

---

# **Modelo Híbrido CNN-Transformer com Descritores Fractais para Classificação de Imagens Histológicas**

---

**Dalmo Scalon Inacio**



UNIVERSIDADE FEDERAL DE UBERLÂNDIA  
FACULDADE DE COMPUTAÇÃO  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**Dalmo Scalon Inacio**

**Modelo Híbrido CNN-Transformer com  
Descritores Fractais para Classificação de  
Imagens Histológicas**

Trabalho de Conclusão de Curso apresentado  
à Faculdade de Computação da Universidade  
Federal de Uberlândia, Minas Gerais, como  
requisito exigido parcial à obtenção do grau de  
Bacharel em Sistemas de Informação.

Área de concentração: Sistemas de Informação

Orientador: Marcelo Zanchetta do Nascimento

Uberlândia - MG

2026

---

# Agradecimentos

Gostaria de agradecer primeiramente à minha família, em especial aos meus pais, Nelio e Alessandra, pelo apoio, incentivo e por sempre proporcionarem as condições necessárias para a conclusão desta etapa da minha formação.

Agradeço também aos meus irmãos e amigos, que estiveram presentes ao longo dessa jornada, oferecendo apoio e incentivo durante os momentos mais desafiadores.

Agradeço ao meu orientador, Prof. Dr. Marcelo Zanchetta do Nascimento, pela orientação, apoio e contribuições fundamentais para o desenvolvimento deste trabalho.

Por fim, agradeço à minha namorada, Mariana, pelo apoio, companheirismo e compreensão durante todo o período de desenvolvimento deste trabalho.

---

# Resumo

A análise de imagens histopatológicas desempenha um papel fundamental no diagnóstico de lesões, sendo frequentemente utilizada para identificar diferentes graus de displasia na cavidade oral. No entanto, a interpretação dessas imagens pode ser um processo complexo e sujeito à variabilidade entre especialistas. Nesse contexto, os sistemas de diagnóstico auxiliado por computador têm sido investigados como ferramentas de apoio à análise de imagens médicas. Este trabalho apresenta uma investigação do uso de modelo híbrido que explora *Vision Transformers* e redes neurais convolucionais para investigar representações baseadas em geometria fractal para a classificação de lesões histológicas de displasia. Os experimentos foram conduzidos utilizando o banco de dados Oral Epithelium DB, que contém imagens histopatológicas classificadas em diferentes graus de displasia. Foram avaliadas diferentes representações das imagens, incluindo imagens histológicas originais, representações sequenciais e representações baseadas em *Recurrence Plot* (Rec-Plot). Os modelos foram treinados e avaliados em diferentes cenários de classificação binária entre as classes saudável e displasias leves, moderadas e severas. Os resultados obtidos indicam que as diferentes representações de imagem influenciam diretamente o desempenho dos modelos de classificação, demonstrando o potencial do uso combinado de técnicas de aprendizado profundo e descritores baseados em geometria fractal para a análise de imagens histopatológicas.

**Palavras-chave:** Histopatologia, aprendizado profundo, geometria fractal, vision transformer, classificação de imagens.

---

## Lista de ilustrações

|                                                                                                                                                                                                                                                                              |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Fluxo de trabalho da histopatologia digital: digitalização de lâminas H&E, extração de patches e análise por redes neurais. . . . .                                                                                                                               | 14 |
| Figura 2 – Pipeline típico de um sistema de Diagnóstico Auxiliado por Computador (CAD) aplicado à análise de imagens histopatológicas (HE et al., 2022). . . . .                                                                                                             | 15 |
| Figura 3 – Arquitetura de um Vision Transformer (ViT). Adaptado de (DOSOVITSKIY et al., 2020). . . . .                                                                                                                                                                       | 19 |
| Figura 4 – Exemplos representativos das quatro classes do <i>Oral Epithelium DB</i> . . . . .                                                                                                                                                                                | 24 |
| Figura 5 – Representações SR geradas a partir dos descritores fractais das quatro classes do <i>Oral Epithelium DB</i> . . . . .                                                                                                                                             | 27 |
| Figura 6 – Representações RecPlot geradas a partir dos descritores fractais das quatro classes do <i>Oral Epithelium DB</i> . . . . .                                                                                                                                        | 28 |
| Figura 7 – Arquitetura híbrida proposta para a etapa de aprendizagem profunda, composta pelos três caminhos paralelos (EfficientNet, ViT Adaptado e CNN Simplificada), seguidos pela concatenação dos vetores de características e classificação por camadas densas. . . . . | 30 |

---

## Lista de abreviaturas e siglas

|          |                                      |
|----------|--------------------------------------|
| ANN      | Artificial Neural Network            |
| CAD      | Diagnóstico Auxiliado por Computador |
| CNN      | Convolutional Neural Network         |
| DL       | Deep Learning                        |
| ML       | Machine Learning                     |
| ViT      | Vision Transformer                   |
| SR       | Sequential Representation            |
| RecPlot  | Recurrence Plot                      |
| ROI      | Region of Interest                   |
| H&E      | Hematoxilina e Eosina                |
| GPU      | Graphics Processing Unit             |
| CPU      | Central Processing Unit              |
| F1-score | F1 Score                             |
| TP       | True Positive                        |
| TN       | True Negative                        |
| FP       | False Positive                       |
| FN       | False Negative                       |
| SVM      | Support Vector Machine               |
| FD       | Fractal Dimension                    |

|     |                                     |
|-----|-------------------------------------|
| AUC | Area Under the Curve                |
| RF  | Random Forest                       |
| WSI | Whole Slide Imaging                 |
| LBP | Local Binary Pattern                |
| XAI | Explainable Artificial Intelligence |
| NLP | Natural Language Processing         |
| UFU | Universidade Federal de Uberlândia  |

---

# Sumário

|            |                                                                  |           |
|------------|------------------------------------------------------------------|-----------|
| <b>1</b>   | <b>INTRODUÇÃO . . . . .</b>                                      | <b>9</b>  |
| <b>1.1</b> | <b>Justificativa . . . . .</b>                                   | <b>10</b> |
| <b>1.2</b> | <b>Objetivos . . . . .</b>                                       | <b>11</b> |
| 1.2.1      | Objetivo Geral . . . . .                                         | 11        |
| 1.2.2      | Objetivos Específicos . . . . .                                  | 11        |
| <b>1.3</b> | <b>Organização da Monografia . . . . .</b>                       | <b>12</b> |
| <b>2</b>   | <b>FUNDAMENTAÇÃO TEÓRICA . . . . .</b>                           | <b>13</b> |
| <b>2.1</b> | <b>Histopatologia . . . . .</b>                                  | <b>13</b> |
| <b>2.2</b> | <b>Sistema de Diagnóstico Auxiliado por Computador . . . . .</b> | <b>14</b> |
| <b>2.3</b> | <b>Geometria Fractal . . . . .</b>                               | <b>16</b> |
| <b>2.4</b> | <b>Redes Neurais Artificiais . . . . .</b>                       | <b>17</b> |
| <b>2.5</b> | <b>Redes Neurais Convolucionais (CNNs) . . . . .</b>             | <b>18</b> |
| <b>2.6</b> | <b>Transformers . . . . .</b>                                    | <b>18</b> |
| <b>2.7</b> | <b>Trabalhos Relacionados . . . . .</b>                          | <b>20</b> |
| <b>3</b>   | <b>METODOLOGIA . . . . .</b>                                     | <b>23</b> |
| <b>3.1</b> | <b>Introdução . . . . .</b>                                      | <b>23</b> |
| <b>3.2</b> | <b>Banco de Imagens . . . . .</b>                                | <b>24</b> |
| <b>3.3</b> | <b>Extração de Descritores Fractais . . . . .</b>                | <b>25</b> |
| 3.3.1      | Métrica de distância e análise multidimensional . . . . .        | 25        |
| 3.3.2      | Descritores extraídos . . . . .                                  | 25        |
| 3.3.3      | Atributos locais e globais . . . . .                             | 26        |
| <b>3.4</b> | <b>Representações 2D de Descritores Fractais . . . . .</b>       | <b>26</b> |
| 3.4.1      | Reorganização Sequencial . . . . .                               | 27        |
| 3.4.2      | Gráfico de Recorrência (GR) . . . . .                            | 27        |
| <b>3.5</b> | <b>Modelo de Aprendizagem Profunda . . . . .</b>                 | <b>28</b> |
| 3.5.1      | Caminho i): EfficientNet para Extração de Detalhes . . . . .     | 29        |



|            |                                                                      |           |
|------------|----------------------------------------------------------------------|-----------|
| 3.5.2      | Caminho ii): ViT Adaptado para Captura de Relações Globais . . . . . | 29        |
| 3.5.3      | Caminho iii): CNN Simplificada . . . . .                             | 29        |
| 3.5.4      | Integração e Classificação . . . . .                                 | 30        |
| 3.5.5      | Variações Estruturais . . . . .                                      | 30        |
| <b>3.6</b> | <b>Configuração Experimental . . . . .</b>                           | <b>31</b> |
| 3.6.1      | Divisão dos Dados . . . . .                                          | 31        |
| 3.6.2      | Parâmetros de Treinamento . . . . .                                  | 31        |
| <b>3.7</b> | <b>Métricas de Avaliação . . . . .</b>                               | <b>32</b> |
| <b>4</b>   | <b>AVALIAÇÃO DOS RESULTADOS . . . . .</b>                            | <b>34</b> |
| 4.1        | Organização dos Experimentos . . . . .                               | 34        |
| 4.2        | Avaliação com Imagens Histológicas Originais . . . . .               | 34        |
| 4.3        | Análise com Reorganização Sequencial (SR) . . . . .                  | 35        |
| 4.4        | Estudo com Gráficos de Recorrência . . . . .                         | 36        |
| 4.5        | Análise de Ablação . . . . .                                         | 37        |
| 4.5.1      | Ablação do Módulo <i>Bottleneck</i> . . . . .                        | 38        |
| 4.5.2      | Ablação do Módulo <i>Neck</i> . . . . .                              | 38        |
| 4.5.3      | Ablação do Módulo <i>Transformer</i> . . . . .                       | 39        |
| 4.6        | Comparação com o Estado da Arte . . . . .                            | 39        |
| <b>5</b>   | <b>CONCLUSÃO . . . . .</b>                                           | <b>41</b> |
| 5.1        | Principais Contribuições . . . . .                                   | 41        |
|            | <b>REFERÊNCIAS . . . . .</b>                                         | <b>43</b> |

---

# Introdução

Lesões histológicas são alterações microscópicas que ocorrem nos tecidos biológicos devido a processos patológicos, como inflamação, degeneração, neoplasia (crescimento anormal de células, como no câncer) ou processos infecciosos. Elas são analisadas através da histopatologia, que utiliza cortes de tecidos corados com substâncias específicas (como Hematoxilina e Eosina – H&E) para permitir a observação ao microscópio. Diante disso, esse trabalho propõe o uso de inteligência artificial para melhorar a precisão e precocidade dos diagnósticos, em busca de contribuir para melhores prognósticos dos pacientes.

A identificação precoce de alterações celulares permite intervenções menos invasivas e melhora significativamente o prognóstico dos pacientes. Por exemplo, no câncer de mama, o diagnóstico em estágios iniciais eleva a taxa de sobrevivência para 98% (PAIS, 2025). Além disso, a detecção precoce de lesões histológicas possibilita tratamentos mais conservadores, reduzindo a necessidade de procedimentos agressivos e melhorando a qualidade de vida dos pacientes. Neste contexto, a aplicação de tecnologias avançadas, como os *Vision Transformers* (ViTs) e as representações de geometria fractal, surge como uma abordagem promissora para aprimorar a precisão e a eficiência na classificação de imagens histológicas.

Diferentes abordagens têm sido exploradas para aprimorar a classificação de imagens histológicas. Faria (2022) investigou o uso de Inteligência Artificial Explicável (XAI) para tornar os modelos de diagnóstico mais interpretáveis por especialistas, enquanto Carvalho (2023) combinou redes neurais convolucionais (CNNs) e descritores fractais para melhorar a análise de displasia oral. Apesar dos avanços, as CNNs possuem limitações na captura de relações espaciais globais, o que justifica o estudo de *Vision Transformers* (ViTs) como alternativa. Assim, este trabalho propõe a combinação de ViTs e geometria fractal para a classificação de lesões histológicas, buscando superar desafios das abordagens anteriores.

Dito isso, o objetivo geral deste trabalho é investigar a eficácia da combinação de *Vision Transformers* (ViTs) e representações baseadas em Geometria Fractal na classificação de lesões histológicas. Dentro desse escopo, alguns objetivos específicos incluem a implementação e o treinamento de modelos ViTs para a classificação de imagens histo-

lógicas combinados à extração de características fractais das imagens, analisando como essas características se correlacionam com os padrões das lesões. Além disso, este estudo propõe comparar o desempenho da abordagem ViTs + geometria fractal com métodos convencionais, como as redes neurais convolucionais (CNNs).

O desenvolvimento deste trabalho envolve a aplicação de redes neurais artificiais, com foco em redes neurais convolucionais (CNNs) e *Transformers*, incluindo *Vision Transformers* (ViTs). Para a implementação dos modelos, será utilizado Python (FOUNDATION, 2025) e as bibliotecas amplamente empregadas em aprendizado profundo, Keras (CHOLLET, 2015) e PyTorch (META AI, 2016), garantindo flexibilidade no treinamento e avaliação dos modelos. O pré-processamento das imagens e a extração de características fractais serão realizados utilizando ferramentas especializadas, como Matlab (MATHWORKS, 2024). Entre os principais algoritmos e técnicas utilizadas estão a descida do gradiente (Gradient Descent), operações de convolução e técnicas de processamento de imagens aplicadas às representações geradas. Embora as Redes Neurais Convolucionais (CNNs) sejam eficazes na extração de características locais, essas arquiteturas possuem um viés local, o que pode dificultar a modelagem direta de dependências espaciais de longo alcance, uma vez que a operação de convolução é limitada pelo tamanho do campo receptivo (KHAN et al., 2022; DOSOVITSKIY et al., 2020).

Nesse contexto, os descritores baseados em Geometria Fractal mostram-se relevantes por sua capacidade de quantificar a complexidade, irregularidade e auto-similaridade das estruturas biológicas, características frequentemente associadas a alterações patológicas nos tecidos. No entanto, parte da literatura concentra-se na utilização de descritores fractais em conjunto com classificadores tradicionais ou CNNs, sendo relativamente limitada a exploração de sua integração com arquiteturas mais recentes, como modelos baseados em *Transformers*, a exemplo dos *Vision Transformers* (ViTs).

Assim, com base nos trabalhos analisados, observa-se que a integração entre representações baseadas em geometria fractal e modelos baseados em *Transformers* ainda é pouco explorada, bem como há uma quantidade limitada de estudos que investigam o impacto dessas representações na capacidade dos modelos de capturar padrões estruturais complexos em imagens histológicas.

## 1.1 Justificativa

A análise de imagens histológicas desempenha um papel fundamental no diagnóstico de diversas patologias, podendo, no entanto, ser influenciada por fatores como a subjetividade inerente à interpretação humana e a alta demanda por especialistas qualificados. Nesse contexto, técnicas de inteligência artificial, especialmente aquelas baseadas em redes neurais, têm sido amplamente exploradas como ferramentas de apoio ao diagnóstico, apresentando resultados promissores.

Entretanto, abordagens tradicionais baseadas em redes neurais convolucionais (CNNs) podem apresentar limitações na captura de relações espaciais globais presentes nas imagens. Por outro lado, modelos baseados em arquiteturas *Transformer*, como os *Vision Transformers* (ViTs), têm demonstrado maior capacidade de modelar dependências de longo alcance. Adicionalmente, descritores baseados em geometria fractal permitem quantificar a complexidade estrutural e a irregularidade dos tecidos, características relevantes para a identificação de lesões.

Apesar desses avanços, ainda são limitadas, na literatura consultada, as investigações sobre a utilização de representações baseadas em geometria fractal em conjunto com modelos como os *Vision Transformers*, bem como sobre o impacto dessa combinação no desempenho da classificação de lesões histológicas, quando comparada a abordagens tradicionais baseadas em imagens originais e CNNs.

Dessa forma, este trabalho busca investigar essa questão, avaliando o impacto da combinação entre representações fractais e arquiteturas baseadas em *Transformers* no desempenho da classificação de lesões histológicas.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

O objetivo geral deste trabalho consiste em investigar o uso de descritores baseados em geometria fractal associados a modelos de aprendizado profundo para a classificação de imagens histológicas do epitélio oral, explorando diferentes formas de representação dessas características e sua integração em arquiteturas híbridas de redes neurais.

### 1.2.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

1. Extrair descritores baseados em geometria fractal a partir de imagens histológicas do conjunto de dados *Oral Epithelium DB*;
2. Investigar diferentes estratégias de representação desses descritores, incluindo reorganização sequencial (SR) e gráficos de recorrência (RecPlot);
3. Desenvolver uma arquitetura híbrida baseada em múltiplos ramos de extração de características, combinando redes convolucionais e mecanismos inspirados em *Transformers*;
4. Avaliar o desempenho do modelo na tarefa de classificação binária entre diferentes graus de displasia epitelial oral;

5. Realizar uma análise de ablação com o objetivo de investigar a contribuição individual dos diferentes componentes da arquitetura proposta.

## 1.3 Organização da Monografia

Esta monografia está organizada em quatro capítulos, descritos a seguir:

**Capítulo 1 – Introdução:** Apresenta o contexto do problema, os objetivos do trabalho, a justificativa da pesquisa e a organização do documento.

**Capítulo 2 – Fundamentação Teórica e Trabalhos Correlatos:** Aborda os principais conceitos relacionados ao tema, incluindo fundamentos sobre histopatologia, aprendizado de máquina, *Vision Transformers* e geometria fractal, além da revisão de trabalhos relacionados presentes na literatura.

**Capítulo 3 – Metodologia:** Descreve as etapas do método de pesquisa adotado, incluindo a descrição do conjunto de dados utilizado, os procedimentos de extração de características fractais, a geração das representações utilizadas nos experimentos e a arquitetura dos modelos empregados.

**Capítulo 4 – Avaliação dos Resultados:** Apresenta os experimentos realizados e a análise dos resultados obtidos, incluindo a comparação entre diferentes representações das imagens, bem como a análise de ablação para avaliar a contribuição dos componentes da arquitetura proposta.

**Capítulo 5 – Conclusão:** Apresenta as considerações finais do trabalho, destacando as principais contribuições alcançadas, limitações identificadas e possíveis direções para trabalhos futuros.

---

## Fundamentação Teórica

Este capítulo dedica-se à apresentação dos principais conceitos teóricos que fundamentam o domínio da pesquisa, bem como das técnicas mais relevantes aplicadas na área. Ademais, são discutidos os trabalhos relacionados que propõem soluções voltadas ao suporte de especialistas na melhoria do diagnóstico das lesões investigadas.

### 2.1 Histopatologia

A histopatologia fundamenta-se no estudo das alterações estruturais em tecidos biológicos decorrentes de processos patológicos, como inflamações, neoplasias e degenerações. O padrão-ouro para essa análise reside na observação microscópica de cortes histológicos, sendo a coloração por Hematoxilina e Eosina (H&E) a técnica mais difundida na rotina clínica. Enquanto a hematoxilina cora os núcleos celulares em tons de azul e roxo, a eosina destaca o citoplasma e a matriz extracelular em tons de rosa, permitindo a diferenciação morfológica essencial para o diagnóstico (BENTAIEB; HAMARNEH, 2018).

Historicamente, essa avaliação é realizada de forma manual por patologistas através da microscopia óptica convencional. Contudo, estudos recentes reforçam que esse método é suscetível a uma significativa variabilidade interobservador (variabilidade entre diferentes profissionais) e intraobservador (variabilidade de um mesmo profissional em análises repetidas), o que pode comprometer a reprodutibilidade dos diagnósticos, especialmente em casos de graduação de tumores complexos (MULLIQI, 2025). Além da subjetividade intrínseca à interpretação humana, a fadiga e a carga de trabalho do especialista são fatores que podem influenciar a precisão da análise manual (NIELSEN; GEORGENSEN et al., 2022).

Nesse cenário, a Histopatologia Digital surge como uma evolução tecnológica, impulsionada pelo desenvolvimento de scanners de lâmina inteira (*Whole Slide Imaging* - WSI). A digitalização das lâminas permite não apenas o armazenamento e o compartilhamento de dados, mas também a aplicação de algoritmos de inteligência artificial (IA) para a automação de tarefas exaustivas, como a contagem de figuras mitóticas e a quantificação de

linfócitos infiltrantes de tumor (do inglês, *tumor-infiltrating lymphocytes* – sTILs), que são células do sistema imunológico associadas à resposta tumoral (THAGAARD; STOVGAARD et al., 2021). Essas abordagens computacionais visam padronizar a análise tecidual, oferecendo métricas quantitativas que auxiliam na redução da subjetividade e na melhoria da estratificação prognóstica dos pacientes (MULLIQI, 2025; NIELSEN; GEORGENSEN et al., 2022).

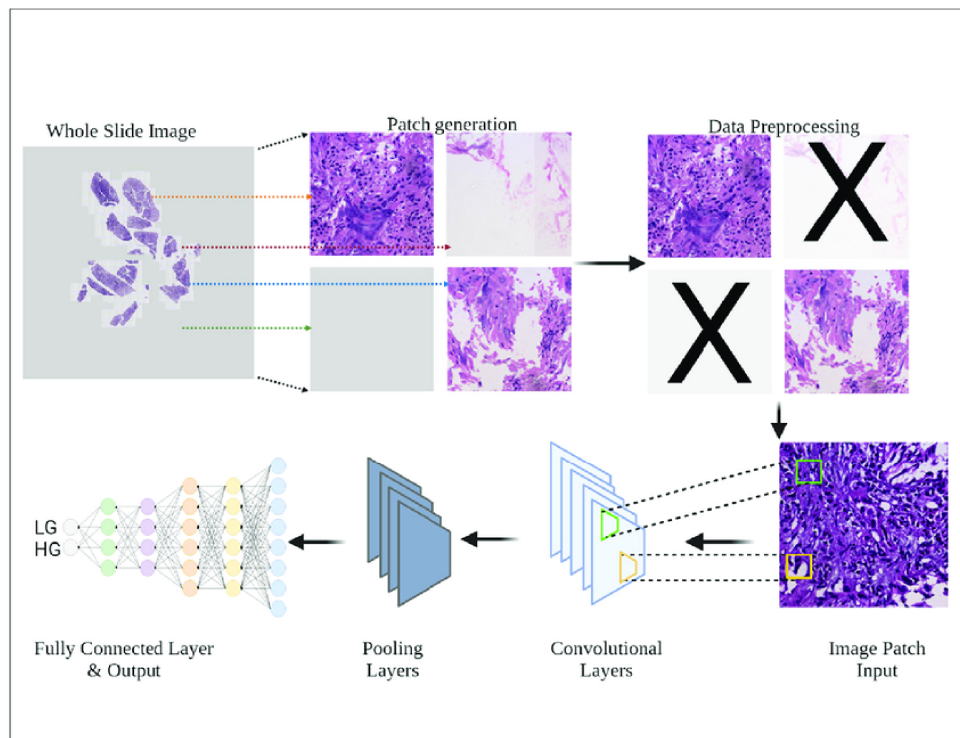


Figura 1 – Fluxo de trabalho da histopatologia digital: digitalização de lâminas H&E, extração de patches e análise por redes neurais.

Fonte: Adaptado de (MUNDHADA et al., 2023).

Como ilustrado na Figura 1, o fluxo de trabalho da histopatologia digital envolve inicialmente a digitalização das lâminas histológicas, seguida pela extração de regiões de interesse (*patches*) a partir das imagens de lâmina inteira. Esses fragmentos são então submetidos a etapas de pré-processamento e analisados por modelos de aprendizado profundo, como redes neurais convolucionais, com o objetivo de realizar tarefas como classificação e detecção de padrões patológicos.

## 2.2 Sistema de Diagnóstico Auxiliado por Computador

O sistema de Diagnóstico Auxiliado por Computador (*Computer-Aided Diagnosis* – CAD) engloba técnicas computacionais destinadas a apoiar médicos e patologistas na detecção e interpretação de anormalidades em imagens clínicas, incluindo aquelas de mi-

croscopia óptica. Esses sistemas complementam, sem substituir, a expertise humana, atuando como segunda opinião para mitigar subjetividade, elevar a reprodutibilidade e aprimorar a consistência diagnóstica (GURCAN et al., 2009).

Tipicamente, um sistema CAD para imagens histológicas segue um *pipeline* estruturado em cinco etapas principais: (1) aquisição e digitalização, (2) pré-processamento, (3) segmentação, (4) extração de atributos e (5) classificação (GURCAN et al., 2009). A aquisição ocorre via microscópios ou escâneres de lâminas inteiras (*Whole Slide Imaging* – WSI), gerando imagens de alta resolução. Como ilustrado na Figura 2, o processo de CAD envolve etapas sequenciais que vão desde a aquisição das imagens até a classificação final.

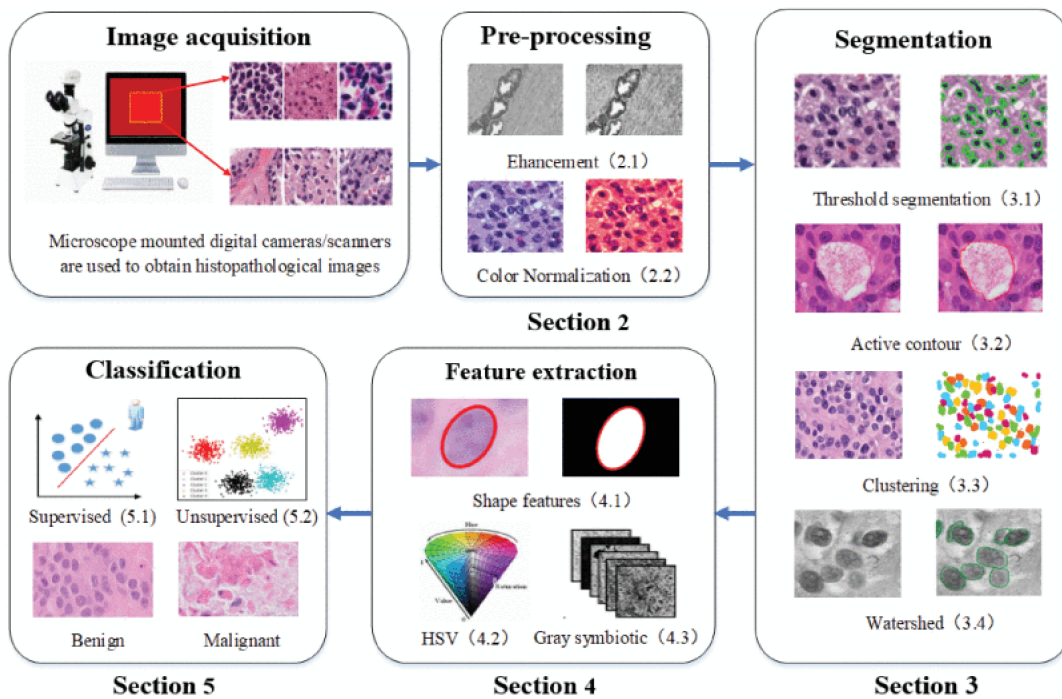


Figura 2 – Pipeline típico de um sistema de Diagnóstico Auxiliado por Computador (CAD) aplicado à análise de imagens histopatológicas (HE et al., 2022).

No pré-processamento, aplicam-se correções para ruído, iluminação e normalização de cor, preparando os dados para análises subsequentes. A segmentação, então, delimita regiões de interesse, como núcleos celulares ou padrões teciduais, empregando métodos como limiarização, *watershed* ou redes neurais convolucionais (GURCAN et al., 2009).

A extração de atributos converte características visuais em vetores numéricos, abrangendo descritores morfológicos (forma, tamanho), texturais (Haralick, Gabor) e fractais (dimensão fractal, lacunarity), ideais para capturar irregularidades em tecidos patológicos (GURCAN et al., 2009). Esses atributos alimentam classificadores de aprendizado supervisionado, como SVMs, redes neurais ou ensembles, cujos desempenhos dependem criticamente da qualidade e discriminatividade dos descritores extraídos.

A adoção de CAD em histopatologia digital tem se expandido, impulsionada pela



capacidade de processar grandes conjunto de dados e minimizar variabilidades interobservador. Integrações recentes de deep learning com descritores manuais, incluindo fractais, elevam a precisão em tarefas como graduação tumoral (ROBERTO, 2021; LITJENS et al., 2017).

## 2.3 Geometria Fractal

A geometria fractal, introduzida por Benoit Mandelbrot (MANDELBROT, 1983), descreve estruturas complexas e irregulares que não podem ser adequadamente representadas pela geometria euclidiana tradicional. Fractais são caracterizados por sua auto-similaridade estatística, exibindo padrões que se repetem em múltiplas escalas e mantendo invariância de escala e rugosidade (BARNSELEY, 1993).

Em imagens histológicas, estruturas como membranas celulares e aglomerados tumorais apresentam propriedades fractais que refletem o estado patológico do tecido (LOPES; BETROUNI, 2009). Para quantificar essa complexidade, utilizam-se descritores fractais capazes de capturar diferentes aspectos estruturais das imagens sob diferentes métricas de distância, como as distâncias de Minkowski (Chessboard), Euclidiana e Manhattan.

A Dimensão Fractal (DF) é uma das principais métricas utilizadas na análise de estruturas complexas, sendo frequentemente estimada pelo método de *box-counting*. Nesse método, uma grade de tamanho  $r$  é sobreposta à imagem, e contabiliza-se o número de caixas  $N(r)$  que interceptam a estrutura. Essa medida é obtida a partir da inclinação da reta no gráfico log-log de  $N(r)$  em função de  $1/r$ , sendo definida conforme a Equação 1:

$$D = \frac{\log N(r)}{\log(1/r)}. \quad (1)$$

Após o cálculo para diferentes valores de  $r$ , é possível construir o gráfico  $\log(N(r)) \times \log(1/r)$ , cujo coeficiente angular corresponde ao valor da dimensão fractal.

A Lacunaridade (LAC) é uma medida complementar à Dimensão Fractal (DF), utilizada para avaliar como o espaço de um fractal é preenchido. Enquanto a DF quantifica a complexidade estrutural, a LAC fornece informações sobre a heterogeneidade e a distribuição dos vazios na estrutura. Essa métrica é calculada a partir de uma matriz de probabilidade  $P(m, L)$ , na qual  $L$  representa o tamanho da janela de análise (escala), enquanto  $m$  corresponde ao número de pixels ocupados (massa) dentro de cada janela. Nessa matriz,  $P(m, L)$  indica a probabilidade de ocorrência de janelas de tamanho  $L$  contendo  $m$  elementos.

A partir dessa matriz, são considerados os momentos estatísticos de primeira e segunda ordem, correspondentes à média e ao segundo momento, respectivamente, definidos pelas Equações 2 e 3.

$$\mu(L) = \sum_{m=1}^N m P(m, L) \quad (2)$$

$$\mu^2(L) = \sum_{m=1}^N m^2 P(m, L) \quad (3)$$

A partir desses valores, a lacunaridade em uma escala  $L$ , denotada por  $\lambda(L)$ , é definida conforme a Equação 4:

$$\lambda(L) = \frac{\mu^2(L) - (\mu(L))^2}{(\mu(L))^2} \quad (4)$$

Assim como na análise por DF, é possível extrair atributos globais a partir da LAC. Esses atributos são obtidos por meio da curva  $\log(L) \times \log(\lambda(L))$ , da qual podem ser derivados descritores como a área sob a curva, a obliquidade, a razão de área e a escala correspondente ao ponto de máximo.

Além disso, a Percolação (PERC) é um descritor relacionado à conectividade estrutural. Neste trabalho, a percolação é tratada de forma estatística, verificando-se a probabilidade de formação de um aglomerado que atravessa a estrutura analisada. Define-se que uma região está em estado de percolação se a ocupação de seus elementos supera o limiar crítico de 0,59275. A métrica final expressa a fração de janelas de análise que atingiram esse estado para cada escala  $r$ :

$$PERC(r) = \frac{\text{Número de caixas percolantes}}{\text{Total de caixas}} \quad (5)$$

Dessa forma, os descritores fractais utilizados, DF, LAC e PERC permitem capturar diferentes aspectos estruturais das imagens, incluindo complexidade, heterogeneidade e conectividade do tecido analisado.

## 2.4 Redes Neurais Artificiais

As redes neurais artificiais (RNAs) são modelos computacionais inspirados na estrutura do sistema nervoso biológico, compostos por unidades denominadas neurônios artificiais, interconectados em camadas. Esses modelos são capazes de aprender a realizar tarefas a partir de exemplos, sendo amplamente utilizados em problemas de classificação, regressão e reconhecimento de padrões (HAYKIN, 2009).

A aplicação de RNAs em imagens médicas tem se mostrado especialmente relevante na análise histopatológica, permitindo identificar características morfológicas complexas associadas a alterações patológicas. Gurcan et al. (2009), ao revisar o estado da arte na análise de imagens histológicas, destacam o uso de redes neurais como ferramentas promissoras para o diagnóstico automatizado, principalmente quando combinadas com métodos robustos de extração de características.

Com a evolução dos dados e da capacidade computacional, surgiram arquiteturas especializadas, como as Redes Neurais Convolucionais (CNNs), e, mais recentemente, os Transformers, os quais serão abordados nas seções seguintes.

## 2.5 Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais (CNNs) são um tipo especializado de arquitetura de rede neural artificial, amplamente utilizadas em tarefas de processamento de imagem. Inspiradas na organização do córtex visual dos seres vivos, as CNNs operam por meio da aplicação de filtros convolucionais que extraem automaticamente características relevantes de uma imagem, como bordas, texturas e formas (GURCAN et al., 2009).

Uma CNN é composta por camadas convolucionais, camadas de *pooling* (subamostragem) e camadas totalmente conectadas. A sequência dessas operações permite a criação de representações hierárquicas dos dados de entrada, facilitando a identificação de padrões complexos em imagens médicas, como as obtidas por microscopia digital. O processo de aprendizado é realizado por meio do ajuste dos pesos dos filtros, com base em dados rotulados e algoritmos de otimização, como a descida do gradiente.

No contexto da histopatologia digital, CNNs têm sido amplamente empregadas em tarefas de segmentação de núcleos celulares, detecção de anomalias e classificação de tecidos saudáveis ou lesionados. Nos trabalhos de (ROBERTO, 2021) e (CARVALHO, 2024), CNNs foram utilizadas como parte de comitês de classificadores, isto é, abordagens que combinam múltiplos modelos de classificação com o objetivo de melhorar o desempenho final, tanto isoladamente quanto integradas com descritores fractais. Os resultados obtidos demonstraram alta acurácia na classificação de imagens histológicas, reforçando o papel das CNNs como ferramenta robusta no apoio ao diagnóstico automatizado.

Apesar de sua eficácia, estudos recentes apontam limitações das CNNs na modelagem de dependências globais em imagens. Nesse contexto, surgem os *Transformers* como arquiteturas mais flexíveis, capazes de capturar relações espaciais de longo alcance por meio de mecanismos de autoatenção, conforme proposto por (DOSOVITSKIY et al., 2020) e discutido em (KHAN et al., 2022). Essa nova abordagem será explorada na seção seguinte.

## 2.6 Transformers

Os *Transformers* são modelos de aprendizado profundo originalmente propostos por (VASWANI et al., 2017) para tarefas de processamento de linguagem natural (NLP). Diferentemente de arquiteturas baseadas em convoluções ou recorrências, os *Transformers* utilizam mecanismos de autoatenção (*self-attention*) para capturar dependências de

longo alcance entre os elementos de entrada, permitindo o processamento paralelo e a modelagem de relações contextuais globais.

A arquitetura *Transformers* é composta, em sua forma clássica, por camadas de atenção multi-cabeça (*multi-head attention*) e camadas *feed-forward* posicionais, organizadas em blocos empilhados. Esses componentes permitem que o modelo aprenda representações em diferentes níveis de abstração, destacando relações relevantes entre os elementos da entrada.

No contexto da visão computacional, os *Transformers* foram adaptados por (DOSOVITSKIY et al., 2020) na forma dos *Vision Transformers* (ViTs), nos quais as imagens são divididas em pequenas regiões denominadas *patches*. Esses *patches* são linearizados e tratados como uma sequência de entrada, de maneira análoga ao processamento de tokens em linguagem natural. Em seguida, são processados por camadas do tipo *Transformer Encoder*, que utilizam mecanismos de autoatenção para capturar relações espaciais globais entre diferentes regiões da imagem.

A Figura 3 ilustra a arquitetura geral de um *Vision Transformer*, destacando as etapas de divisão da imagem em *patches*, projeção linear e processamento por blocos de atenção.

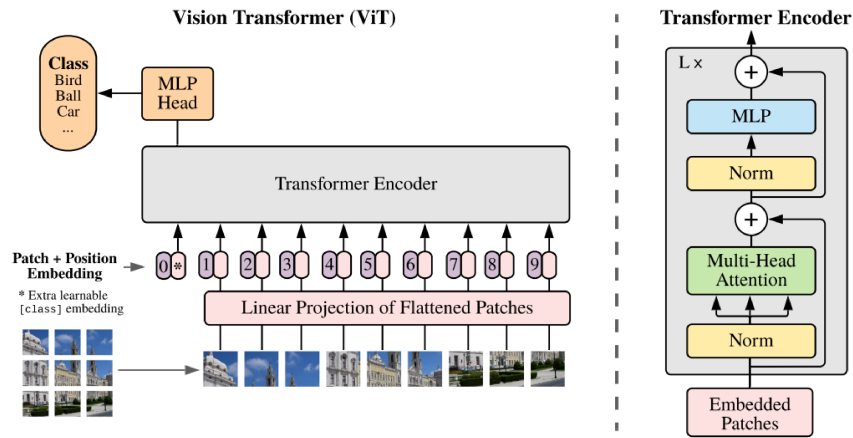


Figura 3 – Arquitetura de um Vision Transformer (ViT). Adaptado de (DOSOVITSKIY et al., 2020).

Apesar do alto desempenho dos ViTs em diversas tarefas de classificação de imagens, sua aplicação em cenários com bases de dados reduzidas pode ser limitada, devido à alta complexidade do modelo e à necessidade de grandes volumes de dados para treinamento eficiente.

Nesse contexto, o presente trabalho adota uma abordagem inspirada na filosofia dos *Vision Transformers*, porém com simplificações arquiteturais. Em vez de utilizar explicitamente mecanismos de autoatenção e *embeddings* posicionais, a imagem é dividida em regiões menores, simulando o conceito de *patches*, e processada por uma sequência de camadas convolucionais.

Essa estratégia permite preservar o processamento localizado das regiões da imagem, ao mesmo tempo em que reduz o custo computacional e torna o modelo mais adequado para cenários com conjuntos de dados limitados, como no caso das imagens histológicas utilizadas neste trabalho.

## 2.7 Trabalhos Relacionados

Roberto (2021) apresentou uma abordagem baseada na associação entre descritores fractais e aprendizado profundo para a classificação de imagens histológicas. Inicialmente, foram extraídos 300 atributos fractais das imagens, os quais foram reorganizados em diferentes representações bidimensionais para compor imagens RGB. Essas representações foram utilizadas como entrada para uma rede neural convolucional (*CNN*), enquanto outra *CNN* foi alimentada com as imagens histológicas originais. Os resultados das redes foram posteriormente combinados em um comitê de classificadores juntamente com classificadores baseados em atributos manuais. Os experimentos foram conduzidos em diferentes conjuntos de dados histológicos, alcançando acurácias entre 88,45% e 99,77%, demonstrando que a associação entre descritores fractais e aprendizado profundo pode melhorar o desempenho na classificação de tecidos histológicos.

Em (FARIA, 2022), o autor desenvolveu uma metodologia para a classificação multi-classe de imagens histológicas de linfomas e displasias epiteliais orais, com foco em Inteligência Artificial Explicável (XAI). A abordagem fundamenta-se na extração de atributos interpretáveis (características cuja representação pode ser compreendida por especialistas, como forma, tamanho e textura), divididos em descritores morfológicos (relacionados à forma dos núcleos celulares) e não-morfológicos (baseados na intensidade de cor e brilho). Esses atributos alimentam classificadores como *Gradient Boosting Decision Tree* (GBDT) e *Support Vector Machine* (SVM). Além da classificação, o trabalho propõe um estimador de confiança para categorizar as predições e utiliza os métodos SHAP e *Anchor*s para fornecer justificativas visuais e quantitativas aos diagnósticos. Os resultados alcançaram acurácias de 93,7% para linfomas e 91,8% para displasias orais, destacando a viabilidade de sistemas CAD mais transparentes para o especialista médico.

(CARVALHO, 2024) propôs um sistema de classificação para displasia epitelial oral baseado na utilização de descritores derivados da geometria fractal combinados com comitês de classificadores. Nesse estudo, as imagens histológicas foram processadas para extrair atributos fractais, incluindo a dimensão fractal e a *lacunarity*, os quais foram utilizados para gerar representações capazes de capturar características estruturais do tecido epitelial. A partir dessas representações, foram avaliados diferentes classificadores, como redes neurais artificiais e máquinas de vetores de suporte (*SVM*), bem como estratégias de comitê baseadas em regras de combinação, como soma, média e votação. Os resultados obtidos demonstraram que a utilização de comitês de classificadores pode melhorar

o desempenho da tarefa de classificação, alcançando acurácia de até 95,6% no conjunto de dados *Oral Dysplasia*. Além disso, o trabalho fornece implementações para a extração dos descritores fractais, que serviram como base metodológica para a geração das representações fractais utilizadas no presente estudo.

O trabalho de (DIAS, 2025) investigou a aplicação de modelos de aprendizado profundo para a classificação de imagens histológicas de displasia oral utilizando diretamente as imagens originais do conjunto de dados. O estudo implementou e avaliou uma arquitetura baseada na combinação de redes neurais convolucionais e *Vision Transformers* (ViT), explorando a capacidade desses modelos em capturar padrões estruturais presentes nas imagens histológicas. Os experimentos demonstraram que a utilização conjunta dessas arquiteturas pode contribuir para a identificação de diferentes níveis de displasia, evidenciando o potencial do aprendizado profundo no desenvolvimento de sistemas de diagnóstico assistido por computador em histopatologia.

Apesar dos trabalhos apresentarem avanços relevantes, observa-se que diferentes abordagens são adotadas para a classificação de imagens histológicas. Enquanto (DIAS, 2025) exploram diretamente modelos baseados em aprendizado profundo, como CNNs e *Vision Transformers*, os trabalhos de (ROBERTO, 2021), (CARVALHO, 2024) ao passo que (FARIA, 2022) adotam uma abordagem voltada à Inteligência Artificial Explicável (XAI), priorizando a interpretabilidade por meio de atributos morfológicos e não morfológicos e do uso de modelos de classificação mais transparentes.

Além disso, nota-se que parte dos estudos utiliza imagens originais como entrada, enquanto outros transformam os dados em representações derivadas, como mapas fractais. No entanto, no contexto dos trabalhos que utilizam a base de dados Oral Epithelium DB, observa-se que poucos estudos investigam de forma sistemática a integração entre diferentes tipos de representações e arquiteturas modernas, como *Transformers*, no domínio de imagens histológicas.

Diante desse cenário, nos trabalhos analisados com essa base de dados, identificam-se algumas limitações. Primeiramente, observa-se a ausência, até onde se estende a literatura consultada, de estudos que explorem de forma conjunta representações baseadas em geometria fractal e modelos baseados em *Vision Transformers*. Além disso, há uma carência de análises comparativas entre diferentes tipos de entrada, como imagens originais e representações derivadas, avaliando o impacto dessas variações no desempenho dos modelos.

A Tabela 1 evidencia que, embora existam avanços tanto no uso de descritores fractais quanto na aplicação de arquiteturas baseadas em *Transformers*, ainda não foram identificados, no contexto dessa base de dados, trabalhos que explorem de forma integrada essas abordagens, reforçando a motivação deste estudo.

Tabela 1 – Comparação entre os trabalhos relacionados.

| Trabalho        | Entrada                | Abordagem        | Fractal | ViT |
|-----------------|------------------------|------------------|---------|-----|
| Roberto (2021)  | Fractal + original     | CNN + Comitê     | Sim     | Não |
| Faria (2022)    | Atributos morfológicos | GBDT + SVM (XAI) | Não     | Não |
| Carvalho (2024) | Fractal                | CNN              | Sim     | Não |
| Dias (2025)     | Imagem original        | CNN + ViT        | Não     | Sim |

---

## Metodologia

Este capítulo descreve o fluxo metodológico adotado neste trabalho para investigação do uso de representações fractais e modelo baseado em *Vision Transformer* (ViT) na classificação de lesões histológicas. A metodologia foi estruturada em etapas sequenciais, compreendendo: (i) organização do banco de imagens histológicas; (ii) extração de descritores fractais multiescala; (iii) geração de representações 2D a partir dos vetores de atributos; (iv) treinamento e avaliação de arquiteturas de redes neurais profundas; e (v) análise comparativa dos resultados obtidos.

### 3.1 Introdução

As etapas e estratégias adotadas neste trabalho caracterizam-se como uma aprofundamento do modelo proposto por (DIAS, 2025), ampliando sua investigação por meio da incorporação de representações fractais e da realização de novos experimentos, incluindo estudos de ablação para avaliação isolada dos módulos da arquitetura com representações baseadas em representação fractal.

Inicialmente, apresenta-se a base de dados utilizada, incluindo as características das imagens histológicas originais e os procedimentos de organização e padronização aplicados. Em seguida, descreve-se o processo de extração das características fractais a partir das imagens originais, realizado com base na implementação desenvolvida por (CARVALHO, 2024), o que permite a geração de representações bidimensionais derivadas de descritores fractais. Ressalta-se que as dimensões finais das representações variam conforme a estratégia adotada: na Reorganização Sequencial obtêm-se imagens  $10 \times 10 \times 3$ , enquanto na abordagem RecPlot geram-se imagens  $100 \times 100 \times 3$ . Posteriormente, são detalhadas as arquiteturas de redes neurais empregadas, incluindo o modelo originalmente proposto por (DIAS, 2025) — composto por redes convolucionais e um *Vision Transformer* — bem como as variações experimentais adotadas neste trabalho. Também são apresentados o protocolo de treinamento, as estratégias de aumento de dados e as métricas de avaliação utilizadas para a comparação sistemática entre as abordagens investigadas.



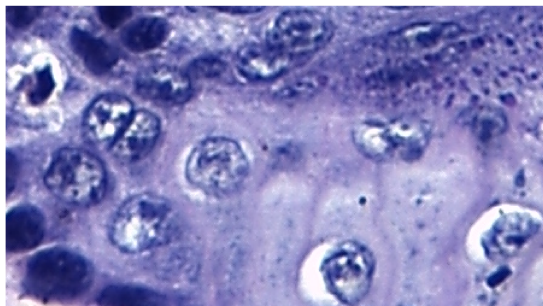
## 3.2 Banco de Imagens

Foi utilizado o banco de dados público *Oral Epithelium DB*, disponibilizado pelo grupo LIPAI no repositório GitHub<sup>1</sup>. O conjunto é composto por imagens histológicas de tecidos da cavidade oral de camundongos da linhagem C57BL/6 submetidos experimentalmente ao agente carcinógeno 4-nitroquinolina-N-óxido (4NQO).

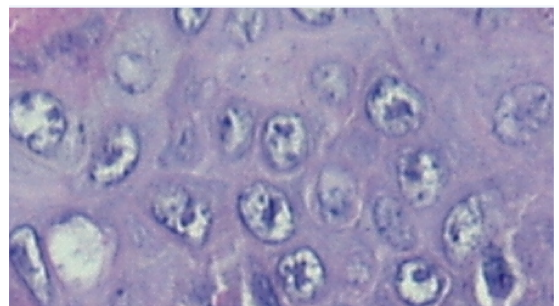
As lâminas histológicas foram digitalizadas com microscopia óptica em aumento de 400× e classificadas por especialistas de acordo com os critérios estabelecidos por (LUMERMAN; FREEDMAN; KERPEL, 1995), sendo organizadas nas classes: tecido saudável, displasia leve, moderada e severa.

Neste trabalho, foram utilizadas as imagens originais do banco como ponto de partida tanto para o treinamento direto dos modelos de aprendizado profundo quanto para a geração de representações derivadas baseadas em descritores fractais, conforme descrito na próxima subseção.

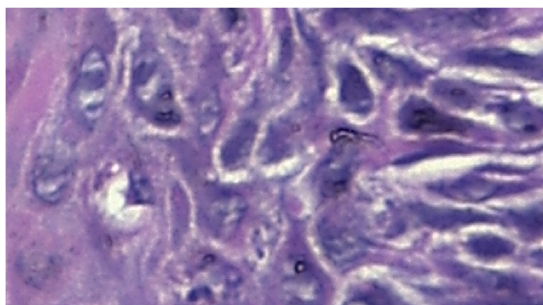
Como forma de ilustrar as características visuais das diferentes classes do banco *Oral Epithelium DB*, são apresentados exemplos de imagens histológicas originais, conforme mostrado na Figura 4.



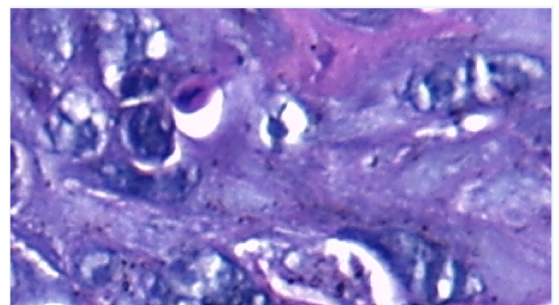
(a) Tecido saudável



(b) Displasia leve



(c) Displasia moderada



(d) Displasia severa

Figura 4 – Exemplos representativos das quatro classes do *Oral Epithelium DB*.

<sup>1</sup> <<https://github.com/LIPAIGroup/OralEpitheliumDB>>

### 3.3 Extração de Descritores Fractais

A extração dos descritores fractais foi realizada a partir das imagens histológicas do banco *Oral Epithelium DB*, utilizando a implementação proposta por (CARVALHO, 2024), baseada no algoritmo *gliding-box*. Esse método permite analisar a imagem em múltiplas escalas, capturando propriedades estruturais locais e globais do tecido (ROBERTO, 2021).

Inicialmente, todas as imagens foram padronizadas para a resolução de  $224 \times 224$  pixels. Em seguida, foram analisadas janelas quadradas de tamanho  $L \times L$ , com  $L \in \{3, 5, \dots, 41\}$ , garantindo uma análise multiescala da organização estrutural da imagem.

#### 3.3.1 Métrica de distância e análise multidimensional

Durante a aplicação do algoritmo *gliding-box*, cada pixel da janela é comparado ao pixel central por meio de uma métrica de distância. Essa etapa é fundamental, pois define quais pixels pertencem à região analisada, influenciando diretamente a construção da matriz de probabilidade utilizada no cálculo dos descritores fractais (ROBERTO, 2021; CARVALHO, 2024).

Neste trabalho, foram utilizadas três métricas de distância:

- ❑ **Distância de Minkowski (Chessboard):** considera a maior diferença entre as coordenadas dos pixels, sendo sensível à conectividade máxima na vizinhança;
- ❑ **Distância Euclidiana:** mede a distância geométrica direta entre dois pixels, considerando a relação espacial contínua;
- ❑ **Distância Manhattan:** calcula a soma das diferenças absolutas entre coordenadas, refletindo deslocamentos ao longo de eixos ortogonais.

A utilização dessas métricas permite diferentes definições de vizinhança e conectividade entre os pixels, possibilitando a captura de diferentes padrões estruturais presentes nas imagens histológicas. Essa abordagem multidimensional amplia a capacidade de representação dos descritores fractais, conforme discutido por (ROBERTO, 2021).

#### 3.3.2 Descritores extraídos

A partir das matrizes obtidas, foram extraídas três categorias de descritores fractais:

- ❑ **Dimensão Fractal (DF):** quantificada conforme a relação apresentada na Equação 1, representando a complexidade e o grau de ocupação espacial da estrutura;
- ❑ **Lacunaridade (LAC):** calculada conforme a Equação 4, sendo associada à heterogeneidade e à distribuição de lacunas na imagem;

- ❑ **Percolação (PERC)**: definida conforme a Equação 5, avaliando a conectividade estrutural e a presença de caminhos conectados ao longo da imagem.

Esses descritores permitem capturar diferentes aspectos estruturais da imagem, incluindo complexidade, heterogeneidade e conectividade do tecido analisado (ROBERTO, 2021).

### 3.3.3 Atributos locais e globais

A extração dos descritores fractais resulta em dois tipos de atributos:

- ❑ **Atributos locais**: correspondem aos valores de DF, LAC e PERC calculados para cada escala  $L$ , formando curvas multiescala;
- ❑ **Atributos globais**: são obtidos a partir dessas curvas, representando o comportamento geral dos descritores ao longo das escalas.

Os atributos globais são derivados da análise das curvas em função do tamanho da janela, incluindo medidas como área sob a curva, inclinação (no caso da dimensão fractal) e estatísticas associadas à distribuição dos valores (CARVALHO, 2024). Esses atributos permitem sintetizar a informação estrutural da imagem em um conjunto reduzido de características representativas.

Ao final do processo, os atributos locais e globais foram organizados em vetores de características, os quais foram posteriormente utilizados na construção das representações bidimensionais e na etapa de classificação.

## 3.4 Representações 2D de Descritores Fractais

A partir dos vetores de descritores fractais extraídos na etapa anterior, foram geradas representações bidimensionais com o objetivo de permitir sua utilização como entrada em arquiteturas de redes neurais profundas. A conversão dos descritores em representações 2D foi realizada conforme metodologia descrita em (CARVALHO, 2024).

Inicialmente, os atributos multiescala foram organizados em três subconjuntos associados às métricas de distância Minkowski, Euclidiana e Manhattan. Cada subconjunto foi tratado como um canal de cores independente (R, G e B), possibilitando a construção de imagens coloridas derivadas dos descritores fractais.

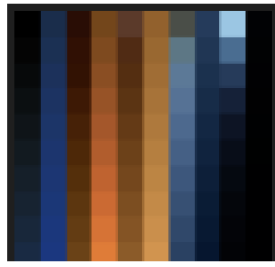
Antes da conversão para o domínio espacial, os atributos foram normalizados por blocos, de modo a reduzir variações de escala entre os diferentes grupos de características. Em seguida, foram adotadas duas estratégias de representação: reorganização sequencial e gráfico de recorrência.

### 3.4.1 Reorganização Sequencial

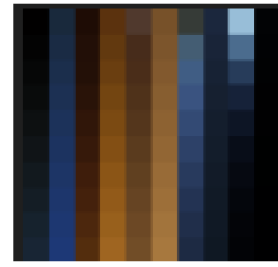
Na abordagem denominada Reorganização Sequencial (do inglês, Sequential Reshape - SR), os 300 atributos fractais locais (100 para cada distância: Minkowski, Euclidiana e Manhattan) foram organizados de forma sequencial em uma matriz tridimensional  $10 \times 10 \times 3$ . Cada conjunto de 100 atributos foi mapeado diretamente para uma matriz  $10 \times 10$ , sendo atribuído a um canal de cor específico da imagem RGB: Minkowski (R), Euclidiana (G) e Manhattan (B).

Essa estratégia consiste em uma simples reestruturação espacial do vetor unidimensional de características, preservando a ordem sequencial dos descritores, sem introduzir relações matemáticas adicionais entre os elementos.

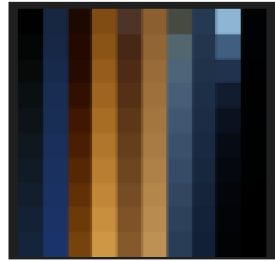
Como forma de ilustrar a aplicação da abordagem Sequencial, são apresentados exemplos das representações geradas para as quatro classes do banco *Oral Epithelium DB*, conforme mostrado na Figura 5.



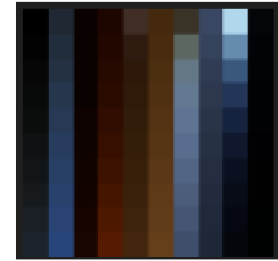
(a) Tecido saudável



(b) Displasia leve



(c) Displasia moderada



(d) Displasia severa

Figura 5 – Representações SR geradas a partir dos descritores fractais das quatro classes do *Oral Epithelium DB*.

### 3.4.2 Gráfico de Recorrência (GR)

Na abordagem denominada Gráfico de Recorrência (GR), também conhecida como *Recurrence Plot* (RP), o vetor unidimensional de atributos fractais é projetado em uma matriz bidimensional quadrada de dimensão  $N \times N$ .

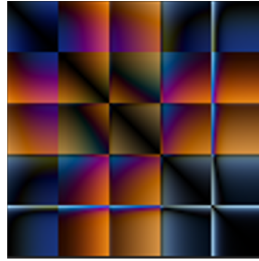
Cada elemento da matriz  $R_{i,j}$  é definido como:

$$R_{i,j} = \|x_i - x_j\| \quad (6)$$

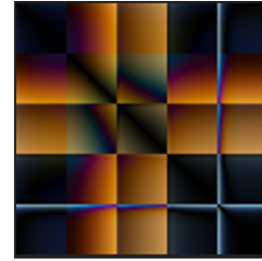
onde  $x_i$  e  $x_j$  representam os descritores fractais nas posições  $i$  e  $j$  do vetor original.

No presente trabalho, considerando vetores com 100 atributos por canal, obteve-se uma matriz  $100 \times 100$  para cada distância (Minkowski, Euclidiana e Manhattan), que foram atribuídas aos canais R, G e B, respectivamente, resultando em uma imagem RGB  $100 \times 100 \times 3$ .

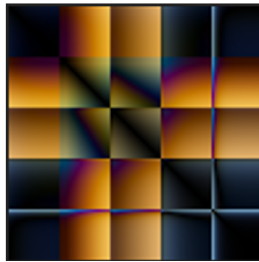
Como forma de ilustrar a aplicação da abordagem Plot de Recorrência, são apresentados exemplos das representações geradas para as quatro classes do banco *Oral Epithelium DB*, conforme mostrado na Figura 6.



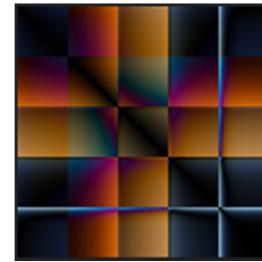
(a) Tecido saudável



(b) Displasia leve



(c) Displasia moderada



(d) Displasia severa

Figura 6 – Representações RecPlot geradas a partir dos descritores fractais das quatro classes do *Oral Epithelium DB*.

Após experimentos preliminares, observou-se que a representação baseada em RecPlot apresentou desempenho superior na tarefa de classificação, motivo pelo qual os estudos posteriores concentraram-se majoritariamente nessa estratégia.

### 3.5 Modelo de Aprendizagem Profunda

O modelo de aprendizagem profunda adotado neste trabalho baseia-se na arquitetura híbrida proposta por (DIAS, 2025), sendo ampliado com a incorporação de representações fractais e a realização de análises estruturais adicionais. A arquitetura funciona como um modelo híbrido, no qual diferentes módulos analisam a imagem sob perspectivas complementares, combinando informações locais e globais para a tomada de decisão final.

A estrutura é composta por três caminhos principais: (i) um módulo baseado em EfficientNet, responsável pela extração de padrões locais detalhados; (ii) um módulo inspirado no Vision Transformer (ViT) adaptado, voltado à captura de relações de longo alcance; e (iii) uma CNN simplificada, denominada *neck*, projetada para extrair padrões

básicos e texturas fundamentais. As saídas vetoriais desses três módulos são concatenadas e encaminhadas para camadas densas responsáveis pela classificação final.

### 3.5.1 Caminho i): EfficientNet para Extração de Detalhes

O primeiro caminho emprega a arquitetura EfficientNetV2B0 (Bottleneck), escolhida por sua eficiência computacional e capacidade de extrair características discriminativas com menor custo de parâmetros quando comparada a arquiteturas tradicionais. A família EfficientNet utiliza um escalonamento composto (compound scaling), que equilibra profundidade, largura e resolução da rede, permitindo melhor desempenho com menor complexidade (TAN; LE, 2021).

Além disso, o uso de pesos pré-treinados no ImageNet favorece a transferência de aprendizado, permitindo capturar padrões visuais relevantes mesmo em cenários com menor volume de dados, como é comum em aplicações histopatológicas. Nesse contexto, a arquitetura é particularmente adequada para a extração de detalhes locais e texturas finas, características importantes na distinção entre diferentes estágios de lesões.

### 3.5.2 Caminho ii): ViT Adaptado para Captura de Relações Globais

O segundo caminho implementa um módulo inspirado na filosofia do Vision Transformer, porém em versão adaptada, adequada ao contexto experimental e ao tamanho das imagens utilizadas. A imagem de entrada é reorganizada para simular a divisão em *patches*, permitindo a modelagem de dependências espaciais de longo alcance.

O módulo é composto por camadas convolucionais sucessivas, com ativações não lineares, responsáveis por extrair representações progressivamente mais abstratas. Operações de *MaxPooling2D* são aplicadas para redução dimensional, e a saída é transformada em vetor unidimensional por meio da operação de *Flatten*.

### 3.5.3 Caminho iii): CNN Simplificada

O terceiro caminho corresponde a uma CNN de arquitetura enxuta (Neck), projetada para atuar como extrator complementar de padrões básicos. Este módulo inicia com camada convolucional de maior largura de filtro, seguida por normalização e ativação PReLU. Operações de *MaxPooling2D* e *Flatten* finalizam a geração do vetor de características.

A arquitetura proposta neste trabalho é ilustrada na Figura 7, na qual são apresentados os três caminhos paralelos responsáveis pela extração de características em diferentes níveis de abstração, seguidos pela etapa de integração e classificação final.

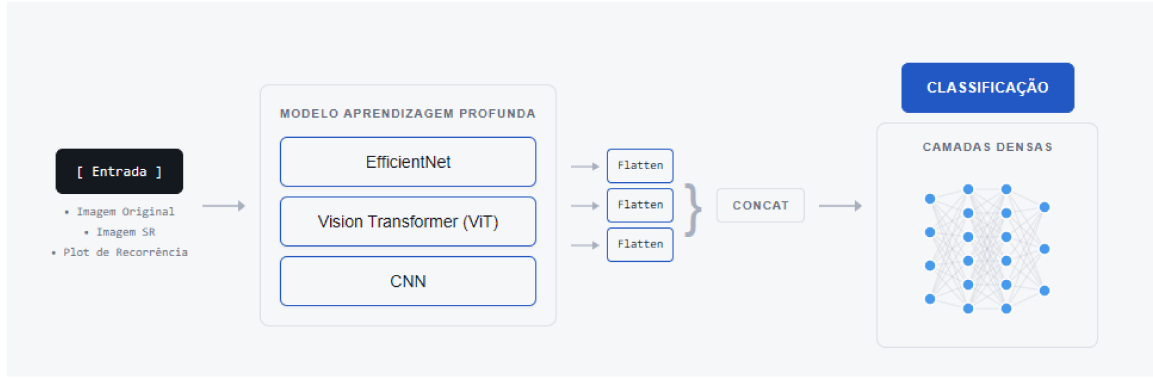


Figura 7 – Arquitetura híbrida proposta para a etapa de aprendizagem profunda, composta pelos três caminhos paralelos (EfficientNet, ViT Adaptado e CNN Simplificada), seguidos pela concatenação dos vetores de características e classificação por camadas densas.

### 3.5.4 Integração e Classificação

Os vetores produzidos pelos três caminhos são concatenados em uma única representação vetorial, que é então processada por camadas densas totalmente conectadas para a tarefa de classificação. Essa estratégia híbrida permite combinar padrões locais detalhados, relações globais e texturas fundamentais, proporcionando maior robustez em cenários com variabilidade morfológica e base de dados limitada.

### 3.5.5 Variações Estruturais

Além da configuração completa da arquitetura híbrida, foram realizadas variações estruturais com o objetivo de analisar a contribuição individual de cada módulo. Para isso, experimentos foram conduzidos com a desativação isolada dos seguintes componentes:

- ❑ Desativação do módulo Vision Transformer (ViT);
- ❑ Desativação do módulo da CNN simplificada (*neck*);
- ❑ Desativação do bloco EfficientNet (*bottleneck*);

Essas variações permitiram avaliar o impacto de cada caminho na performance global do modelo, evidenciando a contribuição relativa das representações locais, globais e básicas na tarefa de classificação. Adicionalmente, o modelo foi avaliado sob diferentes configurações de entrada, incluindo:

- ❑ Imagens histológicas originais;
- ❑ Representações por Reorganização Sequencial (SR);
- ❑ Representações obtidas por meio da técnica Recurrence Plot (*RecPlot*);

Essa análise permitiu investigar a influência das representações derivadas de geometria fractal no desempenho das arquiteturas profundas, bem como comparar sua eficácia em relação às imagens originais.

## 3.6 Configuração Experimental

Esta seção descreve os procedimentos adotados para treinamento, validação e avaliação das arquiteturas propostas, bem como as diferentes configurações experimentais investigadas.

### 3.6.1 Divisão dos Dados

O conjunto de imagens foi dividido inicialmente em 70% para treinamento e 30% para teste, mantendo a proporção estratificada entre as classes.

Posteriormente, o conjunto de treinamento foi novamente subdividido em 70% para treinamento efetivo e 30% para validação, resultando em aproximadamente 49% das amostras para treino, 21% para validação e 30% para teste.

A divisão foi realizada com semente fixa (*random\_state*), garantindo reprodutibilidade dos experimentos. O conjunto de teste permaneceu completamente isolado durante o treinamento.

### 3.6.2 Parâmetros de Treinamento

Os modelos foram treinados utilizando mini-batches com tamanho fixo de 32 amostras, função de perda categórica (*categorical cross-entropy*) e o otimizador Adam, com taxa de aprendizado inicial de 0,001,  $\beta_1 = 0,9$ ,  $\beta_2 = 0,999$  e  $\epsilon = 10^{-7}$ .

As imagens de entrada foram redimensionadas para  $224 \times 224$  pixels, garantindo compatibilidade com as arquiteturas empregadas. O modelo foi configurado para classificação multiclasse com quatro categorias.

Durante o treinamento, foram empregadas estratégias de regularização com o objetivo de evitar sobreajuste e melhorar a estabilidade do processo de convergência. O método *Early Stopping* foi utilizado com monitoramento da perda de validação (*val\_loss*), restaurando os melhores pesos obtidos e utilizando um limite máximo de 100 épocas. Adicionalmente, foi aplicada a técnica *ReduceLROnPlateau*, com fator de redução de 0,2 e paciência de 5 épocas, permitindo o ajuste dinâmico da taxa de aprendizado ao longo do treinamento.

Os experimentos foram executados em um ambiente computacional com sistema operacional Windows 10 Pro, equipado com processador Intel (arquitetura x64), 16 GB de memória RAM e GPU NVIDIA GeForce RTX 2060, com suporte à CUDA 12.6. O treinamento foi realizado utilizando as bibliotecas TensorFlow 2.21 e Keras 3.13.



O tempo médio de treinamento por experimento variou entre aproximadamente 17 e 25 minutos.

Durante o treinamento, foram monitoradas métricas de desempenho como acurácia, precisão, recall e AUC, permitindo acompanhar o comportamento do modelo ao longo das épocas.

Os hiperparâmetros adotados foram definidos empiricamente com base em experimentos preliminares, buscando um equilíbrio entre desempenho, estabilidade e custo computacional. Além disso, o ambiente computacional foi mantido fixo para todos os experimentos, garantindo consistência na avaliação dos resultados.

### 3.7 Métricas de Avaliação

A avaliação do desempenho dos modelos foi realizada por meio das métricas clássicas de classificação: Acurácia (Accuracy), Precisão (Precision), Sensibilidade (Recall) e F1-score. Além das métricas quantitativas, foram analisadas matrizes de confusão para examinar o comportamento do modelo entre as diferentes classes de displasia.

O cálculo das métricas foi realizado com o auxílio da biblioteca *scikit-learn*, por meio da função `classification_report`, que fornece os valores individuais por classe, bem como médias macro e ponderadas. Considerando que os experimentos foram conduzidos em cenários de classificação binária, as métricas foram calculadas diretamente a partir da matriz de confusão correspondente a cada par de classes analisado.

As métricas são derivadas da **matriz de confusão**, estrutura que organiza as predições do modelo em relação aos rótulos reais. A partir dessa matriz, definem-se:

- ❑ **VP (Verdadeiros Positivos)**: amostras corretamente classificadas como pertencentes a uma determinada classe;
- ❑ **VN (Verdadeiros Negativos)**: amostras corretamente classificadas como não pertencentes à classe analisada;
- ❑ **FP (Falsos Positivos)**: amostras incorretamente classificadas como pertencentes à classe;
- ❑ **FN (Falsos Negativos)**: amostras pertencentes à classe que foram incorretamente classificadas como outra classe.

Com base nessas definições, as métricas são formalmente descritas a seguir.

A Acurácia (Acc) mede a proporção total de predições corretas:

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (7)$$

A Precisão (Prec) indica a proporção de amostras classificadas como positivas que realmente pertencem à classe:

$$Prec = \frac{VP}{VP + FP} \quad (8)$$

O Recall (Sensibilidade) mede a capacidade do modelo em identificar corretamente as amostras da classe:

$$Recall = \frac{VP}{VP + FN} \quad (9)$$

O F1-score representa a média harmônica entre Precisão e Recall, sendo particularmente útil quando há desbalanceamento entre classes:

$$F1 = 2 \cdot \frac{Prec \cdot Recall}{Prec + Recall} \quad (10)$$

---

## Avaliação dos Resultados

### 4.1 Organização dos Experimentos

Os experimentos foram conduzidos em cenários de classificação binária, considerando diferentes pares de classes do banco *Oral Epithelium DB*, a saber: saudável vs displasia leve, saudável vs displasia moderada, saudável vs displasia severa, displasia leve vs displasia moderada, displasia leve vs displasia severa e displasia moderada vs displasia severa. Para cada par de classes, foram avaliadas três configurações de entrada: (i) imagens histológicas originais; (ii) representações por Reorganização Sequencial (SR); e (iii) representações baseadas em *Recurrence Plot* (RecPlot). Além disso, a arquitetura híbrida completa foi comparada com variações estruturais obtidas por meio da desativação isolada de cada módulo (EfficientNet, ViT Adaptado e CNN simplificada), caracterizando o estudo de ablação.

A divisão dos dados foi mantida fixa para todas as configurações experimentais, garantindo comparabilidade entre os resultados. As métricas reportadas correspondem à Acurácia, Precisão, Recall e F1-score, calculadas a partir do conjunto de teste previamente isolado.

### 4.2 Avaliação com Imagens Histológicas Originais

Nesta seção são apresentados os resultados obtidos a partir da utilização direta das imagens histológicas originais do banco *Oral Epithelium DB*, sem aplicação de transformações baseadas em descritores fractais.

Os experimentos foram conduzidos em cenários de classificação binária, considerando diferentes pares de classes do conjunto de dados. A Tabela 2 apresenta os valores de Acurácia, Precisão, Recall e F1-score obtidos para cada comparação.

Tabela 2 – Resultados obtidos com imagens histológicas originais

| Comparação          | Accuracy (%)         | Precision (%)        | Recall (%)           | F1-score (%)         |
|---------------------|----------------------|----------------------|----------------------|----------------------|
| Saudável × Leve     | 98,55                | 98,61                | 98,53                | 98,55                |
| Saudável × Moderado | 98,55                | 98,61                | 98,53                | 98,55                |
| Saudável × Severo   | 94,20                | 94,20                | 94,20                | 94,20                |
| Leve × Moderado     | 72,46                | 74,11                | 72,65                | 72,09                |
| Leve × Severo       | 75,36                | 76,46                | 75,21                | 75,03                |
| Moderado × Severo   | 52,17                | 75,37                | 52,86                | 39,07                |
| <b>Média ± DP</b>   | <b>81,88 ± 16,94</b> | <b>86,23 ± 11,03</b> | <b>82,00 ± 16,72</b> | <b>79,58 ± 21,02</b> |

Além dos valores médios apresentados, a análise do desvio padrão permite avaliar a estabilidade do modelo entre os diferentes cenários. Observa-se que, nos casos envolvendo a classe saudável, os resultados apresentam menor variabilidade, indicando maior consistência no desempenho do modelo.

Em contrapartida, cenários mais desafiadores, como leve × moderado e moderado × severo, apresentam maior variação nas métricas, refletindo a dificuldade do modelo em distinguir classes com características histológicas semelhantes.

De modo geral, observa-se que os cenários envolvendo a classe saudável apresentaram os maiores valores de desempenho, indicando que as diferenças morfológicas entre tecidos saudáveis e tecidos displásicos são suficientemente evidentes para permitir uma separação eficaz pelo modelo.

Por outro lado, os cenários envolvendo classes adjacentes de displasia, como leve × moderado e moderado × severo, apresentaram menor desempenho. Esse comportamento é esperado, uma vez que esses estágios da doença apresentam características histológicas progressivas e morfolologicamente semelhantes, o que torna a tarefa de classificação mais desafiadora.

Os resultados obtidos apresentam comportamento semelhante ao reportado por (DIAS, 2025), que também observou maior facilidade na separação entre tecidos saudáveis e displásicos quando comparados aos cenários envolvendo diferentes graus de displasia.

Essa análise estabelece uma base de comparação para os experimentos apresentados nas seções seguintes, nos quais são investigadas representações alternativas baseadas em descritores fractais.

### 4.3 Análise com Reorganização Sequencial (SR)

Nesta seção são apresentados os resultados obtidos utilizando as representações derivadas por Reorganização Sequencial (Sequential Reshape – SR), geradas a partir dos descritores fractais extraídos das imagens histológicas do banco *Oral Epithelium DB*.

Nesse processo, os vetores de atributos fractais são reorganizados sequencialmente em matrizes bidimensionais, formando imagens RGB que preservam a estrutura dos descri-

tores em formato espacial, permitindo sua utilização como entrada para arquiteturas de redes neurais convolucionais.

A tabela 3 apresenta os valores de Acurácia, Precisão, Recall e F1-score obtidos para cada par de classes considerado nos experimentos.

Tabela 3 – Resultados obtidos com imagens utilizando Reorganização Sequencial (SR)

| Comparação          | Accuracy (%)        | Precision (%)       | Recall (%)          | F1-score (%)        |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| Saudável × Leve     | 97,10               | 97,30               | 97,06               | 97,10               |
| Saudável × Moderado | 100,00              | 100,00              | 100,00              | 100,00              |
| Saudável × Severo   | 88,41               | 90,56               | 88,41               | 88,23               |
| Leve × Moderado     | 88,41               | 90,48               | 88,57               | 88,29               |
| Leve × Severo       | 84,06               | 84,83               | 83,95               | 83,94               |
| Moderado × Severo   | 75,36               | 77,19               | 75,55               | 75,03               |
| <b>Média ± DP</b>   | <b>88,89 ± 8,14</b> | <b>90,06 ± 7,58</b> | <b>88,92 ± 8,09</b> | <b>88,77 ± 8,25</b> |

De modo geral, observa-se que o modelo foi capaz de extrair padrões relevantes a partir das representações derivadas dos descritores fractais, apresentando desempenho consistente em diferentes cenários de classificação. Comparativamente aos resultados obtidos com as imagens histológicas originais, nota-se que a representação por Reorganização Sequencial mantém a capacidade discriminativa das características do tecido, indicando que os descritores fractais capturam informações estruturais importantes para a tarefa de classificação.

Observa-se também uma melhoria nos cenários envolvendo classes adjacentes de displasia quando comparados aos resultados obtidos com as imagens originais, sugerindo que a reorganização sequencial dos descritores fractais contribui para evidenciar padrões estruturais associados às alterações progressivas do tecido.

Entretanto, ainda se nota maior dificuldade na distinção entre as classes moderado e severo, que apresentam padrões histológicos mais semelhantes, embora também tenha sido observada uma melhora em relação aos resultados obtidos com as imagens originais. Esses resultados reforçam o potencial das representações derivadas de descritores fractais como alternativa para a análise de imagens histológicas, servindo também como base para a avaliação da abordagem baseada em gráficos de recorrência apresentada na próxima seção.

## 4.4 Estudo com Gráficos de Recorrência

Nesta seção são apresentados os resultados obtidos utilizando representações baseadas em gráficos de recorrência (*Recurrence Plots* – RecPlot), geradas a partir dos descritores fractais extraídos das imagens histológicas do banco *Oral Epithelium DB*.

A Tabela 4 apresenta os valores de Acurácia, Precisão, Recall e F1-score obtidos para cada par de classes considerado nos experimentos.

Tabela 4 – Resultados obtidos com Gráficos de Recorrência (RecPlot)

| Comparação          | Accuracy (%)        | Precision (%)       | Recall (%)          | F1-score (%)        |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| Saudável × Leve     | 100,00              | 100,00              | 100,00              | 100,00              |
| Saudável × Moderado | 98,55               | 98,57               | 98,57               | 98,55               |
| Saudável × Severo   | 98,55               | 98,61               | 98,53               | 98,55               |
| Leve × Moderado     | 92,75               | 92,80               | 92,73               | 92,75               |
| Leve × Severo       | 97,10               | 97,30               | 97,06               | 97,10               |
| Moderado × Severo   | 95,65               | 95,69               | 95,65               | 95,65               |
| <b>Média ± DP</b>   | <b>97,10 ± 2,37</b> | <b>97,16 ± 2,36</b> | <b>97,09 ± 2,37</b> | <b>97,10 ± 2,37</b> |

De modo geral, observa-se que a utilização dos gráficos de recorrência permitiu destacar padrões estruturais presentes nos descritores fractais, contribuindo para um desempenho superior do modelo nos diferentes cenários de classificação avaliados. Em comparação às representações anteriores, os resultados indicam que essa abordagem preserva de forma mais eficaz informações discriminativas relevantes para a identificação das diferentes classes histológicas.

Os maiores valores de desempenho foram observados nos cenários que envolvem a classe saudável, indicando que as diferenças estruturais entre tecidos saudáveis e tecidos displásicos são capturadas com elevada precisão pelo modelo. Mesmo nas comparações entre classes adjacentes de displasia, que apresentam maior similaridade morfológica, os resultados obtidos permanecem elevados, indicando que a representação baseada em gráficos de recorrência é capaz de evidenciar padrões relevantes para a distinção entre os diferentes graus da doença.

De forma geral, os resultados demonstram que os gráficos de recorrência constituem uma representação particularmente eficaz dos descritores fractais, permitindo alcançar alto desempenho médio na tarefa de classificação proposta neste trabalho.

Considerando o desempenho superior obtido com essa representação, as imagens baseadas em RecPlot foram utilizadas nos experimentos de análise de ablação apresentados na próxima seção, nos quais diferentes componentes da arquitetura do modelo são desativados com o objetivo de avaliar individualmente a contribuição de cada rede para o desempenho final do sistema.

## 4.5 Análise de Ablação

A análise de ablação foi conduzida com o objetivo de investigar a contribuição individual dos diferentes componentes da arquitetura proposta. Para isso, foram realizados experimentos adicionais nos quais cada um dos módulos do modelo foi desativado individualmente, permitindo avaliar o impacto de cada rede no desempenho final do sistema.

Considerando que as representações baseadas em gráficos de recorrência (RecPlot) apresentaram os melhores resultados nos experimentos anteriores, essa representação foi

utilizada como base para os testes de ablação. Dessa forma, todos os experimentos desta seção foram realizados utilizando as imagens RecPlot como entrada do modelo.

Inicialmente, foi avaliado o desempenho do modelo completo, com os três ramos da arquitetura ativos. Em seguida, novos experimentos foram realizados desativando individualmente os componentes *Bottleneck*, *Neck* e *Transformer*. Essa estratégia permite observar como a remoção de cada módulo afeta o processo de extração de características e, conseqüentemente, o desempenho do classificador.

#### 4.5.1 Ablação do Módulo *Bottleneck*

A Tabela 5 apresenta os resultados obtidos com a desativação do módulo *Bottleneck*, mantendo os demais componentes da arquitetura ativos e utilizando representações baseadas em RecPlot como entrada.

Tabela 5 – Resultados da análise de ablação com desativação do módulo *Bottleneck*.

| Comparação          | Acurácia (%)        | Precisão (%)        | Recall (%)          | F1-score (%)        |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| Saudável × Leve     | 97,10               | 97,22               | 97,14               | 97,10               |
| Saudável × Moderado | 95,65               | 95,67               | 95,67               | 95,65               |
| Saudável × Severo   | 95,65               | 95,95               | 95,71               | 95,65               |
| Leve × Moderado     | 95,65               | 95,67               | 95,67               | 95,65               |
| Leve × Severo       | 91,30               | 91,30               | 91,30               | 91,30               |
| Moderado × Severo   | 94,20               | 94,20               | 94,20               | 94,20               |
| <b>Média ± DP</b>   | <b>94,92 ± 1,82</b> | <b>95,00 ± 1,87</b> | <b>94,95 ± 1,84</b> | <b>94,92 ± 1,82</b> |

Observa-se que a remoção do módulo *Bottleneck* resulta em uma leve redução no desempenho global, com acurácia média de 94,92%. O impacto mais significativo ocorre na comparação leve × severo, indicando maior dificuldade na distinção entre classes com características mais semelhantes.

#### 4.5.2 Ablação do Módulo *Neck*

A Tabela 6 apresenta os resultados obtidos com a desativação do módulo *Neck*, mantendo os demais componentes da arquitetura ativos e utilizando representações RecPlot como entrada.

Tabela 6 – Resultados da análise de ablação com desativação do módulo *Neck*.

| Comparação          | Acurácia (%)        | Precisão (%)        | Recall (%)          | F1-score (%)        |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| Saudável × Leve     | 100,00              | 100,00              | 100,00              | 100,00              |
| Saudável × Moderado | 98,55               | 98,57               | 98,57               | 98,55               |
| Saudável × Severo   | 84,06               | 87,78               | 84,29               | 83,72               |
| Leve × Moderado     | 98,55               | 98,61               | 98,53               | 98,55               |
| Leve × Severo       | 97,10               | 97,10               | 97,10               | 97,10               |
| Moderado × Severo   | 89,86               | 89,88               | 89,86               | 89,85               |
| <b>Média ± DP</b>   | <b>94,69 ± 5,78</b> | <b>95,32 ± 4,71</b> | <b>94,73 ± 5,71</b> | <b>94,63 ± 5,88</b> |

A remoção do módulo *Neck* resulta em maior variabilidade no desempenho, embora a acurácia média permaneça próxima à observada com o modelo completo. Nota-se uma queda mais acentuada na comparação saudável  $\times$  severo, indicando maior dificuldade do modelo em manter o equilíbrio entre as classes nesse cenário. Por outro lado, o desempenho permanece elevado em casos como saudável  $\times$  leve, sugerindo que os demais componentes da arquitetura são suficientes para capturar diferenças mais evidentes. Esses resultados indicam que o *Neck* atua como um componente auxiliar, contribuindo principalmente para a estabilidade e consistência das predições.

### 4.5.3 Ablação do Módulo *Transformer*

A Tabela 7 apresenta os resultados obtidos com a desativação do módulo *Transformer*, mantendo os demais componentes da arquitetura ativos e utilizando representações RecPlot como entrada.

Tabela 7 – Resultados da análise de ablação com desativação do módulo *Transformer*.

| Comparação                       | Acurácia (%)                       | Precisão (%)                       | Recall (%)                         | F1-score (%)                       |
|----------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Saudável $\times$ Leve           | 97,10                              | 97,22                              | 97,14                              | 97,10                              |
| Saudável $\times$ Moderado       | 100,00                             | 100,00                             | 100,00                             | 100,00                             |
| Saudável $\times$ Severo         | 98,55                              | 98,57                              | 98,57                              | 98,55                              |
| Leve $\times$ Moderado           | 98,55                              | 98,57                              | 98,57                              | 98,55                              |
| Leve $\times$ Severo             | 89,86                              | 90,78                              | 89,75                              | 89,78                              |
| Moderado $\times$ Severo         | 84,06                              | 84,76                              | 84,06                              | 83,96                              |
| <b>Média <math>\pm</math> DP</b> | <b>94,69 <math>\pm</math> 5,78</b> | <b>94,98 <math>\pm</math> 5,45</b> | <b>94,68 <math>\pm</math> 5,80</b> | <b>94,66 <math>\pm</math> 5,82</b> |

A remoção do módulo *Transformer* resulta em uma redução do desempenho em comparações envolvendo classes com maior proximidade estrutural, como leve  $\times$  severo e moderado  $\times$  severo. Apesar de o modelo manter desempenho elevado em cenários mais simples, observa-se maior variabilidade nos resultados, indicando perda na capacidade de capturar relações globais entre os padrões.

Esses resultados sugerem que o módulo *Transformer* desempenha papel relevante na modelagem de dependências espaciais de longo alcance, contribuindo principalmente para a distinção entre classes mais complexas. Sua remoção evidencia uma redução na capacidade do modelo em generalizar padrões mais sutis, reforçando sua importância na arquitetura proposta.

## 4.6 Comparação com o Estado da Arte

Com o objetivo de posicionar o método proposto em relação aos trabalhos existentes na literatura, foi realizada uma análise comparativa considerando abordagens baseadas em descritores fractais, redes neurais convolucionais e arquiteturas baseadas em Transformers.



A Tabela 8 apresenta uma síntese dos principais trabalhos relacionados, destacando suas abordagens e os desempenhos reportados.

Tabela 8 – Comparação entre diferentes abordagens da literatura e o método proposto para classificação de imagens histológicas.

| Trabalho               | Abordagem                               | Acc (%)      |
|------------------------|-----------------------------------------|--------------|
| (CARVALHO, 2024)       | Descritores fractais + SVM/ANN          | 95,60        |
| (FARIA, 2022)          | Atributos morfológicos + GBDT/SVM (XAI) | 94,20        |
| (ROBERTO, 2021)        | Fractal + CNN + Ensemble                | 99,77        |
| (DIAS, 2025)           | CNN + Vision Transformer                | 90,96        |
| <b>Método Proposto</b> | RecPlot + CNN + ViT                     | <b>97,10</b> |

Observa-se que trabalhos como (CARVALHO, 2024) exploram descritores fractais combinados com classificadores tradicionais e redes neurais convolucionais, evidenciando a eficácia dessas representações na captura da complexidade estrutural dos tecidos. Por outro lado, (FARIA, 2022) adota uma abordagem voltada à Inteligência Artificial Explicável, baseada na utilização de atributos morfológicos e não morfológicos interpretáveis associados a modelos de classificação, enquanto (DIAS, 2025) investiga o uso de arquiteturas baseadas em Vision Transformers, destacando a capacidade desses modelos em capturar relações globais nas imagens.

O trabalho de (ROBERTO, 2021) apresenta uma abordagem híbrida, combinando descritores fractais com redes neurais convolucionais e comitês de classificadores, alcançando elevados níveis de desempenho. No entanto, tais abordagens não exploram a integração com arquiteturas baseadas em Transformers. Nesse contexto, o método proposto diferencia-se por integrar representações baseadas em RecPlot com arquiteturas híbridas compostas por CNNs e Vision Transformers, permitindo a exploração conjunta de padrões locais e globais.

Embora alguns trabalhos da literatura apresentem acurácias superiores, essa comparação deve ser analisada com cautela, considerando diferenças nos conjuntos de dados, estratégias de validação e protocolos experimentais. Ainda assim, os resultados obtidos indicam que a abordagem proposta apresenta desempenho competitivo, ao mesmo tempo em que introduz uma estratégia diferenciada de integração entre representações estruturais e modelos de deep learning modernos.

---

## Conclusão

Este trabalho teve como objetivo investigar o uso de diferentes formas de representação de informações derivadas de imagens histológicas, com ênfase na utilização de descritores fractais e suas transformações, para a tarefa de classificação de lesões epiteliais.

Ao longo do desenvolvimento, foi possível atingir o objetivo geral proposto, uma vez que diferentes abordagens de representação foram implementadas, avaliadas e comparadas de forma sistemática. Os experimentos realizados permitiram analisar o impacto de cada tipo de entrada no desempenho do modelo, evidenciando a influência direta da forma de representação na qualidade da classificação.

Em relação aos objetivos específicos, inicialmente foi estabelecida uma linha de base a partir da utilização das imagens histológicas originais, permitindo compreender o comportamento do modelo diante dos dados brutos. Em seguida, foram exploradas representações derivadas de descritores fractais por meio da Reorganização Sequencial (SR), o que possibilitou avaliar a capacidade dessas características em preservar informações estruturais relevantes do tecido. Posteriormente, a utilização de gráficos de recorrência (RecPlot) permitiu investigar uma forma mais complexa de representação dos descritores fractais, evidenciando ganhos expressivos de desempenho. Por fim, a realização da análise de ablação possibilitou compreender a contribuição individual dos diferentes componentes da arquitetura proposta, permitindo uma avaliação mais aprofundada do modelo.

Dessa forma, o desenvolvimento do trabalho atendeu aos objetivos propostos, fornecendo evidências experimentais consistentes sobre a eficácia das abordagens investigadas.

### 5.1 Principais Contribuições

As principais contribuições deste trabalho estão relacionadas à investigação e validação de diferentes estratégias de representação de informações para a classificação de lesões histológicas. Inicialmente, destaca-se a análise comparativa entre três formas de entrada distintas: imagens histológicas originais, representações derivadas por Reorganização Sequencial (SR) e representações baseadas em gráficos de recorrência (RecPlot).

Essa comparação permitiu evidenciar de forma clara o impacto da representação dos dados no desempenho do modelo.

Os resultados experimentais demonstraram que a utilização de descritores fractais, quando organizados em estruturas adequadas, contribui significativamente para a melhoria da capacidade discriminativa do modelo. Em especial, a abordagem baseada em RecPlot apresentou desempenho superior, com altos valores de acurácia, precisão, recall e F1-score em todos os cenários avaliados, além de menor variabilidade nos resultados.

Outra contribuição relevante refere-se à utilização de uma arquitetura híbrida composta por múltiplos ramos de extração de características. A análise de ablação evidenciou que os diferentes componentes do modelo atuam de forma complementar, contribuindo para a melhoria do desempenho global. Dessa forma, os resultados obtidos validam a hipótese de que a combinação de descritores fractais com representações estruturais derivadas, aliada a arquiteturas híbridas (CNNs, e Transformers), constitui uma estratégia relevante para a classificação de lesões histológicas.

Como trabalhos futuros, sugere-se a utilização de bases de dados maiores e mais diversificadas, bem como a adoção de estratégias mais robustas de validação, como a divisão dos dados por paciente. Outra possibilidade de investigação consiste no aprimoramento da arquitetura baseada em Transformer utilizada neste trabalho. A incorporação de mecanismos mais avançados de *self-attention* pode contribuir para uma melhor captura de relações globais entre os padrões presentes nas representações RecPlot.

---

## Referências

BARNSLEY, M. F. **Fractals Everywhere**. 2nd. ed. Boston: Academic Press, 1993. ISBN 9780120790623. Citado na página 16.

BENTAIEB, A.; HAMARNEH, G. Adversarial stain transfer for histopathology image analysis. **IEEE Transactions on Medical Imaging**, v. 37, n. 3, p. 792–802, 2018. Citado na página 13.

CARVALHO, R. H. d. O. **Classificação de displasia da cavidade oral baseada em descritores fractais e ensemble learning**. Dissertação (Mestrado) — Universidade Federal de Uberlândia (UFU), 2024. [Online; acessado em 08 abril 2025]. Disponível em: <<https://repositorio.ufu.br/handle/123456789/43129>>. Citado 7 vezes nas páginas 18, 20, 21, 23, 25, 26 e 40.

CARVALHO, T. M. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação), **Explorando modelos em cascata de CNNs e fractais para classificação de displasia oral**. Uberlândia: [s.n.], 2023. Disponível em: <<https://repositorio.ufu.br/handle/123456789/39840>>. Acesso em: 14 abr. 2025. Citado na página 9.

CHOLLET, F. **Keras: Deep Learning for Python**. 2015. <[https://keras.io/getting\\_started/](https://keras.io/getting_started/)>. [Online; accessed 08-Março-2025]. Citado na página 10.

DIAS, G. R. C. **Título completo do trabalho**. 2025. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação), Universidade Federal de Uberlândia. Disponível em: <<https://repositorio.ufu.br/handle/123456789/47646>>. Citado 5 vezes nas páginas 21, 23, 28, 35 e 40.

DOSOVITSKIY, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. **arXiv preprint arXiv:2010.11929**, 2020. Disponível em: <<https://arxiv.org/abs/2010.11929>>. Citado 4 vezes nas páginas 4, 10, 18 e 19.

FARIA, T. M. **Metodologia para a Classificação Multiclasse de Imagens Histológicas baseada em Inteligência Artificial Explicável**. Tese (Doutorado) — Universidade Federal de Uberlândia (UFU), 2022. [Online; acessado em 10-Março-2025]. Disponível em: <<https://repositorio.ufu.br/handle/123456789/35475>>. Citado 4 vezes nas páginas 9, 20, 21 e 40.

FOUNDATION, P. S. **Linguagem de Programação**. 2025. <<https://www.python.org/>>. [Online; accessed 08-Março-2025]. Citado na página 10.

GURCAN, M. N. et al. Histopathological image analysis: A review. **IEEE Reviews in Biomedical Engineering**, v. 2, p. 147–171, 2009. Citado 3 vezes nas páginas 15, 17 e 18.

HAYKIN, S. **Neural Networks and Learning Machines**. 3rd. ed. [S.l.]: Prentice Hall, 2009. Citado na página 17.

HE, W. et al. Progress of machine vision in the detection of cancer cells in histopathology. **IEEE Access**, v. 10, p. 46753–46771, 2022. Citado 2 vezes nas páginas 4 e 15.

KHAN, S. et al. Transformers in vision: A survey. **ACM Computing Surveys**, v. 54, n. 10s, p. 1–41, 2022. Citado 2 vezes nas páginas 10 e 18.

LITJENS, G. et al. A survey on deep learning in medical image analysis. **Medical Image Analysis**, v. 42, p. 60–88, 2017. ISSN 1361-8415. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1361841517301135>>. Citado na página 16.

LOPES, R.; BETROUNI, N. Fractal and multifractal analysis of digital images of granulometric materials. **Journal of Microscopy**, Wiley, v. 233, n. 3, p. 498–505, 2009. Citado na página 16.

LUMERMAN, H.; FREEDMAN, P.; KERPEL, S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. **Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology**, v. 79, n. 3, p. 321–329, 1995. Disponível em: <[https://www.oooojournal.net/article/S1079-2104\(05\)80226-4/abstract](https://www.oooojournal.net/article/S1079-2104(05)80226-4/abstract)>. Citado na página 24.

MANDELBROT, B. B. **The Fractal Geometry of Nature**. Revised and enlarged edition. San Francisco: W. H. Freeman, 1983. ISBN 9780716711865. Disponível em: <<https://books.google.com/books?id=0R2LkE3N7-oC>>. Citado na página 16.

MATHWORKS. **MATLAB R2024b: The Language of Technical Computing**. 2024. <<https://www.mathworks.com/products/matlab.html>>. [Online; acessado em 8-Março-2025]. Citado na página 10.

META AI. **PyTorch: An Open Source Machine Learning Framework**. 2016. <<https://pytorch.org/get-started/locally/>>. [Online; accessed 8-Março-2025]. Citado na página 10.

MULLIQI, N. Towards clinically reliable artificial intelligence for prostate cancer pathology. **Digital Pathology Insights**, 2025. Disponível em: <<https://scispace.com/papers/towards-clinically-reliable-artificial-intelligence-for-wk4plw3lpqgl>>. Citado 2 vezes nas páginas 13 e 14.

MUNDHADA, A. et al. Differentiation of urothelial carcinoma in histopathology images using deep learning and visualization. **Journal of Pathology Informatics**, v. 14, p. 100155, 2023. Fonte da imagem de workflow de histopatologia digital. Citado na página 14.

NIELSEN, P. S.; GEORGEN, J. B. et al. Computer-assisted annotation of digital h&e/sox10 dual stains generates high-performing convolutional neural network for calculating tumor burden in h&e-stained cutaneous melanoma. **International Journal**

of **Environmental Research and Public Health**, v. 19, n. 21, p. 14327, 2022. Citado 2 vezes nas páginas 13 e 14.

PAIS, E. **El país, deteccion cancer precoce**. 2025. <<https://elpais.com/salud-y-bienestar/el-medico-de-cerca/2025-01-29/la-importancia-del-cribado-en-la-deteccion-del-cancer.html>>. [Online; accessed 11-Março-2025]. Citado na página 9.

ROBERTO, G. F. **Associação entre atributos manuais e aprendizado profundo baseada em geometria fractal para classificação de imagens histológicas**. Tese (Doutorado) — Universidade Federal de Uberlândia (UFU), 2021. [Online; acessado em 08 abril 2025]. Disponível em: <<https://repositorio.ufu.br/handle/123456789/33905>>. Citado 7 vezes nas páginas 16, 18, 20, 21, 25, 26 e 40.

TAN, M.; LE, Q. Efficientnetv2: Smaller models and faster training. In: MEILA, M.; ZHANG, T. (Ed.). **Proceedings of the 38th International Conference on Machine Learning**. PMLR, 2021. (Proceedings of Machine Learning Research, v. 139), p. 10096–10106. Disponível em: <<https://proceedings.mlr.press/v139/tan21a.html>>. Citado na página 29.

THAGAARD, J.; STOVGAARD, E. S. et al. Automated quantification of stil density with h&e-based digital image analysis has prognostic potential in triple-negative breast cancers. **Cancers**, v. 13, n. 12, p. 3050, 2021. Citado na página 14.

VASWANI, A. et al. Attention is all you need. In: **Advances in Neural Information Processing Systems (NeurIPS)**. [s.n.], 2017. Disponível em: <[https://papers.nips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)>. Citado na página 18.