, a interface pode exibir a máscara sobreposta à imagem, o polígono, a caixa delimitadora, os pontos selecionados, a descrição retornada e a classificação anotacional auxiliar.

Figura 19 – Análise de região anotada presente no conjunto de validação do *dataset*.

Arquivo enviado: 02_val_agnor_Nucleus_10007.png

Imagem e pontos

Clique na imagem para adicionar um ponto positivo. Use Shift, Ctrl ou Alt + clique para adicionar ponto negativo.

☐ máscara ☒ polígono ☒ bounding box ☒ pontos

Imagem: carregada Máscara: loaded URL da máscara: /masks/6f65e721-ebfd-46f2-a410-318fee7e3f8.png



Pontos selecionados:

- positivo: x=49, y=46

2. Ações

Analisar região Remover último ponto Limpar pontos

3. Descrição retornada

Com base no classificador por embeddings visuais do DAM, a região selecionada corresponde a categoria anotacional AgNOR para núcleo isolado ou região nuclear principal (confiança: 0.5363). A imagem citopatológica mostra uma região de tecido com um centro de núcleo, que é o objeto central da imagem. A região é circundada por um tecido com uma cor marrom-branca, e há uma textura que sugere uma estrutura fibrótica. A região é isolada e não está em contato com outras estruturas visíveis na imagem. Observação: esta descrição é visual e localizada, gerada a partir da região segmentada, e não corresponde a diagnóstico clínico.

Classificação anotacional

Classe sugerida: agnor_Nucleus (54%)

Fonte: Elaborado pela autora.

Para arquivos CZI, a interface permite carregar o arquivo, listar *tiles* disponíveis, exportar um *tile* em formato PNG para visualização no componente de visualização e acionar

a conversão para DICOM. A visualização interativa, entretanto, permanece baseada na imagem raster exportada, e não no arquivo DICOM gerado. Essa delimitação mantém a interface compatível com o objetivo do protótipo e abre possibilidade para evolução posterior, considerando a visualização de imagens médicas de alta dimensão.

3.4 LIMITAÇÕES DE IMPLEMENTAÇÃO

As limitações de implementação foram tratadas como delimitações metodológicas do protótipo. A primeira limitação refere-se à execução de modelos fundacionais. O uso do SAM2 depende da configuração correta de *checkpoint*, arquivo de configuração e ambiente computacional compatível. De forma semelhante, o uso do DAM depende de um servidor local ou externo compatível com a interface utilizada pela API, além de recursos computacionais adequados para inferência e extração de *embeddings*.

O suporte a arquivos CZI e à conversão para DICOM depende de bibliotecas externas e da estrutura interna dos arquivos utilizados. Por esse motivo, os resultados deste módulo foram reportados separadamente no capítulo de resultados, indicando a entrada utilizada, a saída gerada e as limitações observadas.

Outra limitação está relacionada aos dados disponíveis para os experimentos exploratórios com DAM. O UFSC Multi-Cyto fornece imagens, máscaras e classes anotacionais, mas não descrições textuais especializadas para cada região (Santos et al., 2025). Assim, os experimentos de adaptação e classificação foram estruturados a partir das classes do *dataset*. Essa escolha permite avaliar o comportamento computacional do modelo somente sobre classes anotacionais.

A interface web também possui escopo restrito. Ela demonstra o fluxo principal com imagens comuns e com *tiles* extraídos de arquivos CZI, mas não implementa um visualizador completo para arquivos DICOM. Essa decisão reduz a complexidade da interface e mantém o foco do trabalho na arquitetura da API, na integração dos módulos principais e na geração da saída de interoperabilidade. Os resultados apresentados nesse trabalho devem ser lidos como avaliação técnica e exploratória de uma arquitetura de apoio à anotação, segmentação e descrição localizada de regiões em imagens citopatológicas.

4 RESULTADOS E EXPERIMENTOS

Este capítulo apresenta a avaliação técnica do protótipo desenvolvido. A avaliação foi organizada em três etapas: verificação funcional dos módulos da API, preparação dos dados utilizados nos experimentos exploratórios e análise das estratégias de adaptação e uso do DAM.

4.1 AMBIENTE EXPERIMENTAL

Os experimentos foram conduzidos em ambiente Linux. O Quadro 10 apresenta o resumo da configuração.

Quadro 10 – Resumo do ambiente experimental

Item	Configuração
Sistema operacional	Linux, arquitetura x86_64.
API	Python 3.10.14; FastAPI 0.136.3.
Experimentos DAM	Python 3.10.14; PyTorch 2.6.0+cu124; NumPy 1.26.4; Pillow 10.4.0.
Frontend	Node.js 20.19.2; npm 9.2.0; React 18.3.1; TypeScript 5.4.5; Vite 4.5.14.
Graphics Processing Unit (GPU)s disponíveis	2x NVIDIA GeForce RTX 3090, 24 GB; 1x NVIDIA GeForce RTX 5080, 16 GB.

Fonte: Elaborado pela autora.

As dependências do protótipo foram organizadas em requisitos da API, requisitos do servidor DAM e dependências da interface web. A API utiliza bibliotecas para serviços web, manipulação de imagens, processamento numérico, comunicação com o servidor DAM, leitura de arquivos CZI e conversão para DICOM. Os experimentos de classificação sobre *embeddings* visuais utilizaram *checkpoints* armazenados no diretório de experimentos do projeto.

4.2 VALIDAÇÃO FUNCIONAL DA API

A análise de validação funcional da API considerou os pontos de integração necessários para upload de arquivos, acesso a regiões, segmentação, descrição localizada, classificação anotacional auxiliar e acompanhamento de *jobs*. O *backend* possui um teste automatizado para o endpoint de saúde, definido em `backend/tests/test_health.py`. O Quadro 11 resume a situação dos módulos funcionais.

Quadro 11 – Validação funcional dos módulos da API

Módulo	Estado na avaliação	Observação
Saúde da API	Implementado	Há teste automatizado para <code>GET /api/v1/health</code> ; a ampliação da suíte de testes permanece como limitação da validação.
Arquivos e metadados	Implementado	O fluxo de upload, registro local, metadados e <i>tiles</i> está implementado e integrado aos demais módulos da API.
Segmentação interativa	Implementado	O endpoint de segmentação está integrado ao serviço de SAM2, retornando máscara, polígono, caixa delimitadora, área e metadados.
Descrição localizada	Implementado	O endpoint de descrição prepara a região segmentada, comunica-se com o servidor DAM e retorna descrição textual controlada.
Classificação anotacional auxiliar	Implementado	O servidor DAM foi estendido localmente para retornar classe sugerida, confiança e lista <i>top-k</i> , com base no classificador treinado sobre <i>embeddings</i> visuais.
Pipeline integrado	Implementado	O endpoint <code>POST /api/v1/pipeline/analyze-region</code> consolida segmentação, descrição e classificação auxiliar no <i>pipeline</i> .
Conversão de CZI para DICOM	Validado funcionalmente	O fluxo foi executado com arquivo CZI real, gerando saída DICOM em arquivo <code>.dcm</code> ; a validação permanece limitada a uma amostra.
Interface web	Implementada	A interface demonstra upload, seleção de pontos, execução do <i>pipeline</i> , visualização de <i>tile</i> exportado de CZI e acionamento da conversão de CZI para DICOM.

Fonte: Elaborado pela autora.

4.3 AVALIAÇÃO DOS MÓDULOS IMPLEMENTADOS

Os módulos implementados foram avaliados de acordo com sua responsabilidade no fluxo do protótipo. O módulo de arquivos fornece a entrada de dados e o registro local; o módulo de segmentação delimita a região de interesse; o módulo de descrição transforma a região segmentada em texto visual localizado; o módulo de classificação anotacional fornece uma hipótese computacional sobre classes do *dataset*; e o módulo de conversão de CZI para DICOM trata da interoperabilidade.

4.3.1 Arquivos, Tiles e Metadados

O fluxo de arquivos permite registrar imagens comuns, identificar o tipo de arquivo e recuperar a imagem completa como *tile full*. Para arquivos CZI, realiza extração de metadados, listagem de *tiles* e exportação de regiões como imagem PNG. O acesso a regiões de imagens de alta dimensão é tratado como etapa de entrada para os modelos, enquanto a conversão de CZI para DICOM permanece como módulo paralelo de interoperabilidade.

Na avaliação funcional, o fluxo de arquivos foi testado com imagens de citologia e com um arquivo CZI real. Para o arquivo CZI, o protótipo registrou o upload, listou regiões disponíveis e exportou um *tile* em formato PNG para uso na visualização interativa. Esse teste teve finalidade técnica e não substitui uma avaliação ampla de compatibilidade com diferentes arquivos de microscopia.

4.3.2 Segmentação Interativa

Para o módulo de segmentação interativa, a avaliação ocorreu pela integração com o fluxo de dados. A entrada recebe arquivo, *tile* opcional e pontos, enquanto a saída é estruturada com máscara, polígono, caixa delimitadora, área e metadados. A execução com SAM2 depende de *checkpoint* e arquivo de configuração compatíveis com o ambiente.

4.3.3 Descrição Localizada com DAM

O módulo de descrição localizada foi avaliado quanto à preparação da região segmentada e ao uso de *prompt* restritivo. A região é preparada em formato RGBA, mantendo a máscara no canal alfa, e enviada ao servidor DAM com instrução para gerar uma descrição visual em português brasileiro. A resposta foi delimitada para descrever atributos observáveis, como forma, contorno, coloração, textura, sobreposição, contraste e qualidade da região.

Os experimentos quantitativos apresentados nas seções seguintes não avaliam diretamente a qualidade linguística das descrições livres geradas pelo DAM. Esses experimentos avaliam, de forma separada, estratégias exploratórias de adaptação e classificação anotacional associadas ao uso do DAM no *dataset* UFSC Multi-Cyto.

4.3.4 Validação Funcional da Conversão CZI-DICOM

Para validar funcionalmente o módulo de interoperabilidade, foi utilizado um arquivo em formato CZI disponibilizado no conjunto de imagens associado ao trabalho de Ramirez Puebla, Mark Welch e Borisy (2024). Esse arquivo foi empregado para verificar leitura, extração de *tile* e conversão de formato, não sendo parte dos experimentos citopatológicos com o *dataset* UFSC Multi-Cyto.

O arquivo CZI foi registrado pela API e utilizado para testar a listagem e a exportação de *tiles*. Como a amostra possui estrutura multidimensional e canais em escala de cinza, a visualização no navegador exigiu a exportação de uma região para PNG, com normalização de contraste para exibição no componente de visualização. Esse procedimento permitiu utilizar o *tile* exportado no mesmo componente interativo empregado para imagens comuns.

Em seguida, foi executado o fluxo de conversão de CZI para DICOM por meio do endpoint de conversão do protótipo. A conversão foi concluída com sucesso e produziu um arquivo *.dcm*, disponibilizado como saída do *job*. Esse resultado indica funcionamento do módulo para a amostra testada, sem representar validação ampla para diferentes fabricantes, configurações de aquisição ou estruturas internas de arquivos CZI.

A visualização direta do arquivo DICOM gerado não foi incorporada ao componente de visualização da interface web. A interface atual visualiza imagens convencionais ou *tiles* exportados em PNG, utilizados para seleção de pontos e análise regional. Assim, a conversão para DICOM deve ser interpretada como funcionalidade de interoperabilidade e geração de saída, enquanto a visualização direta de arquivos DICOM permanece como limitação do protótipo.

4.4 DATASET EXPERIMENTAL

Os experimentos exploratórios com DAM utilizaram o *dataset* UFSC Multi-Cyto, disponibilizado no Mendeley Data (Santos et al., 2025). O conjunto contém imagens citopatológicas com anotações por instância e classes anotacionais associadas a diferentes colorações. No recorte utilizado neste projeto, foram consideradas 22 classes anotacionais, derivadas das anotações disponíveis na estrutura do *dataset*.

A preparação dos exemplos experimentais partiu das anotações em formato COCO. Para cada anotação selecionada, o processo recortou a região da imagem original, aplicou a máscara como canal alfa e salvou a região em formato PNG com quatro canais. Em seguida, os exemplos foram registrados em um manifesto JSONL, contendo caminho da imagem, classe anotacional e metadados da região. Esse procedimento permitiu utilizar as mesmas regiões tanto em experimentos de adaptação textual quanto na extração de *embeddings* visuais.

O Quadro 12 resume os principais manifestos utilizados nos experimentos quantita-

tivos. A validação foi mantida com limite de até 300 exemplos por classe; após remoções e filtros aplicados no *pipeline* de extração de características, os relatórios de avaliação registraram 4.875 exemplos válidos na validação. Após a seleção do *checkpoint* principal, também foi preparado um recorte do *split* de teste com limite de até 300 exemplos por classe, resultando em 4.931 regiões, das quais 4.925 permaneceram válidas após a extração dos *embeddings* visuais.

Quadro 12 – Manifestos experimentais utilizados

Manifesto	Treino	Validação	Finalidade
500/300 fixo	9.781 exemplos	4.879 exemplos	Comparar desempenho com limite menor de treino por classe.
1000/300 fixo	18.759 exemplos	4.879 exemplos	Configuração principal por equilíbrio entre acurácia e macro-F1.
2000/300 fixo	32.804 exemplos	4.879 exemplos	Avaliar efeito de aumentar o limite de treino para até 2.000 exemplos por classe.
1000/300 com aumento	24.186 exemplos combinados	Mantida a validação fixa	Avaliar aumento de dados em classes raras do treino.

Fonte: Elaborado pela autora.

A distribuição por classe permaneceu desbalanceada mesmo após a seleção balanceada por limite máximo, pois algumas classes possuíam poucos exemplos disponíveis. No manifesto 1000/300 **fixo**, as classes de *papanicolaou* em divisão, não definida e reacional possuíam, respectivamente, 44, 63 e 245 exemplos de treino. Na validação, essas mesmas classes possuíam 4, 9 e 85 exemplos. Essa característica influencia a interpretação das métricas e reforça a necessidade de analisar resultados por classe, além das médias globais.

4.5 EXPERIMENTOS COM DESCRIBE ANYTHING MODEL

Os experimentos com DAM foram conduzidos em duas rotas principais. A primeira consistiu em uma adaptação textual leve com LoRA, aplicada ao componente linguístico do modelo. A segunda utilizou o DAM congelado como extrator de *embeddings* visuais e treinou uma cabeça classificadora MLP para prever as 22 classes anotacionais do UFSC Multi-Cyto. Essa segunda rota produziu os principais resultados quantitativos deste capítulo.

4.5.1 Adaptação Textual com LoRA

A adaptação textual com LoRA teve caráter exploratório. O objetivo foi verificar se o componente linguístico do DAM poderia aprender um formato de resposta mais aderente às classes anotacionais do *dataset*, sem ajustar de ponta a ponta os componentes visuais do modelo. Essa abordagem se enquadra em ajuste eficiente em parâmetros e não deve ser descrita como *fine-tuning* completo do DAM (Hu et al., 2021; Lialin et al., 2023).

Os testes dessa rota demonstraram funcionamento técnico do *pipeline* de preparação de dados, carregamento do modelo, aplicação dos adaptadores e comparação qualitativa entre respostas. Entretanto, a adaptação textual não resolveu de forma consistente a discriminação visual entre classes. Esse resultado é coerente com a limitação metodológica da própria estratégia, pois ao ajustar o componente textual, a capacidade de distinguir visualmente classes próximas ou raras continua dependente das representações visuais congeladas.

Por esse motivo, a adaptação textual com LoRA foi mantida como experimento exploratório. A rota quantitativa principal passou a ser a classificação sobre *embeddings* visuais extraídos do DAM.

4.5.2 DAM Congelado como Extrator Visual

Na segunda rota, o DAM foi mantido congelado e utilizado para extrair *embeddings* visuais das regiões anotadas. Em seguida, uma cabeça classificadora MLP foi treinada de forma supervisionada para prever as classes anotacionais. A configuração principal utilizou *pooling mean*, normalização 12 e camada oculta de dimensão 1024. O treinamento adotou *dropout* de 0,2, taxa de aprendizado de 0,0005, decaimento de peso de 0,0001, lote de 128 exemplos e 80 épocas.

As métricas reportadas foram acurácia, macro-F1 e weighted-F1. O F1 combina precisão e *recall* em uma média harmônica; em avaliação multiclasse, a média macro calcula a média não ponderada dos F1 por classe, enquanto a média ponderada considera o suporte de cada classe (F1_SCORE..., 2026). No *script* de avaliação do projeto, a macro-F1 reportada corresponde à média das classes com suporte maior que zero, e a weighted-F1 corresponde à média dos F1 ponderada pelo suporte de cada classe. A Tabela 1 apresenta os principais *checkpoints* comparados durante a seleção do modelo. Esses resultados foram calculados no *split* de validação, com 4.875 regiões válidas e 22 classes anotacionais.

O *checkpoint* com até 1.000 exemplos por classe foi adotado como candidato principal por apresentar a maior macro-F1 entre as configurações comparadas. O *checkpoint* com até 2.000 exemplos por classe obteve a maior acurácia global, mas apresentou macro-F1 e weighted-F1 ligeiramente inferiores. Dado o desbalanceamento das classes, a macro-F1 é relevante porque reduz a tendência de avaliar o desempenho apenas pelas classes com

Tabela 1 – Resultados de validação dos classificadores treinados sobre embeddings visuais do DAM

Configuração	Validação	Acurácia	Macro-F1	Weighted-F1
Até 500 exemplos por classe	4.875	0,7050	0,6631	0,7009
Até 1.000 exemplos por classe	4.875	0,7245	0,6743	0,7242
Até 2.000 exemplos por classe	4.875	0,7274	0,6727	0,7230
1.000 por classe com aumento para classes raras	4.875	0,7171	0,6637	0,7186

Fonte: Elaborado pela autora.

Tabela 2 – Comparação entre validação e teste do checkpoint principal

Split	Regiões válidas	Acurácia	Macro-F1	Weighted-F1
Validação	4.875	0,7245	0,6743	0,7242
Teste	4.925	0,7202	0,6639	0,7183

Fonte: Elaborado pela autora.

maior suporte.

Após a escolha do *checkpoint* principal, foi realizada uma avaliação adicional no *split* de teste. O resultado no teste apresentou acurácia de 0,7202, macro-F1 de 0,6639 e weighted-F1 de 0,7183, valores próximos aos obtidos na validação. A pequena redução em macro-F1 indica perda moderada de desempenho no conjunto de teste, mantendo a interpretação de que os *embeddings* visuais do DAM carregam sinal discriminativo para parte das classes anotacionais.

No treinamento da configuração principal, o melhor resultado de validação ocorreu na época 72, com perda de treino de 0,7127, acurácia de treino de 0,7333, macro-F1 de treino de 0,6937, perda de validação de 0,7400, acurácia de validação de 0,7245 e macro-F1 de validação de 0,6743. Esses valores indicam que os *embeddings* visuais do DAM contêm sinal discriminativo para parte das classes anotacionais, mas não demonstram validação clínica nem desempenho suficiente para uso diagnóstico.

As matrizes de confusão normalizadas permitem observar, para cada classe real, a proporção de exemplos atribuída a cada classe predita. Essa visualização complementa as métricas globais da Tabela 1, pois evidencia diferenças de comportamento entre as configurações avaliadas. As Figuras 20, 21, 22 e 23 apresentam as matrizes correspondentes aos quatro casos comparados.

Além da distribuição completa dos erros, foram gerados gráficos com as principais

confusões de cada configuração, isto é, os pares de classe real e classe predita com maior número absoluto de ocorrências incorretas. As Figuras 24, 25, 26 e 27 apresentam esses resultados.

Figura 20 – Matriz de confusão normalizada do classificador com até 500 exemplos por classe.



Fonte: Elaborado pela autora.

Figura 21 – Matriz de confusão normalizada do classificador com até 1.000 exemplos por classe.



Fonte: Elaborado pela autora.

Figura 22 – Matriz de confusão normalizada do classificador com até 2.000 exemplos por classe.



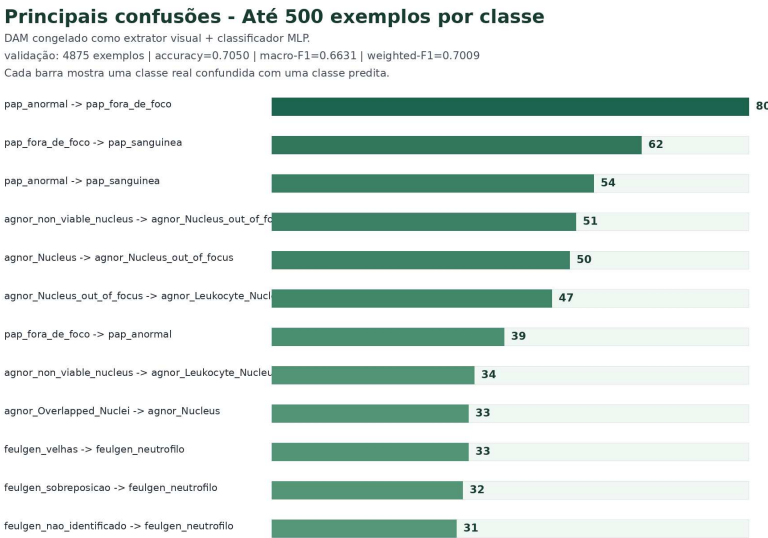
Fonte: Elaborado pela autora.

Figura 23 – Matriz de confusão normalizada do classificador com aumento de dados para classes raras.



Fonte: Elaborado pela autora.

Figura 24 – Principais confusões do classificador com até 500 exemplos por classe.



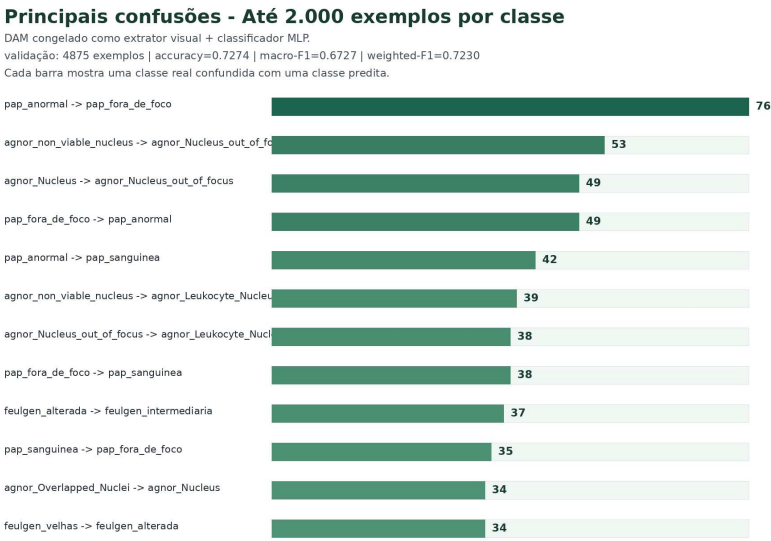
Fonte: Elaborado pela autora.

Figura 25 – Principais confusões do classificador com até 1.000 exemplos por classe.



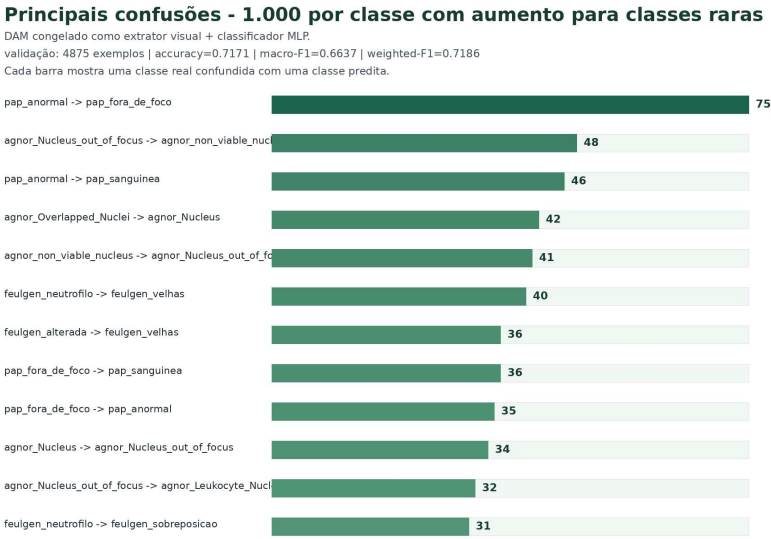
Fonte: Elaborado pela autora.

Figura 26 – Principais confusões do classificador com até 2.000 exemplos por classe.



Fonte: Elaborado pela autora.

Figura 27 – Principais confusões do classificador com aumento de dados para classes raras.



Fonte: Elaborado pela autora.

4.6 EXPERIMENTOS COM CLASSES ANOTACIONAIS

A análise por classe mostrou diferenças entre grupos anotacionais. Algumas classes, como `agnor_Cluster` e `agnor_Satellite`, apresentaram F1 elevado na configuração principal, com valores de 0,9801 e 0,9834, respectivamente. Por outro lado, classes raras ou visualmente próximas apresentaram desempenho inferior ou instável, especialmente quando o suporte de validação era muito baixo.

A Tabela 3 apresenta exemplos de métricas por classe no *checkpoint* principal. As classes foram selecionadas para ilustrar tanto casos de bom desempenho quanto classes problemáticas.

Tabela 3 – Métricas por classe no checkpoint principal

Classe anotacional	Suporte	Precisão	Recall	F1
<code>agnor_Cluster</code>	300	0,9768	0,9833	0,9801
<code>agnor_Satellite</code>	300	0,9801	0,9867	0,9834
<code>feulgen_sujeira</code>	66	0,8889	0,7273	0,8000
<code>pap_anormal</code>	299	0,5099	0,6020	0,5521
<code>pap_em_divisao</code>	4	0,0000	0,0000	0,0000
<code>pap_nao_definida</code>	9	0,8000	0,4444	0,5714
<code>pap_reacional</code>	85	0,5000	0,3412	0,4056

Fonte: Elaborado pela autora.

As principais confusões do *checkpoint* principal ocorreram entre classes que podem compartilhar características visuais ou de qualidade de imagem. O maior número de erros foi de `pap_fora_de_foco` predito como `pap_anormal`, com 81 ocorrências. Também foram observadas confusões entre `agnor_Nucleus` e `agnor_Nucleus_out_of_focus`, entre `agnor_non_viable_nucleus` e `agnor_Nucleus_out_of_focus`, e entre `pap_anormal`, `pap_fora_de_foco` e `pap_sanguinea`. Esses padrões indicam que parte dos erros estão em grupos com atributos visuais próximos, o que limita a interpretação isolada das métricas globais.

4.6.1 Aumento de Dados para Classes Raras

Foi avaliada uma estratégia de aumento de dados no split de treino, mantendo a validação fixa. O objetivo foi elevar para 300 exemplos de treino as classes `pap_em_divisao`, `pap_nao_definida` e `pap_reacional`. Foram aplicadas transformações geométricas, incluindo espelhamentos, rotações de 90, 180 e 270 graus, transposição e transposição reversa.

O manifesto resultante combinou 23.638 exemplos originais com 548 exemplos aumentados, totalizando 24.186 exemplos.

Apesar de melhorar a classe `pap_em_divisao`, cujo F1 passou para 0,0417, a estratégia de aumento não superou o *checkpoint* principal. A acurácia caiu de 0,7245 para 0,7171, a macro-F1 caiu de 0,6743 para 0,6637 e a weighted-F1 caiu de 0,7242 para 0,7186. Além disso, `pap_nao_definida` passou de F1 0,5714 para 0,4000, e `pap_reacional` passou de 0,4056 para 0,3947. Esse resultado sugere que o aumento geométrico simples, no recorte avaliado, não foi suficiente para melhorar o desempenho global ou o equilíbrio entre classes.

4.7 ANÁLISE QUALITATIVA DOS RESULTADOS

Os experimentos indicam que o DAM congelado pode atuar como extrator de representações visuais úteis para algumas classes anotacionais do UFSC Multi-Cyto. O resultado do *checkpoint* principal, com acurácia de 0,7245 e macro-F1 de 0,6743 na validação, e acurácia de 0,7202 e macro-F1 de 0,6639 no teste, indica sinal discriminativo nos *embeddings* visuais, especialmente em classes com maior separação visual no recorte avaliado. Entretanto, as confusões observadas entre classes de Papanicolaou e entre classes relacionadas a foco, viabilidade e sobreposição indicam que as representações não resolvem completamente a discriminação visual exigida pelo *dataset*.

Do ponto de vista metodológico, a principal contribuição foi documentar um caminho de adaptação controlado, através da preparação de regiões anotadas, extração de *embeddings* visuais, treinamento de classificador MLP, avaliação por métricas globais e análise por classe. No *pipeline* integrado, a classificação anotacional foi tratada como contexto computacional para a descrição visual localizada. Mesmo quando o classificador sugere uma classe com confiança elevada, a saída não deve ser convertida em diagnóstico. A formulação textual do sistema deve permanecer restrita a características observáveis e a avisos de limitação.

4.8 DISCUSSÃO DAS LIMITAÇÕES

A primeira limitação diz respeito a ausência de validação clínica. Os resultados foram calculados sobre classes anotacionais do *dataset* UFSC Multi-Cyto e não sobre laudos diagnósticos independentes, categorias clínicas finais ou avaliação descritiva de especialistas. Portanto, os valores de acurácia, macro-F1 e weighted-F1 não devem ser interpretados como desempenho diagnóstico.

A segunda limitação está relacionada ao desbalanceamento e ao baixo suporte de algumas classes. Classes como `pap_em_divisao` e `pap_nao_definida` possuem poucos exemplos nos recortes de treino, validação e teste, o que torna suas métricas instáveis.

Mesmo quando há melhora numérica em uma classe rara, a conclusão deve ser cautelosa, pois poucos exemplos podem alterar substancialmente a métrica.

A terceira limitação diz respeito ao escopo dos experimentos com DAM. A adaptação textual com LoRA demonstrou viabilidade técnica, mas não resolveu a discriminação visual entre classes. Já a abordagem com DAM congelado e MLP treinado sobre *embeddings* visuais produziu o melhor resultado quantitativo, mas não corresponde a *fine-tuning* completo do DAM. Essa distinção é central para evitar extrapolação metodológica.

Por fim, a conversão de CZI para DICOM foi validada funcionalmente com uma amostra real, mas permanece limitada quanto à abrangência da avaliação. O teste documenta a execução bem-sucedida do fluxo para um arquivo específico, mas não avalia compatibilidade com diferentes origens, fabricantes, metadados ou estruturas internas de arquivos CZI. Além disso, o protótipo não implementa visualização direta de arquivos DICOM na interface web, já que a análise interativa é realizada sobre *tiles* exportados em formato de imagem convencional.

5 CONCLUSÃO

O trabalho desenvolveu e avaliou tecnicamente um protótipo computacional para apoio à análise regional de imagens citopatológicas. A abordagem implementou um fluxo de recebimento e processamento de imagens, suporte a arquivos CZI, interoperabilidade com DICOM aplicada a WSI, segmentação interativa com SAM2, descrição visual localizada com DAM e uma interface web de demonstração.

Do ponto de vista de software, o protótipo foi organizado em arquitetura modular, com separação entre domínio, serviços, infraestrutura, rotas de API e interface web. Essa organização permitiu implementar e documentar fluxos de upload, consulta de metadados, extração de *tiles*, segmentação, descrição, análise regional integrada e conversão de CZI para DICOM. A validação funcional indicou viabilidade técnica dos módulos principais e registrou a conversão bem-sucedida de uma amostra CZI real para saída DICOM, ainda que essa evidência permaneça limitada a um arquivo específico.

Do ponto de vista experimental, o trabalho investigou o uso do DAM em regiões anotadas do *dataset* UFSC Multi-Cyto. A adaptação textual com LoRA apresentou funcionamento técnico, mas o melhor desempenho quantitativo foi obtido com o DAM congelado como extrator de *embeddings* visuais e uma cabeça MLP treinada para classes anotacionais. A configuração principal apresentou acurácia de 0,7245, macro-F1 de 0,6743 e weighted-F1 de 0,7242 no conjunto de validação, o que sugere utilidade computacional dos *embeddings* visuais do DAM para classificação anotacional auxiliar.

Em síntese, o trabalho demonstrou a viabilidade de uma abordagem para análise regional em imagens citopatológicas, articulando segmentação interativa, descrição visual localizada e classificação anotacional em um protótipo único.

5.1 TRABALHOS FUTUROS

Como continuidade desse trabalho, recomenda-se ampliar a avaliação do módulo de conversão de CZI para DICOM, utilizando arquivos provenientes de diferentes equipamentos, fabricantes e configurações de aquisição. Essa ampliação permitiria verificar a preservação dos dados visuais e dos metadados em diferentes estruturas de arquivo. Também se propõe integrar ao protótipo um visualizador compatível com DICOM, associado a um servidor para armazenamento e recuperação dos objetos gerados, possibilitando a inspeção das imagens convertidas no próprio fluxo da aplicação.

Outra frente consiste na avaliação das descrições visuais geradas pelo DAM com a participação de especialistas em citopatologia. Essa avaliação pode considerar critérios como correspondência com a região segmentada, adequação da terminologia, relevância das características descritas, clareza e presença de informações inconsistentes. Tal procedimento permitiria compreender com maior precisão as possibilidades e limitações do modelo como ferramenta de apoio à anotação citopatológica.

Em relação à classificação anotacional, trabalhos posteriores podem explorar diferentes arquiteturas sobre os *embeddings* visuais extraídos pelo DAM, bem como estratégias de balanceamento, aumento de dados e tratamento específico das classes com menor número de exemplos. Também podem ser investigadas estratégias de adaptação do modelo que considerem conjuntamente as representações visuais e as características das diferentes colorações presentes no *dataset*.

Por fim, propõe-se a construção de bases multimodais que associem regiões citopatológicas segmentadas, máscaras, classes anotacionais e descrições textuais elaboradas ou revisadas por especialistas. A disponibilidade desse tipo de conjunto de dados permitiria realizar adaptações supervisionadas, estabelecer critérios quantitativos para a avaliação das descrições e investigar modelos mais especializados na análise regional de imagens citopatológicas.

REFERÊNCIAS

- ANDREÓLI PETROLINI, Vinícius et al. Collaborative Telepathology in a Statewide Telemedicine Environment: First Tests in the Context of the Brazilian Public Healthcare System. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). [S.l.]: IEEE Computer Society, 2019. p. 684–689. DOI: 10.1109/CBMS.2019.00139. Disponível em: <https://doi.ieeecomputersociety.org/10.1109/CBMS.2019.00139>. Acesso em: 1 dez. 2024.
- ASIF, S.; WENHUI, Y.; UR-REHMAN, S. et al. Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision. **Archives of Computational Methods in Engineering**, 2024. DOI: 10.1007/s11831-024-10148-w. Acesso em: 2 dez. 2024.
- BOARETTO, Naiara et al. Câncer: Uma Revisão Integrativa por Estudantes de Medicina. **Revista Brasileira de Cancerologia**, v. 9, n. 2, e-20246402, 2024. DOI: 10.32963/bcmufsc.v9i2.6402. Acesso em: 30 nov. 2024.
- BRASIL. **Câncer**. Ministério da Saúde, Instituto Nacional de Câncer. 2021. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer>. Acesso em: 9 jun. 2024.
- BRUNDAGE, David et al. **Whole Slide Image to DICOM Conversion as Event-Driven Cloud Infrastructure**. [S.l.: s.n.], 2022. DOI: 10.48550/arXiv.2203.13888. arXiv: 2203.13888 [cs.DC]. Disponível em: <https://arxiv.org/abs/2203.13888>. Acesso em: 11 jun. 2026.
- CARL ZEISS MICROSCOPY. **CZI Image File Format**. 2026. Disponível em: <https://www.zeiss.com/microscopy/en/products/software/zeiss-zen/czi-image-file-format.html>. Acesso em: 11 jun. 2026.
- CFM regulamenta a transmissão de imagens patológicas e garante maior segurança ao diagnóstico de doenças. Conselho Federal de Medicina. 2019. Disponível em: <https://portal.cfm.org.br/noticias/cfm-regulamenta-a-transmissao-de-imagens-patologicas-e-garante-maior-seguranca-ao-diagnostico-de-doencas>. Acesso em: 1 dez. 2024.
- DICOM STANDARDS COMMITTEE. **DICOM Standard: Current Edition**. National Electrical Manufacturers Association. 2026. Disponível em: <https://www.dicomstandard.org/current/>. Acesso em: 11 jun. 2026.
- EVARISTO, Eric et al. **Systematic Literature Review on Multimodal Large Language Models applied to Medical Image Data**. [S.l.: s.n.], maio 2025. DOI: 10.13140/RG.2.2.32280.10244.
- RESOLUÇÃO define telepatologia como especialidade médica. Federação dos Hospitais, Clínicas, Casas de Saúde, Laboratórios de Pesquisas e Análises Clínicas e Demais

Estabelecimentos de Serviços de Saúde do Estado de São Paulo. 2019. Disponível em: <https://fehoesp360.org.br/sindsuzano/resolucao-define-telepatologia-como-especialidade-medica/>. Acesso em: 1 dez. 2024.

FOWLER, JR; MAANI, EV; DUNTON, CJ et al. **Cervical Cancer**. [Updated 2023 Nov 12]. National Center for Biotechnology Information. Jan. 2024. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK431093/>. Acesso em: 1 dez. 2024.

GIARNIERI, Enrico; SCARDAPANE, Simone. Towards Artificial Intelligence Applications in Next Generation Cytopathology. **Biomedicines**, v. 11, n. 8, 2023. ISSN 2227-9059. DOI: 10.3390/biomedicines11082225. Disponível em: <https://www.mdpi.com/2227-9059/11/8/2225>. Acesso em: 2 dez. 2024.

GOODE, Adam et al. OpenSlide: A Vendor-Neutral Software Foundation for Digital Pathology. **Journal of Pathology Informatics**, v. 4, n. 1, p. 27, 2013. DOI: 10.4103/2153-3539.119005. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3815078/>. Acesso em: 11 jul. 2026.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: MIT Press, 2016. Disponível em: <https://www.deeplearningbook.org/>. Acesso em: 11 jun. 2026.

HU, Edward J. et al. **LoRA: Low-Rank Adaptation of Large Language Models**. [S.l.: s.n.], 2021. DOI: 10.48550/arXiv.2106.09685. arXiv: 2106.09685 [cs.CL]. Disponível em: <https://arxiv.org/abs/2106.09685>. Acesso em: 11 jun. 2026.

JOHNS HOPKINS MEDICINE. **Cytology**. 2024. Disponível em: <https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/cytology>. Acesso em: 30 nov. 2024.

KANTSEMAL, Anthony. **How AI Helps in Whole Slide Image Analysis for Cancer Detection**. 2024. Disponível em: <https://mobidev.biz/blog/ai-cancer-detection-approaches-challenges-opportunities>. Acesso em: 2 dez. 2024.

KUMAR, N.; GUPTA, R.; GUPTA, S. Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. **Journal of Digital Imaging**, v. 33, n. 4, p. 1034–1040, ago. 2020. DOI: 10.1007/s10278-020-00351-z. Acesso em: 2 dez. 2024.

LIALIN, Vladislav et al. **Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning**. [S.l.: s.n.], 2023. DOI: 10.48550/arXiv.2303.15647. arXiv: 2303.15647 [cs.CL]. Disponível em: <https://arxiv.org/abs/2303.15647>. Acesso em: 11 jun. 2026.

LIAN, Long et al. **Describe Anything: Detailed Localized Image and Video Captioning**. [S.l.: s.n.], 2025. DOI: 10.48550/arXiv.2504.16072. arXiv: 2504.16072 [cs.CV]. Disponível em: <https://arxiv.org/abs/2504.16072>. Acesso em: 11 jun. 2026.

MASSON, de. **Untitled Item**. Jul. 2022. DOI: 10.6084/m9.figshare.20288619.v1. Disponível em: https://figshare.com/articles/figure/Untitled_Item/20288619. Acesso em: 1 dez. 2024.

MEDICINE, National Library of. **Collection Development Guidelines of the National Library of Medicine: Cytology**. National Center for Biotechnology Information. 2003. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK518827/>. Acesso em: 30 nov. 2024.

PANGARKAR, M. A. The Bethesda System for reporting cervical cytology. **Cytojournal**, v. 19, p. 28, abr. 2022. DOI: 10.25259/CMAS_03_07_2021. Acesso em: 2 dez. 2024.

PDQ ADULT TREATMENT EDITORIAL BOARD. **Lip and Oral Cavity Cancer Treatment (PDQ): Patient Version**. National Cancer Institute. 16 maio 2025. Disponível em: <https://www.cancer.gov/types/head-and-neck/patient/adult/lip-mouth-treatment-pdq>. Acesso em: 11 jul. 2026.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

RAMIREZ PUEBLA, Shamayim Tabita; MARK WELCH, Jessica; BORISY, Gary. **Improved Visualization of Oral Microbial Consortia: Associated Images**. Dados de imagem associados ao artigo publicado no Journal of Dental Research, DOI: 10.1177/00220345241251784. Zenodo. 2024. DOI: 10.5281/zenodo.11042811. Disponível em: <https://zenodo.org/records/11042811>. Acesso em: 15 jun. 2026.

RAVI, Nikhila et al. **SAM 2: Segment Anything in Images and Videos**. [S.l.: s.n.], 2024. DOI: 10.48550/arXiv.2408.00714. arXiv: 2408.00714 [cs.CV]. Disponível em: <https://arxiv.org/abs/2408.00714>. Acesso em: 11 jun. 2026.

SANTOS, Luís Otávio et al. **UFSC Multi-Cyto: Multi-Stain Cytopathology Image Dataset**. Versão V1. Mendeley Data. 2025. DOI: 10.17632/b8bygdxr7r.1. Disponível em: <https://data.mendeley.com/datasets/b8bygdxr7r/1>. Acesso em: 11 jun. 2026.

SARTORI, Beatriz; HILLESHEIN, Klisman Drescher; RIBEIRO, Osmar Damasceno. TELEPATOLOGIA. **Anais de Medicina**, v. 1, n. 1, p. 38, maio 2014. Disponível em: <https://periodicos.unoesc.edu.br/anaisdemedicina/article/view/4659>. Acesso em: 1 dez. 2024.

F1_SCORE. Scikit-learn Developers. 2026. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html. Acesso em: 11 jun. 2026.

SHAFI, S.; PARWANI, A.V. Artificial intelligence in diagnostic pathology. **Diagnostic Pathology**, v. 18, n. 109, 2023. DOI: 10.1186/s13000-023-01375-z. Acesso em: 2 dez. 2024.

STANFORD PATHOLOGY. **Cytopathology**. 2024. Disponível em: <https://stanfordlab.com/anatomic-pathology/cytopathology.html>. Acesso em: 30 nov. 2024.

TREVISAN REZENDE, Mariana et al. **CRIC Cervix Microscope Slide Image 157**. [S.l.: s.n.], 2020. DOI: 10.6084/m9.figshare.12230081.v2. Disponível em: https://figshare.com/articles/figure/CRIC_Cervix_Microscope_Slide_Image_157/12230081. Acesso em: 30 nov. 2024.

VICTÓRIA MATIAS, André et al. Segmentation, Detection and Classification of Cell Nuclei on Oral Cytology Samples Stained with Papanicolaou. **SN Computer Science**, Springer Science e Business Media LLC, v. 2, n. 4, maio 2021. ISSN 2661-8907. DOI: 10.1007/s42979-021-00676-8. Disponível em: <http://dx.doi.org/10.1007/s42979-021-00676-8>. Acesso em: 9 jun. 2024.

VICTÓRIA MATIAS, André et al. What is the state of the art of computer vision-assisted cytology? A Systematic Literature Review. **Computerized Medical Imaging and Graphics**, v. 91, p. 101934, 2021. ISSN 0895-6111. DOI: <https://doi.org/10.1016/j.compmedimag.2021.101934>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0895611121000835>. Acesso em: 1 dez. 2024.

ZHANG, B.; SHI, H.; WANG, H. Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach. **Journal of Multidisciplinary Healthcare**, v. 16, p. 1779–1791, 2023. DOI: 10.2147/JMDH.S410301. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10312208/>. Acesso em: 9 jun. 2024.

ZHUANG, Fuzhen et al. **A Comprehensive Survey on Transfer Learning**. [S.l.: s.n.], 2020. DOI: 10.48550/arXiv.1911.02685. arXiv: 1911.02685 [cs.LG]. Disponível em: <https://arxiv.org/abs/1911.02685>. Acesso em: 11 jun. 2026.

Apêndices

APÊNDICE A – REPOSITÓRIO

O código-fonte associado a este trabalho encontra-se organizado nos repositórios abaixo. Para reprodução dos experimentos principais, utilizar as instruções presentes no arquivo README.md.

- <https://github.com/tainadacruz/cytopathology-image-descriptor>
- <https://codigos.ufsc.br/taina.cruz/cytopathology-image-descriptor>

APÊNDICE B – ARTIGO EM FORMATO SBC

Desenvolvimento de um Protótipo Computacional para Apoio à Análise Citopatológica: uma Abordagem com *Describe Anything Model* para Descrição de Regiões Celulares

Tainá da Cruz¹, Luís Otávio Santos¹, Aldo von Wangenheim¹,
Alexandre Sherlley Casimiro Onofre¹

¹Departamento de Informática e Estatística – Centro Tecnológico
Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brasil

`taina.cruz@grad.ufsc.br, luis.otaviosantoss@gmail.com,`

`aldo.vw@ufsc.br, alexandre.onofre@ufsc.br`

Abstract. *This work develops and technically evaluates a computational prototype to support the annotation and visual analysis of cytopathological images, integrating the Describe Anything Model (DAM) for localized visual description and annotational classification. The prototype combines image reception and processing, CZI-to-DICOM conversion for Whole Slide Imaging, point-based interactive segmentation with SAM2, and localized description with DAM, over a FastAPI backend and a React web interface. On the UFSC Multi-Cyto dataset, using the frozen DAM as a visual feature extractor, the main classifier reached 0.7245 accuracy and 0.6743 macro-F1 on validation, indicating exploratory viability without characterizing diagnostic validation.*

Keywords: *Cytopathological Images. Describe Anything Model. Interactive Segmentation. Localized Visual Description. Annotation-Based Classification.*

Resumo. *Este trabalho desenvolve e avalia tecnicamente um protótipo computacional para apoio à anotação e à análise visual de imagens citopatológicas, integrando o Describe Anything Model (DAM) para descrição visual localizada e classificação anotacional. O protótipo reúne recebimento e processamento de imagens, conversão CZI para DICOM aplicada a Whole Slide Imaging, segmentação interativa por pontos com SAM2 e descrição localizada com DAM, sobre um backend em FastAPI e uma interface web em React. No dataset UFSC Multi-Cyto, usando o DAM congelado como extrator visual, o classificador principal alcançou acurácia de 0,7245 e macro-F1 de 0,6743 na validação, indicando viabilidade exploratória, sem caracterizar validação diagnóstica.*

Palavras-chave: *Imagens Citopatológicas. Describe Anything Model. Segmentação Interativa. Descrição Visual Localizada. Classificação Anotacional.*

1. Introdução

O câncer consiste em um grupo de mais de 100 doenças caracterizadas pelo crescimento desordenado de células, capazes de se disseminar para outros tecidos por meio de metástase [Brasil 2021]. A capacidade de invasão e proliferação distingue os tumores malignos dos benignos [Boaretto et al. 2024]. A detecção precoce é fundamental para o tratamento,

e os estágios iniciais podem ser identificados por alterações no núcleo das células afetadas; contudo, a análise manual é demorada e sujeita a variações de interpretação entre profissionais [Victória Matias et al. 2021]. Nesse contexto, a citopatologia examina células obtidas de tecidos ou fluidos corporais [Johns Hopkins Medicine 2024], e é amplamente empregada no rastreamento do câncer cervical [Fowler et al. 2024].

A digitalização de lâminas permite que regiões celulares sejam analisadas em ambiente computacional [Kumar et al. 2020]. Entretanto, a manipulação dessas imagens envolve formatos específicos de microscopia, como o *Carl Zeiss Image* (CZI), e padrões de interoperabilidade em imagem médica, como o DICOM aplicado a *Whole Slide Imaging* (WSI) [Carl Zeiss Microscopy 2026, DICOM Standards Committee 2026, Brundage et al. 2022]. Em revisão sistemática conduzida no contexto deste trabalho, [Evaristo et al. 2025] analisaram *Multimodal Large Language Models* (MLLMs) aplicados a imagem médica e apontaram limitações relacionadas à escassez de dados anotados, custo computacional, confiabilidade, alucinações e integração clínica. Esses resultados reforçam a relevância de investigar, de forma exploratória, o uso de modelos multimodais nesse domínio [Zhang et al. 2023].

Diante disso, este trabalho investiga uma alternativa computacional baseada em descrição visual localizada e segmentação interativa orientada por pontos. Modelos como o *Describe Anything Model* (DAM) foram propostos para gerar descrições detalhadas de regiões específicas em imagens e vídeos [Lian et al. 2025], enquanto o *Segment Anything Model 2* (SAM2) atua como modelo fundacional de segmentação promptável [Ravi et al. 2024]. A avaliação considera testes funcionais, exemplos de entrada e saída, análise qualitativa e experimentos exploratórios com classes anotacionais do *dataset* UFSC Multi-Cyto [Santos et al. 2025].

1.1. Objetivos

O objetivo geral é desenvolver e avaliar um protótipo computacional para apoio à análise de imagens derivadas de exames citológicos, integrando processamento de imagens, segmentação interativa com SAM2, descrição visual localizada com DAM, classificação anotacional baseada em *embeddings* visuais e suporte exploratório à interoperabilidade CZI-DICOM. Como objetivos específicos, o trabalho busca: (a) implementar uma API para recebimento, registro e processamento de imagens citopatológicas; (b) implementar suporte a arquivos CZI com conversão para DICOM; (c) integrar segmentação interativa por pontos com SAM2, com retorno estruturado de máscara, polígono, caixa delimitadora e metadados; (d) integrar descrição visual localizada de regiões segmentadas com DAM; (e) investigar o uso do DAM como extrator visual para classificação anotacional; e (f) avaliar o protótipo por validação funcional, análise qualitativa e experimentos com o *dataset* UFSC Multi-Cyto.

2. Fundamentação Teórica

2.1. Citopatologia, Colorações e Interoperabilidade

A citologia estuda a morfologia, a estrutura e as alterações das células, e o termo citopatologia é empregado quando o foco está no processo diagnóstico, em que células obtidas por esfoliação, raspagem ou aspiração são processadas em

lâminas e examinadas microscopicamente para identificar tumores e lesões pré-cancerosas [National Library of Medicine 2003, Stanford Pathology 2024]. Por ser um procedimento de baixo custo e não invasivo, é utilizado no apoio ao diagnóstico de diversas doenças, embora a realização manual permaneça demorada e vulnerável à interpretação humana [Victória Matias et al. 2021].

A análise citológica emprega diferentes técnicas de coloração que destacam estruturas celulares de interesse [Victória Matias et al. 2021]. O *dataset* UFSC Multi-Cyto reúne imagens coradas por Papanicolaou (cavidade oral), Feulgen e *Argyrophilic Nucleolar Organizer Regions* (AgNOR), estas duas provenientes do colo do útero [Santos et al. 2025]. No rastreamento cervical, associado à infecção persistente por tipos oncogênicos do Papilomavírus Humano (HPV), o exame de Papanicolaou permite identificar alterações citológicas anormais [Fowler et al. 2024]. No câncer oral, de maioria escamosa, a citologia esfoliativa apoia a identificação de anormalidades [PDQ Adult Treatment Editorial Board 2025]. Para padronizar relatórios cervicais, o Sistema Bethesda estabelece categorias que refletem a gravidade das lesões [Pangarkar 2022].

O WSI permite a digitalização completa de lâminas em alta resolução, gerando lâminas virtuais que preservam detalhes morfológicos e podem ser integradas a algoritmos de IA [Kumar et al. 2020, Kantsemlal 2024]. Nesse domínio, coexistem formatos proprietários de fabricantes de *scanners* [Goode et al. 2013]. Destacam-se, para este trabalho, o formato CZI, voltado à aquisição e ao armazenamento de imagens multidimensionais de microscopia [Carl Zeiss Microscopy 2026], e o padrão DICOM aplicado a WSI, que favorece a padronização e a interoperabilidade de imagens de lâmina inteira [DICOM Standards Committee 2026, Brundage et al. 2022]. A telepatologia, regulamentada no Brasil em 2019, reforça a importância de integridade, confidencialidade e rastreabilidade no tratamento de imagens médicas digitais [Sartori et al. 2014, FEHOESP 2019, CFM 2019].

2.2. Segmentação Interativa e Descrição Visual Localizada

Avanços recentes em IA têm automatizado tarefas antes manuais, como classificação celular e análise morfológica, otimizando o processamento de grandes volumes de dados [Shafi and Parwani 2023, Asif et al. 2024]. Algoritmos de *Deep Learning* possibilitam a automatização de processos complexos, como detecção e segmentação celular [Giarnieri and Scardapane 2023]. A segmentação difere da classificação por delimitar espacialmente os elementos de interesse, frequentemente em nível de *pixel*; na segmentação interativa, a delimitação é orientada por entradas do usuário, como cliques, pontos, caixas ou máscaras [Ravi et al. 2024].

O SAM2 é descrito como um modelo fundacional para segmentação visual promptável em imagens e vídeos, tratando a imagem como caso particular de vídeo com um único quadro; emprega arquitetura baseada em transformadores com *streaming memory*, capaz de armazenar informações sobre o objeto segmentado, os *prompts* fornecidos e as previsões anteriores [Ravi et al. 2024]. O DAM, por sua vez, é proposto para *detailed localized captioning*, isto é, a geração de descrições textuais detalhadas de regiões específicas. Para tanto, combina um *focal prompt*, que representa a região com maior densidade de informação visual, e uma arquitetura visual localizada, que incorpora a

máscara da região de forma espacialmente alinhada e integra o contexto global da imagem [Lian et al. 2025].

O aprendizado por transferência reaproveita conhecimento de um domínio de origem para reduzir a dependência de grandes volumes de dados no domínio-alvo [Zhuang et al. 2020]. Entre os métodos de *Parameter-Efficient Fine-Tuning* (PEFT), que treinam apenas um subconjunto reduzido de parâmetros [Lialin et al. 2023], o *Low-Rank Adaptation* (LoRA) congela os pesos pré-treinados e insere matrizes treináveis de baixo posto [Hu et al. 2021]. Alternativamente, um modelo congelado pode ser usado como extrator de *embeddings*, que alimentam um classificador supervisionado, como um *Multilayer Perceptron* (MLP) [Goodfellow et al. 2016].

3. Trabalhos Relacionados

[Andreóli Petrolini et al. 2019] desenvolveram um sistema de telepatologia colaborativa para o Sistema Único de Saúde brasileiro, integrando diagnóstico remoto, ambiente virtual que simula um microscópio e geração padronizada de relatórios em DICOM *Structured Reporting*. A avaliação, conduzida com 21 participantes por meio do modelo ADE-QUATE [Alves 2017], demonstrou resultados positivos em efetividade e conforto de uso. O presente trabalho se relaciona a essa proposta por atuar sobre imagens patológicas digitais e interoperabilidade, mas se diferencia por não propor telepatologia colaborativa nem diagnóstico remoto, e sim um protótipo de API independente para apoio à anotação e à análise visual localizada.

[Evaristo et al. 2025] apresentaram uma revisão sistemática sobre MLLMs aplicados a dados de imagem médica, analisando publicações de 2023 a março de 2025. Os autores observaram diversificação das aplicações, com apoio diagnóstico especializado, sistemas interativos, geração de relatórios e localização visual, mas destacaram limitações como escassez de dados anotados em larga escala, custo de anotação especializada, custos computacionais, risco de alucinações e desafios de integração clínica. Esses resultados se relacionam diretamente a este trabalho, pois reforçam a dificuldade de obter bases estruturadas com pares de imagem médica e descrição textual especializada.

4. Desenvolvimento

4.1. Método de Pesquisa

O trabalho caracteriza-se como pesquisa aplicada, de caráter exploratório e descritivo, conduzida por meio do desenvolvimento e da avaliação técnica de um protótipo computacional [Prodanov and Freitas 2013]. O *backend* foi desenvolvido em Python com o *framework* FastAPI, e a interface web de demonstração em TypeScript com React e Vite. Foram utilizadas bibliotecas para armazenamento local, manipulação de imagens, leitura de arquivos CZI, conversão para DICOM, segmentação com SAM2 e descrição com DAM [Ravi et al. 2024, Lian et al. 2025]. O objetivo metodológico não foi validar clinicamente um sistema de diagnóstico, mas desenvolver e avaliar uma API para apoio à anotação de imagens citopatológicas.

4.2. Conjunto de Dados

Os experimentos exploratórios utilizaram o UFSC Multi-Cyto, disponibilizado no *Mendeley Data* [Santos et al. 2025]. O conjunto contém 17.422 fragmentos de imagens citopatológicas com anotações por instância, fornecidos nos formatos COCO e YOLO e divididos

em treinamento (12.502), validação (2.448) e teste (2.472). Durante o desenvolvimento, foram identificadas 22 classes anotacionais, agrupadas pelos prefixos *pap_* (Papanicolaou), *agnor_* (AgNOR) e *feulgen_* (Feulgen). Para os experimentos com DAM, o processo recorta a região anotada, aplica a máscara da anotação como canal alfa e registra os exemplos em um manifesto estruturado. Para os testes de interoperabilidade, foi utilizado um arquivo CZI do conjunto associado ao trabalho de [Ramirez Puebla et al. 2024], empregado apenas na validação funcional de leitura, extração de *tile* e conversão para DICOM.

4.3. Arquitetura do Sistema

O protótipo foi organizado em camadas, separando entidades de domínio, orquestração de casos de uso, roteadores HTTP e adaptadores concretos para armazenamento, leitura de imagens, conversão, segmentação e descrição. A aplicação FastAPI registra os roteadores de saúde, arquivos, *jobs*, segmentação, descrição e *pipeline* sob o prefixo */api/v1*. O armazenamento local é organizado em diretórios para *uploads*, *tiles* exportados, máscaras, *jobs* e saídas DICOM.

O **módulo de arquivos e interoperabilidade** recebe imagens por *upload* HTTP, registra os arquivos, permite acesso a imagens comuns (via *tile full*) ou a regiões extraídas de arquivos CZI, e oferece conversão para DICOM por meio de *jobs* assíncronos. O **módulo de segmentação interativa** recebe o identificador do arquivo, um *tile* opcional e uma lista de pontos rotulados como positivos (1) ou negativos (0); o adaptador do SAM2 carrega a imagem em RGB, executa a inferência, seleciona a máscara de maior pontuação e a converte em caixa delimitadora, polígono e área em *pixels*. O **módulo de descrição visual localizada** combina a imagem RGB com a máscara em uma imagem RGBA, calcula um recorte com contexto ao redor da caixa delimitadora e envia a região ao servidor DAM; a instrução textual restringe a resposta a características visuais observáveis em português, explicitando que a saída não deve conter diagnóstico clínico ou classificação de malignidade [Lian et al. 2025].

O **pipeline integrado de análise regional**, exposto pelo *endpoint* POST */api/v1/pipeline/analyze-region*, reúne esses módulos em um único fluxo, apresentado na Figura 4 (Apêndice A). A partir do arquivo, dos pontos e de uma instrução textual, o serviço exporta a imagem ou *tile*, executa a segmentação, encaminha a região à classificação anotacional auxiliar e solicita a descrição ao DAM, retornando um objeto JSON com os blocos *segmentation*, *description* e *classification*.

4.4. Adaptação Exploratória e Classificação Anotacional

Além da descrição localizada, foram implementados experimentos exploratórios de adaptação ao domínio citopatológico, motivados pela escassez de bases com pares de imagem e texto especializado [Evaristo et al. 2025]. A primeira rota consistiu na adaptação textual leve com LoRA, aplicada ao componente linguístico do DAM, sem ajuste de ponta a ponta dos componentes visuais [Hu et al. 2021, Lialin et al. 2023]. A segunda rota utilizou o DAM congelado como extrator de *embeddings* visuais das regiões anotadas, com os vetores resultantes servindo ao treinamento de uma cabeça classificadora MLP, em estratégia fundamentada no aprendizado por transferência [Zhuang et al. 2020]. O servidor DAM foi estendido com um *endpoint* local de classificação anotacional auxiliar, cujo

resultado (classe mais provável, confiança e lista top- k) é usado como contexto para a descrição textual.

4.5. Limitações de Implementação

As limitações foram tratadas como delimitações metodológicas. A execução do SAM2 e do DAM depende de *checkpoints*, arquivos de configuração e recursos computacionais compatíveis. O suporte a CZI e a conversão para DICOM dependem de bibliotecas externas e da estrutura interna dos arquivos. O UFSC Multi-Cyto fornece imagens, máscaras e classes anotacionais, mas não descrições textuais especializadas por região, de modo que os experimentos de adaptação e classificação foram estruturados a partir das classes do *dataset* [Santos et al. 2025]. A interface web demonstra o fluxo com imagens comuns e *tiles* de CZI, mas não implementa um visualizador completo para arquivos DICOM.

5. Resultados e Experimentos

Os experimentos foram conduzidos em ambiente Linux (x86_64), com API em Python 3.10.14 e FastAPI, experimentos de DAM em PyTorch 2.6.0+cu124, e duas GPUs NVIDIA GeForce RTX 3090 (24 GB) e uma RTX 5080 (16 GB). A validação funcional confirmou o funcionamento dos módulos de saúde, arquivos e metadados, segmentação interativa, descrição localizada, classificação anotacional auxiliar e *pipeline* integrado. A Figura 1 ilustra uma máscara produzida pelo SAM2 a partir de pontos indicados pelo usuário.

A partir da região segmentada, o *pipeline* solicita ao DAM uma descrição visual localizada. A Figura 2 apresenta um exemplo de saída retornada pelo protótipo, restrita a características visuais observáveis, como forma, contorno, coloração, textura e sobreposição, acompanhada de aviso de que a descrição não corresponde a diagnóstico clínico, além da classe anotacional sugerida pela classificação.

O fluxo de conversão CZI para DICOM foi validado funcionalmente com um arquivo CZI real, gerando uma saída *.dcm* disponibilizada como resultado do *job*. Esse teste documenta a execução bem-sucedida do fluxo para uma amostra, sem representar validação ampla para diferentes fabricantes, configurações de aquisição ou estruturas internas de arquivos CZI.

5.1. Adaptação Textual com LoRA

A adaptação textual com LoRA teve caráter exploratório, verificando se o componente linguístico do DAM poderia aprender um formato de resposta mais aderente às classes anotacionais. Os testes demonstraram funcionamento técnico do *pipeline* de preparação de dados, carregamento do modelo e aplicação dos adaptadores, mas a adaptação não resolveu de forma consistente a discriminação visual entre classes, resultado coerente com o fato de os componentes visuais permanecerem congelados. Por esse motivo, a rota quantitativa principal passou a ser a classificação sobre *embeddings* visuais.

5.2. DAM Congelado como Extrator Visual

Na rota principal, o DAM foi mantido congelado e utilizado para extrair *embeddings* visuais das regiões anotadas, treinando-se uma cabeça MLP para prever as 22 classes anotacionais. A configuração principal adotou *pooling mean*, normalização *l2* e camada

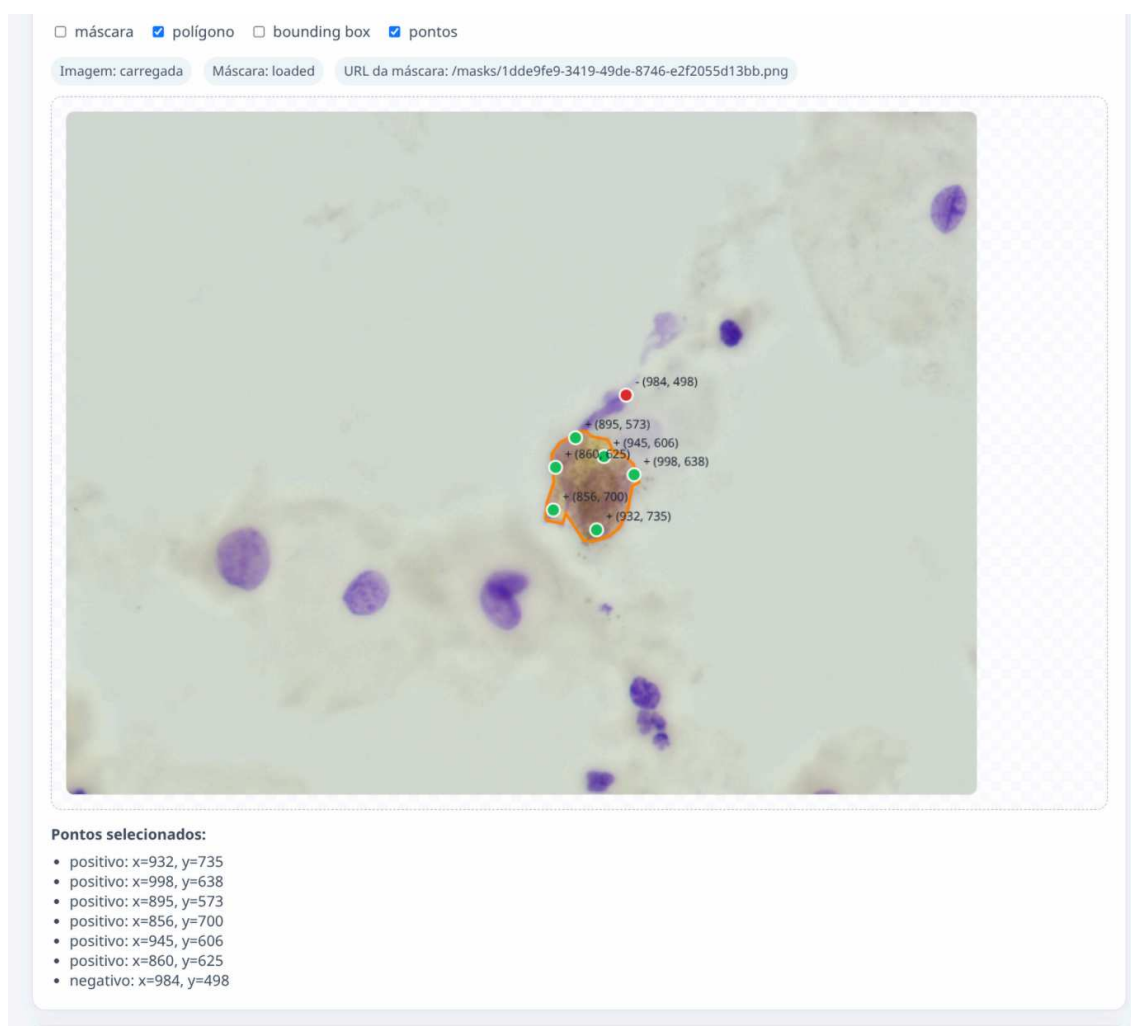


Figura 1. Máscara do SAM2 aplicada a uma região citopatológica selecionada por pontos.

oculta de dimensão 1024, com *dropout* de 0,2, taxa de aprendizado de 0,0005, decaimento de peso de 0,0001, lote de 128 exemplos e 80 épocas. As métricas reportadas foram acurácia, macro-F1 e *weighted-F1*; a média macro reduz a tendência de avaliar o desempenho apenas pelas classes de maior suporte [Scikit-learn Developers 2026]. A Tabela 1 compara as configurações avaliadas sobre 4.875 regiões válidas de validação.

O *checkpoint* com até 1.000 exemplos por classe foi adotado como candidato principal por apresentar a maior macro-F1. Embora a configuração com até 2.000 exemplos tenha obtido acurácia global superior, sua macro-F1 e *weighted-F1* foram inferiores. No *split* de teste, o *checkpoint* principal apresentou acurácia de 0,7202, macro-F1 de 0,6639 e *weighted-F1* de 0,7183, valores próximos aos da validação. A Figura 3 apresenta a matriz de confusão normalizada dessa configuração.

A análise por classe revelou diferenças entre grupos anotacionais. Classes como *agnor_Cluster* e *agnor_Satellite* apresentaram F1 elevado (0,9801 e 0,9834), enquanto classes raras ou visualmente próximas apresentaram desempenho inferior ou instável, sobretudo com baixo suporte de validação, como indica a Tabela 2. As princi-

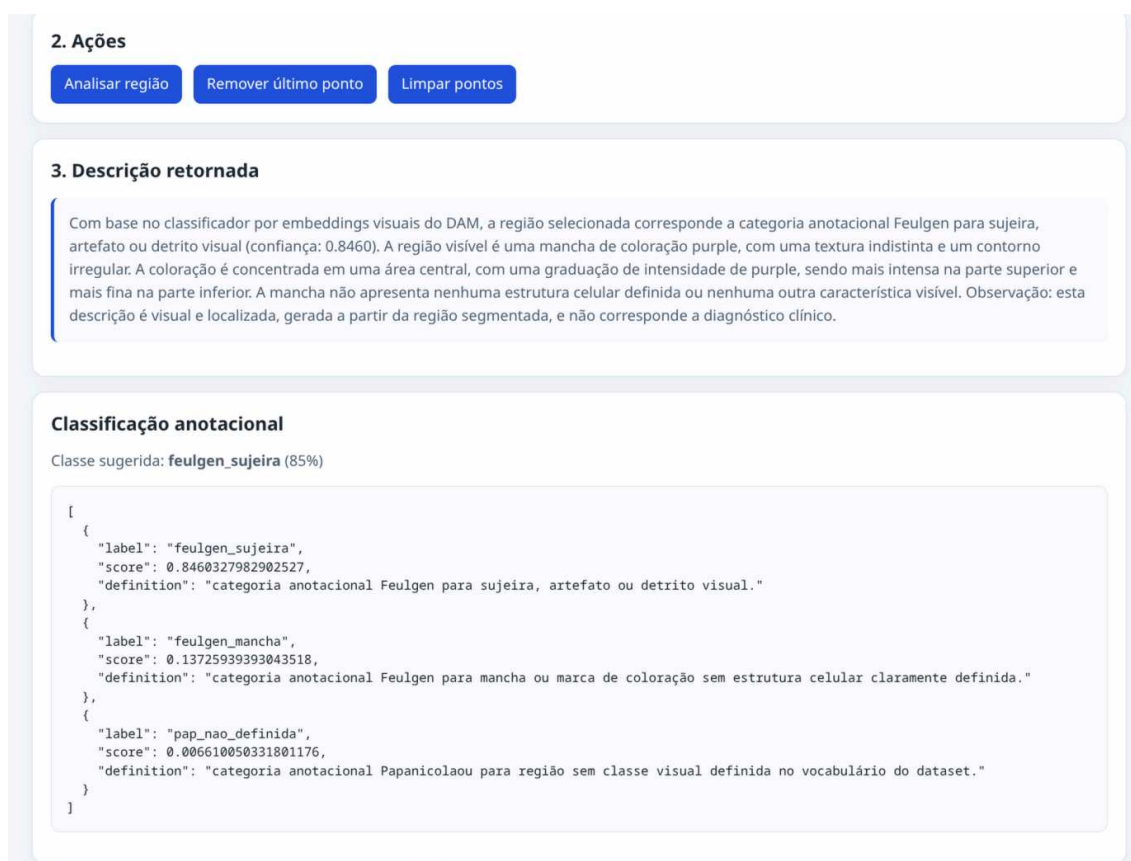


Figura 2. Descrição visual localizada retornada pelo protótipo para uma região selecionada, com a classe anotacional sugerida.

país confusões ocorreram entre `pap_fora_de_foco` e `pap_anormal` (81 ocorrências), entre `agnor_Nucleus` e `agnor_Nucleus_out_of_focus`, e entre classes de Papanicolaou associadas a foco, sangue e anormalidade, indicando que parte dos erros se concentra em grupos com atributos visuais próximos.

Uma estratégia de aumento de dados no *split* de treino, com transformações geométricas para elevar a 300 exemplos as classes `pap_em_divisao`, `pap_nao_definida` e `pap_reacional`, não superou o *checkpoint* principal: a acurácia caiu para 0,7171 e a macro-F1 para 0,6637, sugerindo que o aumento geométrico simples, no recorte avaliado, não foi suficiente para melhorar o equilíbrio entre classes.

5.3. Análise e Limitações

Os experimentos indicam que o DAM congelado pode atuar como extrator de representações visuais úteis para parte das classes anotacionais, especialmente aquelas com maior separação visual. Entretanto, as confusões observadas mostram que as representações não resolvem completamente a discriminação exigida pelo *dataset*. Os resultados foram calculados sobre classes anotacionais, e não sobre laudos diagnósticos independentes ou avaliação de especialistas. Por conta disso, os valores de acurácia, macro-F1 e *weighted-F1* não devem ser interpretados como desempenho diagnóstico. Somam-se a isso o desbalanceamento e o baixo suporte de algumas classes, que tornam suas métricas instáveis, e a validação da conversão CZI-DICOM limitada a uma única amostra.

Tabela 1. Resultados de validação dos classificadores sobre *embeddings* visuais do DAM.

Configuração	Acurácia	Macro-F1	Weighted-F1
Até 500 exemplos por classe	0,7050	0,6631	0,7009
Até 1.000 exemplos por classe	0,7245	0,6743	0,7242
Até 2.000 exemplos por classe	0,7274	0,6727	0,7230
1.000 por classe com aumento (classes raras)	0,7171	0,6637	0,7186

Tabela 2. Exemplos de métricas por classe no *checkpoint* principal.

Classe anotacional	Suporte	Precisão	Recall	F1
agnor_Cluster	300	0,9768	0,9833	0,9801
agnor_Satellite	300	0,9801	0,9867	0,9834
feulgen_sujeira	66	0,8889	0,7273	0,8000
pap_anormal	299	0,5099	0,6020	0,5521
pap_em_divisao	4	0,0000	0,0000	0,0000
pap_nao_definida	9	0,8000	0,4444	0,5714
pap_reacional	85	0,5000	0,3412	0,4056

6. Conclusão

Este trabalho desenvolveu e avaliou tecnicamente um protótipo computacional para apoio à análise regional de imagens citopatológicas, integrando recebimento e processamento de imagens, suporte a CZI, interoperabilidade com DICOM aplicada a WSI, segmentação interativa com SAM2, descrição visual localizada com DAM e uma interface web de demonstração. A arquitetura modular permitiu implementar os fluxos de *upload*, consulta de metadados, extração de *tiles*, segmentação, descrição, análise regional integrada e conversão de CZI para DICOM, incluindo a conversão bem-sucedida de uma amostra real. No eixo experimental, o melhor desempenho quantitativo foi obtido com o DAM congelado como extrator de *embeddings* visuais e uma cabeça MLP, alcançando acurácia de 0,7245 e macro-F1 de 0,6743 na validação. Esses resultados indicam utilidade computacional dos *embeddings* para classificação anotacional auxiliar, sem caracterizar validação clínica.

Como trabalhos futuros, propõe-se ampliar a validação da conversão CZI-DICOM com arquivos de diferentes origens e estruturas internas, integrar ao protótipo um visualizador DICOM associado a um servidor e avaliar as descrições geradas pelo DAM com especialistas em citopatologia. Também se pretende investigar novas arquiteturas de classificação, estratégias de balanceamento de classes e outras formas de adaptação sobre os *embeddings* visuais. Por fim, destaca-se a construção de bases multimodais que associem regiões segmentadas, máscaras, classes anotacionais e descrições especializadas revisadas por profissionais, possibilitando a adaptação e a avaliação mais objetiva de modelos voltados à citopatologia.

Referências

- [Alves 2017] Alves, J. M. (2017). The adequate model: software quality evaluation model for telemedicine and telehealth systems. Master's thesis, Universidade Federal de Santa Catarina (Programa de Pós-Graduação em Ciência da Computação). Disponível em:



Figura 3. Matriz de confusão normalizada do classificador principal (até 1.000 exemplos por classe).

<https://repositorio.ufsc.br/handle/123456789/186763>. Acesso em: 1 dez. 2024.

[Andreóli Petrolini et al. 2019] Andreóli Petrolini, V., Beckhauser, E., Savaris, A., Ines Meurer, M., von Wangenheim, A., and Krechel, D. (2019). Collaborative telepathology in a statewide telemedicine environment. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 684–689. IEEE Computer Society. Disponível em: <https://doi.ieeecomputersociety.org/10.1109/CBMS.2019.00139>. Acesso em: 1 dez. 2024.

[Asif et al. 2024] Asif, S., Wenhui, Y., ur Rehman, S., et al. (2024). Advancements and prospects of machine learning in medical diagnostics: Unveiling the future of diagnostic precision. *Archives of Computational Methods in Engineering*.

[Boaretto et al. 2024] Boaretto, N., e Silva, G. A. C., de Oliveira, F. A., and outros (2024). Câncer: Uma revisão integrativa por estudantes de medicina. *Revista Brasileira de*

Cancerologia, 9(2):e-20246402.

- [Brasil 2021] Brasil (2021). Câncer. Ministério da Saúde, Instituto Nacional de Câncer (INCA). Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer>. Acesso em: 9 jun. 2024.
- [Brundage et al. 2022] Brundage, D., Rosenthal, J., Carelli, R., Rand, S., Umeton, R., Loda, M., and Marchionni, L. (2022). Whole slide image to dicom conversion as event-driven cloud infrastructure. arXiv:2203.13888. Disponível em: <https://arxiv.org/abs/2203.13888>. Acesso em: 11 jun. 2026.
- [Carl Zeiss Microscopy 2026] Carl Zeiss Microscopy (2026). CZI image file format. Disponível em: <https://www.zeiss.com/microscopy/en/products/software/zeiss-zen/czi-image-file-format.html>. Acesso em: 11 jun. 2026.
- [CFM 2019] CFM (2019). Cfm regulamenta a transmissão de imagens patológicas e garante maior segurança ao diagnóstico de doenças. Conselho Federal de Medicina. Disponível em: <https://portal.cfm.org.br/noticias/cfm-regulamenta-a-transmissao-de-imagens-patologicas-e-garante-maior-seguranca-ao-diagnostico-de-doencas>. Acesso em: 1 dez. 2024.
- [DICOM Standards Committee 2026] DICOM Standards Committee (2026). DICOM standard: Current edition. National Electrical Manufacturers Association. Disponível em: <https://www.dicomstandard.org/current/>. Acesso em: 11 jun. 2026.
- [Evaristo et al. 2025] Evaristo, E., da Cruz, T., Ribeiro, R., Santos, L., and von Wangenheim, A. (2025). Systematic literature review on multimodal large language models applied to medical image data.
- [FEHOESP 2019] FEHOESP (2019). Resolução define telepatologia como especialidade médica. Federação dos Hospitais, Clínicas, Casas de Saúde, Laboratórios de Pesquisas e Análises Clínicas e Demais Estabelecimentos de Serviços de Saúde do Estado de São Paulo. Disponível em: <https://fehoesp360.org.br/sindsuzano/resolucao-define-telepatologia-como-especialidade-medica/>. Acesso em: 1 dez. 2024.
- [Fowler et al. 2024] Fowler, J., Maani, E., Dunton, C., et al. (2024). Cervical cancer. [Updated 2023 Nov 12]. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK431093/>. Acesso em: 1 dez. 2024.
- [Giarnieri and Scardapane 2023] Giarnieri, E. and Scardapane, S. (2023). Towards artificial intelligence applications in next generation cytopathology. *Biomedicines*, 11(8). Disponível em: <https://www.mdpi.com/2227-9059/11/8/2225>. Acesso em: 2 dez. 2024.
- [Goode et al. 2013] Goode, A., Gilbert, B., Harkes, J., Jukic, D., and Satyanarayanan, M. (2013). OpenSlide: A vendor-neutral software foundation for digital pathology. *Journal of Pathology Informatics*, 4(1):27. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3815078/>. Acesso em: 11 jul. 2026.
- [Goodfellow et al. 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. Disponível em: <https://www.deeplearningbook.org/>. Acesso em: 11 jun. 2026.

- [Hu et al. 2021] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). LoRA: Low-rank adaptation of large language models. arXiv:2106.09685. Disponível em: <https://arxiv.org/abs/2106.09685>. Acesso em: 11 jun. 2026.
- [Johns Hopkins Medicine 2024] Johns Hopkins Medicine (2024). Cytology. Disponível em: <https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/cytology>. Acesso em: 30 nov. 2024.
- [Kantsemlal 2024] Kantsemlal, A. (2024). How ai helps in whole slide image analysis for cancer detection. Mobidev Blog. Disponível em: <https://mobidev.biz/blog/ai-cancer-detection-approaches-challenges-opportunities>. Acesso em: 2 dez. 2024.
- [Kumar et al. 2020] Kumar, N., Gupta, R., and Gupta, S. (2020). Whole slide imaging (WSI) in pathology: Current perspectives and future directions. *Journal of Digital Imaging*, 33(4):1034–1040.
- [Lialin et al. 2023] Lialin, V., Deshpande, V., Yao, X., and Rumshisky, A. (2023). Scaling down to scale up: A guide to parameter-efficient fine-tuning. arXiv:2303.15647. Disponível em: <https://arxiv.org/abs/2303.15647>. Acesso em: 11 jun. 2026.
- [Lian et al. 2025] Lian, L., Ding, Y., Ge, Y., Liu, S., Mao, H., Li, B., Pavone, M., Liu, M.-Y., Darrell, T., Yala, A., and Cui, Y. (2025). Describe anything: Detailed localized image and video captioning. arXiv:2504.16072. Disponível em: <https://arxiv.org/abs/2504.16072>. Acesso em: 11 jun. 2026.
- [National Library of Medicine 2003] National Library of Medicine (2003). Collection development guidelines of the national library of medicine: Cytology. National Center for Biotechnology Information (NCBI). Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK518827/>. Acesso em: 30 nov. 2024.
- [Pangarkar 2022] Pangarkar, M. A. (2022). The Bethesda system for reporting cervical cytology. *Cytojournal*, 19:28.
- [PDQ Adult Treatment Editorial Board 2025] PDQ Adult Treatment Editorial Board (2025). Lip and oral cavity cancer treatment (PDQ): Patient version. National Cancer Institute, Bethesda, MD. Disponível em: <https://www.cancer.gov/types/head-and-neck/patient/adult/lip-mouth-treatment-pdq>. Acesso em: 11 jul. 2026.
- [Prodanov and Freitas 2013] Prodanov, C. C. and Freitas, E. C. d. (2013). *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. Feevale, Novo Hamburgo, 2 edition.
- [Ramirez Puebla et al. 2024] Ramirez Puebla, S. T., Mark Welch, J., and Borisly, G. (2024). Improved visualization of oral microbial consortia: Associated images. Zenodo. Disponível em: <https://zenodo.org/records/11042811>. Acesso em: 15 jun. 2026.
- [Ravi et al. 2024] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N.,

- Wu, C.-Y., Girshick, R., Dollár, P., and Feichtenhofer, C. (2024). Sam 2: Segment anything in images and videos. arXiv:2408.00714. Disponível em: <https://arxiv.org/abs/2408.00714>. Acesso em: 11 jun. 2026.
- [Santos et al. 2025] Santos, L. O., Ribeiro, R., Tiellet, B. Q., Silva, L. M. E., Onofre, F. B. d. M., Onofre, A. S. C., and von Wangenheim, A. (2025). Ufsc multi-cyto: Multi-stain cytopathology image dataset. Mendeley Data, V1. Disponível em: <https://data.mendeley.com/datasets/b8bygdxr7r/1>. Acesso em: 11 jun. 2026.
- [Sartori et al. 2014] Sartori, B., Hilleshein, K. D., and Ribeiro, O. D. (2014). Telepatologia. *Anais de Medicina*, 1(1):38. Disponível em: <https://periodicos.unoesc.edu.br/anaisdemedicina/article/view/4659>. Acesso em: 1 dez. 2024.
- [Scikit-learn Developers 2026] Scikit-learn Developers (2026). f1_score. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html. Acesso em: 11 jun. 2026.
- [Shafi and Parwani 2023] Shafi, S. and Parwani, A. (2023). Artificial intelligence in diagnostic pathology. *Diagnostic Pathology*, 18(109).
- [Stanford Pathology 2024] Stanford Pathology (2024). Cytopathology. Disponível em: <https://stanfordlab.com/anatomic-pathology/cytopathology.html>. Acesso em: 30 nov. 2024.
- [Vitória Matias et al. 2021] Vitória Matias, A., Atkinson Amorim, J. G., Buschetto Macarini, L. A., Cerentini, A., Casimiro Onofre, A. S., De Miranda Onofre, F. B., Daltoé, F. P., Stemmer, M. R., and von Wangenheim, A. (2021). What is the state of the art of computer vision-assisted cytology? a systematic literature review. *Computerized Medical Imaging and Graphics*, 91:101934. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0895611121000835>. Acesso em: 1 dez. 2024.
- [Vitória Matias et al. 2021] Vitória Matias, A., Cerentini, A., Buschetto Macarini, L. A., Atkinson Amorim, J. G., Perozzo Daltoé, F., and von Wangenheim, A. (2021). Segmentation, detection and classification of cell nuclei on oral cytology samples stained with papanicolaou. *SN Computer Science*, 2(4). Disponível em: <http://dx.doi.org/10.1007/s42979-021-00676-8>. Acesso em: 9 jun. 2024.
- [Zhang et al. 2023] Zhang, B., Shi, H., and Wang, H. (2023). Machine learning and ai in cancer prognosis, prediction, and treatment selection: A critical approach. *Journal of Multidisciplinary Healthcare*, 16:1779–1791. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10312208/>. Acesso em: 9 jun. 2024.
- [Zhuang et al. 2020] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. arXiv:1911.02685. Disponível em: <https://arxiv.org/abs/1911.02685>. Acesso em: 11 jun. 2026.

A. Pipeline Integrado de Análise Regional

A Figura 4 apresenta o fluxo completo do pipeline integrado de análise regional descrito na Seção 4.3.

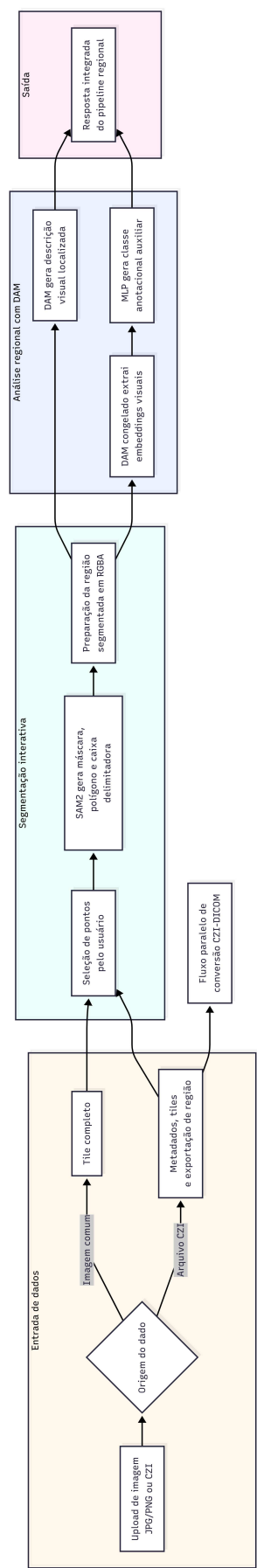


Figura 4. Pipeline integrado de análise regional.