

Uso Combinado de Transformações Complexas e Modelos Híbridos na Classificação de Imagens Histológicas

1 Introdução

Com o avanço da informatização no setor da saúde, a produção de imagens médicas se intensificou significativamente, especialmente as imagens histológicas geradas a partir de biópsias e técnicas de coloração de tecidos. Essas imagens desempenham um papel fundamental na detecção precoce, diagnóstico e monitoramento de diversas doenças, como cânceres e condições inflamatórias. No entanto, sua análise manual é uma tarefa complexa, que demanda alto grau de especialização e está sujeita a variações interobservador. Nesse cenário, métodos computacionais para a classificação automática dessas imagens têm se consolidado como uma alternativa promissora para apoiar a tomada de decisões clínicas, contribuindo para a redução de erros e a otimização de recursos. Essa abordagem ganha ainda mais relevância diante da crescente demanda por diagnósticos rápidos e acessíveis, especialmente em regiões com escassez de profissionais especializados, onde soluções baseadas em inteligência artificial se mostram cada vez mais indispensáveis.

Entre os avanços recentes na área, destacam-se as arquiteturas híbridas que combinam *Convolutional Neural Networks* (CNNs) com *Transformers*, aproveitando diferentes níveis de representação da imagem. Estudos como o de Wu et al. [1] evidenciam a eficácia desses modelos em tarefas de classificação multilabel de imagens médicas, superando abordagens tradicionais como ResNet e *Vision Transformers* (ViT). Apesar desse progresso, ainda há uma lacuna no uso de técnicas de *data augmentation* que explorem plenamente o potencial dessas arquiteturas híbridas. Em especial, observa-se uma escassez de abordagens que considerem transformações complexas como estratégia de geração de dados sintéticos, o que representa uma oportunidade de pesquisa relevante.

As redes complexas têm se mostrado eficazes em tarefas de reconhecimento de padrões e análise de texturas, conforme demonstrado por [2]. Nessa abordagem, a imagem é modelada como um grafo complexo onde os pixels são representados como vértices e as conexões são estabelecidas com base em distância euclidiana e diferença de intensidade na escala de cinza. O processo converte medidas topológicas da rede (como grau e força dos vértices) em novos mapas de características, gerando representações que destacam padrões texturais em múltiplas escalas - desde micro até macroestruturas - conforme o raio de conexão varia. Enquanto trabalhos como [2] utilizam essas representações como entrada para redes neurais randômicas na extração de descritores para classificação, a mesma técnica pode ser adaptada para gerar variações da imagem original, assim caracterizando uma transformação complexa, mantendo suas propriedades estatísticas fundamentais. Essa capacidade é particularmente relevante em domínios com escassez de dados, como imagens médicas [3], onde a geração de novas amostras sintéticas pode melhorar o desempenho de modelos de aprendizado profundo.

Este trabalho propõe a integração de transformações complexas a uma arquitetura híbrida composta por CNNs e *Transformers*, combinando as capacidades complementares dessas redes: enquanto as CNNs são eficazes na identificação de padrões locais, os Transformers se destacam na captura de relações contextuais globais por meio do mecanismo de autoatenção [4]. Essa combinação é especialmente relevante em cenários que exigem múltiplas escalas de análise, como ocorre em imagens médicas transformadas com variações de raio. Dessa forma, busca-se promover uma aprendizagem mais robusta e sensível às nuances presentes nas imagens clínicas, explorando de maneira sinérgica características locais e contextuais.

Diante disso, esta proposta visa investigar a eficácia do uso de transformações complexas como técnica de aumento de dados em modelos híbridos de classificação de imagens histológicas. Ao empregar essa abordagem em conjunto com arquiteturas baseadas em CNNs e Transformers, objetiva-se extrair representações multiescalares mais informativas. Além disso, será realizada uma comparação sistemática com outra técnica contemporânea de aumento de dados amplamente utilizada: as *Generative Adversarial Networks* (GANs). Com isso, pretende-se conduzir uma

análise abrangente da efetividade dessas diferentes estratégias no aprimoramento da classificação médica assistida por inteligência artificial.

2 Objetivos

Neste projeto, a meta é investigar o impacto de técnicas modernas de aumento de dados e arquiteturas híbridas na tarefa de classificação de imagens médicas. Para tanto, pretende-se:

- Aplicar transformações complexas com variações paramétricas para gerar versões alternativas das imagens, destacando padrões locais e globais;
- Utilizar modelos de GANs como técnica comparativa de aumento de dados, a fim de avaliar sua eficácia relativa em relação às transformações complexas;
- Implementar uma arquitetura híbrida de classificação baseada na combinação de CNNs e ViT, explorando sua capacidade de representar múltiplas escalas de informação;
- Avaliar os resultados obtidos por cada estratégia e seus respectivos desempenhos na arquitetura híbrida, visando compreender as justificativas para tais conclusões e sugerir melhorias;
- Identificar os parâmetros mais relevantes das transformações complexas e os limites da abordagem híbrida.

Em complemento, esta proposta tem aderência em áreas prioritárias do Ministério da Ciência, Tecnologia, Inovações e Comunicações (estabelecidas na Portaria MCTIC nº 1.122/2020, com texto alterado pela Portaria MCTIC nº 1.329/2020), especificamente com as Áreas de Tecnologias para Qualidade de Vida (Saúde) e Tecnologias Habilitadoras (Inteligência Artificial). No que tange à lista dos Objetivos do Desenvolvimento Sustentável (ODS), esta pesquisa apresenta aderência direta com a ODS 9 (indústria, inovação e infraestrutura).

3 Metodologia

A pesquisa será conduzida em cinco etapas principais, conforme descritas a seguir.

3.1 Etapa 1 — Revisão Bibliográfica

Será realizada uma revisão sistemática da literatura sobre: (i) técnicas de aumento de dados baseadas em transformações matemáticas, com ênfase em transformações complexas; (ii) métodos de geração de dados sintéticos com GANs; e (iii) arquiteturas híbridas de classificação de imagens que integram CNNs e ViTs. Essa etapa fornecerá embasamento teórico para a definição e implementação dos métodos experimentais.

3.2 Etapa 2 — Geração de Dados com Transformações Complexas

Nesta etapa, será implementado um processo de aumento de dados baseado em transformações complexas, semelhante a abordagem proposta por Ribas et al. [5], porém com o enfoque em geração de dados ao invés de extração de características. A metodologia consiste em modelar cada imagem como uma rede complexa, onde cada pixel é representado por um vértice e as conexões (arestas) são estabelecidas com base em um critério de vizinhança radial.

Mais precisamente, dado um pixel $i = (x_i, y_i)$ em uma imagem I , ele é mapeado como um vértice $v_i \in V$. Dois vértices v_i e v_j são conectados por uma aresta se a distância entre eles for menor que um raio r definido:

$$E = \{\{v_i, v_j\} \mid d((x_i, y_i), (x_j, y_j)) < r\},$$

em que d representa a métrica de distância utilizada. Inicialmente, será empregada a distância Euclidiana, mas também será avaliada a substituição por outras métricas, como a distância de Minkowski (com diferentes valores de p), com o objetivo de verificar o impacto dessa variação na geração de padrões texturais.

O peso da aresta entre v_i e v_j é definido pela diferença absoluta entre as intensidades dos pixels:

$$w_{ij} = |I(x_i, y_i) - I(x_j, y_j)|.$$

A partir dessa modelagem, serão calculadas medidas topológicas locais, como grau e força dos vértices. Essas medidas serão então mapeadas de volta ao espaço da imagem, gerando novas versões da imagem original. Cada imagem transformada reflete uma configuração distinta de parâmetros da rede (como raio de conexão, tipo de métrica e limiar de aresta), permitindo a captura de padrões texturais em múltiplas escalas — raios menores realçam estruturas locais e raios maiores evidenciam padrões mais globais.

O pipeline de geração será composto pelas seguintes etapas: (i) Conversão de imagens originais para escala de cinza; (ii) Definição de um conjunto de valores de raio $r \in \{r_1, r_2, \dots, r_n\}$; (iii) Variação da métrica de distância (ex.: Euclidiana, Minkowski); (iv) Modelagem de cada imagem como uma rede complexa para cada combinação de parâmetros; (v) Cálculo das medidas topológicas dos vértices (grau, força); (vi) Mapeamento das medidas para o plano da imagem, resultando em novas imagens transformadas.

As imagens transformadas serão incorporadas ao conjunto de treinamento, compondo uma base de dados aumentada que será utilizada nas etapas seguintes de treinamento do modelo. A diversidade introduzida pelas diferentes parametrizações pretende enriquecer a capacidade do modelo de aprender representações robustas e sensíveis às variações de textura das imagens histológicas.

3.3 Etapa 3 — Geração de Dados com GANs

Nesta etapa, serão geradas imagens sintéticas por meio de redes adversariais generativas (GANs), com o objetivo de expandir o conjunto de dados de imagens histológicas e mitigar o desbalanceamento entre classes. O procedimento será implementado utilizando o framework PyTorch, com base em arquiteturas de GAN reconhecidas na literatura por sua estabilidade e capacidade de gerar imagens realistas em contextos médicos, conforme discutido por Ruiz-Casado et al. [6].

Inicialmente, será selecionada uma ou mais variantes de GAN com base em critérios como compatibilidade com o conjunto de dados, disponibilidade de implementação e viabilidade de treinamento. As redes serão treinadas com o conjunto original de imagens, e as amostras geradas serão avaliadas quanto à qualidade visual e diversidade estatística.

A métrica principal para avaliação será o *Fréchet Inception Distance* (FID) [7], que compara as distribuições estatísticas das representações extraídas por uma rede Inception-v3 entre as imagens reais e as geradas. O cálculo será realizado com o pacote `pytorch-fid`. Apenas amostras oriundas de modelos com FID satisfatório serão incorporadas ao conjunto de dados final. Adicionalmente, poderá ser realizada uma inspeção visual amostral para verificar a plausibilidade morfológica das imagens geradas.

As imagens sintéticas selecionadas serão rotuladas conforme a classe de origem e integradas ao conjunto de treinamento, garantindo o balanceamento entre as categorias. Esta abordagem servirá como base comparativa à técnica de aumento de dados por transformações complexas, permitindo avaliar o impacto relativo de diferentes estratégias de geração de dados na robustez do modelo híbrido de classificação proposto.

3.4 Etapa 4 — Treinamento do Modelo Híbrido (CNN + ViT)

Será utilizada uma arquitetura híbrida de classificação composta por módulos convolucionais (CNNs) e mecanismos de autoatenção baseados em Transformers (ViT), com o objetivo de explorar simultaneamente características locais e globais das imagens histológicas. Essa combinação busca integrar a capacidade das CNNs de extrair padrões espaciais de baixo nível com a habilidade dos Transformers em capturar dependências de longo alcance por meio da atenção.

A escolha da arquitetura seguirá diretrizes da literatura, considerando modelos que realizam a fusão dessas duas abordagens de forma eficiente, como ilustrado em trabalhos recentes que integram convoluções com mecanismos de atenção, incluindo ResNet-ViT, EfficientNet-ViT e varian-

tes baseadas em fusões adaptativas, como MedViT e outras propostas híbridas com acoplamento local-global [4, 8, 9].

Serão treinadas três versões distintas do modelo híbrido:

- Utilizando apenas o conjunto de dados aumentado por transformações complexas;
- Utilizando apenas o conjunto de dados aumentado por GANs;
- Utilizando o conjunto combinado das duas estratégias de aumento de dados.

O treinamento será conduzido com base em boas práticas da literatura de aprendizado profundo. A divisão do conjunto de dados será realizada de forma estratificada entre treino, validação e teste, garantindo o balanceamento das classes. Os hiperparâmetros serão ajustados por meio de validação cruzada, e serão utilizadas funções de perda apropriadas para tarefas de classificação multiclasse (como a *cross-entropy loss*). O otimizador Adam será empregado com configuração adequada da taxa de aprendizado, e serão aplicadas técnicas de regularização, como *dropout* e *early stopping*, para mitigar o risco de sobreajuste e melhorar a capacidade de generalização.

O desempenho dos modelos será avaliado com base nas métricas estabelecidas na etapa de análise de resultados, permitindo uma comparação objetiva entre as diferentes abordagens de aumento de dados aplicadas ao treinamento do modelo híbrido.

3.5 Etapa 5 — Avaliação e Comparação de Resultados

Serão utilizados os seguintes indicadores de desempenho: acurácia, precisão, *recall*, F1-score e AUC-ROC. A análise incluirá:

- Comparação quantitativa dos desempenhos obtidos pelos modelos treinados com diferentes estratégias de aumento de dados;
- Avaliação do impacto da variação dos parâmetros das transformações complexas no desempenho do modelo;
- Análise qualitativa das imagens sintéticas geradas pelas GANs;
- Discussão das limitações computacionais e da generalização dos modelos.

A análise estatística dos resultados será conduzida por meio de testes de significância apropriados (ex: teste de Wilcoxon para comparação pareada), com o objetivo de identificar diferenças relevantes entre as abordagens.

4 Plano de Trabalho e Cronograma de Execução

Este projeto é parte de outros projetos que estão em pleno desenvolvimento pelo orientador. O plano de trabalho consiste em realizar as etapas descritas na seção 3: Etapa 1 - Revisão Bibliográfica; Etapa 2 - Geração de Dados com Transformações Complexas; Etapa 3 - Geração de Dados com GANs; Etapa 4 - Treinamento do Modelo Híbrido (CNN + ViT); Etapa 5 - Avaliação e Comparação de Resultados.

5 Resultados Esperados

Espera-se que a aplicação das transformações complexas, com variações parametrizadas, contribua significativamente para a geração de imagens sintéticas capazes de enriquecer o conjunto de dados original, promovendo uma melhoria mensurável na acurácia da classificação de imagens médicas. A comparação com a técnica de aumento de dados baseada em GANs deverá evidenciar o potencial das transformações complexas como alternativa viável e eficaz. Adicionalmente, o uso de uma arquitetura híbrida de classificação, combinando CNNs e ViT, deverá apresentar bom desempenho na extração de características locais e globais, visando desfrutar dos benefícios das transformações com variações de parâmetros. Espera-se, ainda, identificar os parâmetros mais

Tabela 1: Cronograma para o atendimento das etapas, considerando o período de 12 meses.

Etapas	Ano 1: 2025-2026											
	1	2	3	4	5	6	7	8	9	10	11	12
1	x		x		x		x		x			
2	x	x	x									
3				x	x							
4					x	x						
5							x	x	x	x		
Escrita: Processos e Resultados		x		x		x		x		x	x	x

influentes nas transformações complexas e levantar possíveis limitações da abordagem híbrida adotada, fornecendo direções para aprimoramentos futuros e contribuindo para o avanço de técnicas de classificação automatizada no campo da saúde.

Referências

- [1] X. Wu, Y. Feng, H. Xu, Z. Lin, T. Chen, S. Li, S. Qiu, Q. Liu, Y. Ma, and S. Zhang, “Ctrascnn: Combining transformer and cnn in multilabel medical image classification,” *Knowledge-Based Systems*, vol. 281, p. 111030, 2023.
- [2] L. C. Ribas, L. F. Scabini, R. H. Condori, and O. M. Bruno, “Color-texture classification based on spatio-spectral complex network representations,” *Physica A: Statistical Mechanics and its Applications*, vol. 635, p. 129518, 2024.
- [3] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [4] Z. Peng, W. Huang, S. Gu, L. Xie, Y. Wang, J. Jiao, and Q. Ye, “Conformer: Local features coupling global representations for visual recognition,” 2021.
- [5] L. C. Ribas, J. J. d. M. S. Junior, L. F. Scabini, and O. M. Bruno, “Fusion of complex networks and randomized neural networks for texture analysis,” *Pattern Recognition*, vol. 103, p. 107189, 2020.
- [6] J. L. Ruiz-Casado, M. A. Molina-Cabello, and R. M. Luque-Baena, “Enhancing histopathological image classification performance through synthetic data generation with generative adversarial networks,” *Sensors*, vol. 24, no. 12, p. 3777, 2024.
- [7] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a local nash equilibrium,” 2018.
- [8] T. Hussain, H. Shouno, A. Hussain, *et al.*, “Effresnet-vit: A fusion-based convolutional and vision transformer model for explainable medical image classification,” *IEEE Access*, vol. 13, pp. 54041–54066, 2025.
- [9] X. Chen, R. Li, Z. Wang, and T. Wang, “Medvit: Hybrid cnn-transformer architecture for robust medical image diagnosis,” *Medical Image Analysis*, vol. 87, p. 102829, 2023.