

# Extração de Deep Features via Autoencoders Variacionais para Análise de Imagens Histológicas

Eduardo Paraíso<sup>1</sup>

<sup>1</sup>Instituto de Ciências Exatas e Informática  
Pontifícia Universidade Católica de Minas Gerais  
Belo Horizonte, MG – Brasil

eduardo.paraíso@sga.pucminas.br

**Abstract.** *This work investigates the use of Variational Autoencoders to extract deep features from histological images of prostate cancer, focusing on the analysis of the learned latent space. Using the DiagSet-A dataset, the study implements an encoder architecture based on ResNet50. After training, analyses were conducted using PCA, t-SNE, K-Means, and DBSCAN. The results showed that the learned latent space exhibited continuity and coherence but low discriminative capacity among Gleason patterns, with class overlap and predominant clustering into two macrostructures. Interpolation revealed smooth transitions, though with subtle morphological variations. The study concludes that incorporating additional morphological information and using supervised models may improve the latent representation and its applicability to clinical tasks.*

**Resumo.** *Este trabalho investiga o uso de Autoencoders Variacionais para extrair deep features de imagens histológicas de câncer de próstata, com foco na análise do espaço latente aprendido. Utilizando o conjunto de dados DiagSet-A, o estudo implementa uma arquitetura baseada na ResNet50 para o encoder. Após treinamento, foram realizadas análises com PCA, t-SNE, K-Means e DBSCAN. Os resultados mostraram que o espaço latente aprendido apresentou continuidade e coerência, mas baixa capacidade discriminativa entre os padrões de Gleason, com sobreposição entre classes e agrupamento predominante em duas macroestruturas. A interpolação indicou transições suaves, mas variações morfológicas sutis. O estudo conclui que a inclusão de informações morfológicas adicionais e o uso de modelos supervisionados podem melhorar a representação latente e sua aplicabilidade em tarefas clínicas.*

## 1. Introdução

A análise automatizada de imagens histológicas desempenha um papel cada vez mais relevante na patologia digital, oferecendo suporte ao diagnóstico clínico por meio de técnicas de aprendizado de máquina [Dalmedico et al., 2024]. A crescente disponibilidade de grandes bancos de dados de imagens biomédicas, aliada aos avanços em inteligência artificial, permitiu o desenvolvimento de métodos capazes de extrair características discriminativas de tecidos com alta acurácia. Entre esses métodos, destaca-se o uso de representações latentes aprendidas por redes neurais profundas.

Os autoencoders variacionais (VAEs, do inglês *Variational Autoencoders*) têm se consolidado como uma abordagem promissora para aprendizado não supervisionado de

representações compactas e semânticas de imagens [Academy, 2025]. Diferentemente dos autoencoders clássicos, os VAEs impõem uma estrutura probabilística sobre o espaço latente, permitindo não apenas a compressão da informação, mas também a geração de novos dados e a modelagem de incertezas. Essa característica é particularmente útil em contextos biomédicos, nos quais a variabilidade morfológica entre tecidos pode ser sutil e complexa.

Neste trabalho, foi investigado a aplicação de autoencoders variacionais para a extração de deep features a partir de imagens histológicas, com o objetivo de avaliar a capacidade dessas representações em capturar padrões relevantes para tarefas subsequentes, como classificação, clusterização ou segmentação. Exploramos diferentes arquiteturas de encoder e analisamos o impacto do espaço latente na qualidade das representações extraídas. Os experimentos são realizados sobre um conjunto de dados histológicos anotados, permitindo a validação quantitativa e qualitativa dos resultados.

## **2. Trabalho Relacionados**

Estudos recentes em patologia computacional têm se concentrado no desenvolvimento de modelos de aprendizado profundo para a análise automatizada de imagens histológicas da próstata, com especial à classificação de padrões de Gleason, segmentação precisa e síntese de imagens interpretáveis.

O desenho e avaliação de arquiteturas de aprendizado profundo voltadas à classificação automatizada de Gleason, possuem abordagens que exploram redes convolucionais, transformadores visuais e modelos híbridos, como ConvNext, EfficientNet e Vision Mamba. Diversos trabalhos recentes como os de Malekmohammadi et al. [2024]; Müller et al. [2024]; Bhattacharyya et al. [2024] e Karthik et al. [2024], destacam o impacto dessas arquiteturas em precisão, escalabilidade e eficiência computacional, incluindo aplicações em imagens de lâminas.

Para contornar a limitação de dados rotulados, muitos estudos adotam estratégias de aprendizado autosupervisionado ou fracamente supervisionado. Essas abordagens envolvem autoencoders, aprendizado contrastivo, múltiplas instâncias e rótulos em nível de slide. Tabatabaei et al. [2023] e Silva-Rodriguez et al. [2021]; Silva-Rodríguez et al. [2021] mostram que essas técnicas mantêm a capacidade discriminativa dos modelos mesmo sob condições de anotação escassa ou ruidosa.

A integração de dados multimodais tem sido empregada para melhorar a robustez diagnóstica e a predição da agressividade tumoral. Estudos como os de Mahapatra et al. [2021] e Rodrigues et al. [2024] combinam informações de histopatologia com ressonância magnética ou radiômica, por meio de técnicas como autocodificadores condicionais e fusão de características entre domínios.

Outro foco importante está na localização espacial de padrões de Gleason por segmentação, com o uso de variantes do U-Net, SegFormer e mecanismos de atenção para identificar regiões específicas com padrões tumorais distintos. Abordagens como as de Chatrian et al. [2019] e Vats et al. [2024] aumentam a interpretabilidade dos modelos ao gerar mapas de segmentação alinhados com o raciocínio clínico.

A variabilidade interobservador e a heterogeneidade nos dados histológicos representam desafios adicionais. Para lidar com isso, alguns estudos propõem normalização de

coloração, treinamento de modelos independentes de domínio e uso de rótulos flexíveis que capturam a incerteza diagnóstica. Trabalhos como o de Mittmann et al. [2024] e Khalid et al. [2024] demonstram que essas estratégias ajudam a melhorar a robustez e a confiança clínica dos modelos.

Alguns modelos mais recentes exploram classificação hierárquica ou pipelines em múltiplos estágios, que integram segmentação preliminar com classificadores binários e multiclasse, ou redes hierárquicas com atenção. Essas estratégias visam capturar transições morfológicas complexas e melhorar a precisão entre padrões Gleason adjacentes, como exemplificado em Hammouda et al. [2022] ou Butt et al. [2024].

Por fim, algumas abordagens recentes propõem o uso de funções de perda especializadas e estratégias de treinamento robustas, como perdas focais, amostragem difícil e destilação de conhecimento. Essas técnicas, como demonstrado em Karthik et al. [2024], têm mostrado melhorar a discriminação entre padrões de Gleason morfológicamente próximos e a estabilidade do treinamento.

Em conjunto, esses estudos demonstram avanços significativos na busca por modelos acurados, interpretáveis e treináveis em cenários de dados limitados, embora ainda persistam desafios em padronização de anotações, validação externa e integração nos fluxos clínicos reais.

### **3. Referencial Teórico**

#### **3.1. Deep Features**

O conceito de deep features refere-se às representações extraídas automaticamente por redes neurais profundas a partir dos dados brutos, como imagens. Diferentemente das características projetadas manualmente, essas representações são aprendidas de forma hierárquica, capturando tanto padrões locais quanto estruturas de alto nível. Em contextos biomédicos, as deep features têm se mostrado altamente eficazes em tarefas como classificação de tipos celulares, detecção de câncer e segmentação de estruturas anatômicas [KiRaz et al., 2023].

#### **3.2. Autoencoders e Autoencoders Variacionais**

Autoencoders [Yang, 2024] são redes neurais não supervisionadas projetadas para aprender uma codificação eficiente dos dados de entrada. A arquitetura clássica é composta por duas partes principais: o encoder, que comprime os dados em um espaço latente de menor dimensão, e o decoder, responsável por reconstruir a entrada a partir dessa codificação. O objetivo do treinamento é minimizar a diferença entre os dados originais e os reconstruídos, promovendo a aprendizagem de representações compactas.

Os VAE, introduzidos por Kingma e Welling [2013], incorporam princípios da inferência bayesiana ao modelo de autoencoder. Em vez de aprender uma representação fixa para cada entrada, o VAE modela a distribuição probabilística dos dados no espaço latente, geralmente assumida como uma normal multivariada. O encoder estima os parâmetros (média e variância) dessa distribuição, enquanto o decoder reconstrói a entrada a partir de amostras latentes extraídas dessa distribuição.

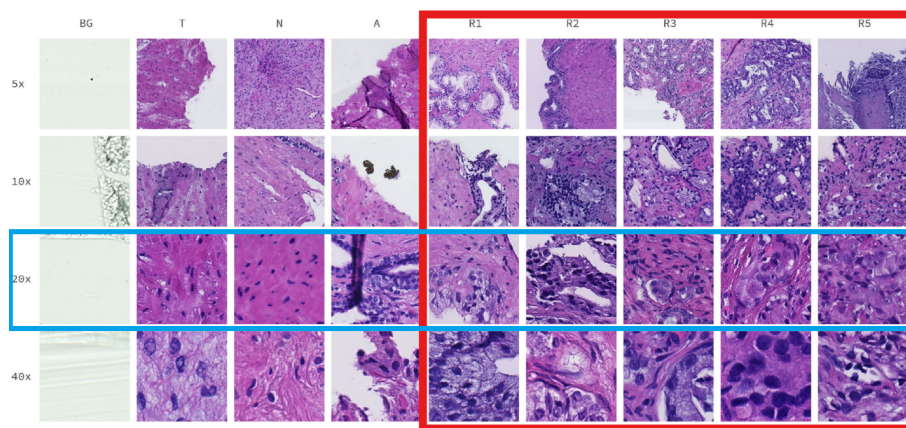
A aplicação de VAEs em imagens médicas tem ganhado destaque na literatura, com resultados promissores em contextos como geração de imagens sintéticas, detecção

de anomalias, segmentação e classificação [Rais et al., 2024]. Em imagens histológicas, os VAEs podem capturar variações sutis na morfologia tecidual que seriam difíceis de detectar por métodos tradicionais. Estudos recentes demonstram que as representações latentes extraídas por VAEs podem ser utilizadas como entradas para classificadores ou analisadas por métodos não supervisionados para identificação de subgrupos patológicos.

## 4. Materiais e Métodos

### 4.1. Conjunto de Dados

O conjunto de dados DiagSet-A, proposto por Koziarski et al. [2024], é voltado à detecção de câncer de próstata a partir de imagens histopatológicas. Ele é composto por aproximadamente 2,6 milhões de patches extraídos com stride fixo de 430 lâminas digitalizadas e totalmente anotadas por patologistas especialistas. As amostras foram geradas em quatro níveis de ampliação (5x, 10x, 20x e 40x); contudo, neste trabalho, foram utilizados exclusivamente patches com magnificação de 20x, a fim de preservar os detalhes morfológicos relevantes e manter a consistência na resolução das imagens.



**Figura 1. Amostragem de imagens por classe e por nível de magnificação da objetiva. Foram selecionadas apenas imagens representativas do câncer de próstata (tecido normal e padrões de Gleason), destacadas em vermelho. Em azul, a magnificação escolhida para este estudo.**

Para contemplar a heterogeneidade dos padrões histológicos e garantir representatividade das diferentes classes, foram consideradas apenas as seguintes categorias: Normal, R1, R2, R3, R4 e R5 (tecido benigno e os diferentes padrões de agressividade do câncer de próstata). De cada uma dessas classes, foram amostrados aproximadamente 5.000 patches, totalizando cerca de 30.000 imagens. Esse procedimento teve como objetivo reduzir o desbalanceamento de classes presente na distribuição original do dataset, que é composta por aproximadamente 35,3% de tecido não canceroso, 0,7% para os padrões R1 e R2, 6,1% para R3, 15,4% para R4 e 3,0% para R5.

### 4.2. Pré-processamento

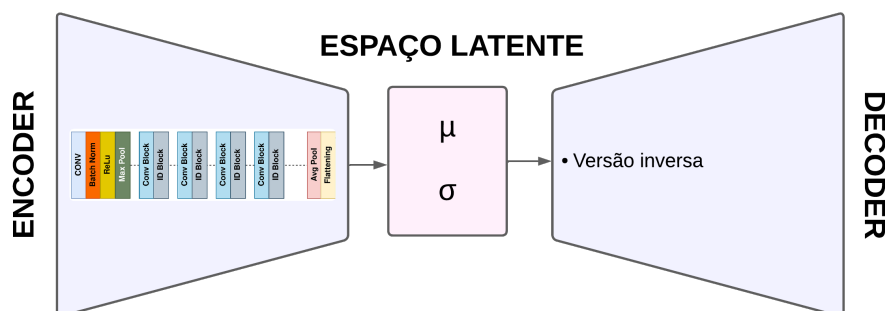
Os arquivos disponibilizados estão no formato NumPy (.npy), organizados em quatro dimensões {N, 256, 256, 3}, onde N representa o número de patches armazenados no arquivo, e cada patch é uma imagem RGB de 256×256 pixels.

Para garantir padronização na entrada dos modelos e compatibilidade com arquiteturas convolucionais pré-treinadas na base de dados ImageNet, as imagens foram individualizadas e redimensionadas para  $224 \times 224$  pixels. Esse tamanho é adotado por padrão em diversas redes amplamente utilizadas, como ResNet-50, DenseNet e InceptionV3, cujas camadas iniciais foram otimizadas para esse formato. Essa adaptação permite o uso direto de pesos pré-treinados, acelerando a convergência do treinamento e, frequentemente, melhorando a generalização do modelo, especialmente em cenários como em imagens médicas.

Além do redimensionamento, os valores dos pixels foram normalizados para o intervalo  $[0, 1]$ , convertendo a escala original  $[0, 255]$  para ponto flutuante. Essa normalização é uma etapa essencial para estabilizar o treinamento de redes neurais profundas, pois reduz a variância das ativações e gradientes entre as camadas, facilitando o processo de otimização com métodos baseados em gradiente.

### 4.3. Arquitetura

Para a modelagem do espaço latente, foi implementada uma VAE, cuja arquitetura de codificação foi baseada na ResNet50, previamente treinada na base ImageNet. A utilização desse backbone permitiu a extração de representações visuais de alto nível, mantendo características morfológicas relevantes das imagens histopatológicas.



**Figura 2.** O Autoencoder Variacional é composto por três partes: a codificadora (encoder), o espaço latente e a decodificadora (decoder). Na etapa de codificação, foi empregada uma ResNet50 pré-treinada na ImageNet. O espaço latente é definido por dois vetores:  $\mu$  (média) e  $\log \sigma^2$  (logaritmo da variância), que representam a distribuição probabilística das variáveis latentes. Na etapa de decodificação, foram utilizadas camadas convolucionais transpostas para reconstruir a imagem original.

Diversas arquiteturas de codificador foram avaliadas durante a fase experimental; no entanto, os melhores desempenhos foram obtidos com a ResNet50, motivo pelo qual foi adotada na versão final do modelo. A etapa de decodificação da VAE foi composta por camadas convolucionais transpostas (transposed convolutions), organizadas de modo a reconstruir as imagens no mesmo formato original das amostras de entrada, ou seja,  $224 \times 224 \times 3$  pixels.

O modelo foi implementado na biblioteca PyTorch e treinado com batch size igual a 128, utilizando o otimizador Adam, com taxa de aprendizado inicial de  $10^{-3}$  e espaço latente de dimensão 128. O treinamento foi planejado para ocorrer ao longo de 100 épocas; contudo, observou-se estabilização precoce das métricas de reconstrução e

perda de evidência variacional (ELBO) a partir da 36ª época, o que motivou a interrupção do treinamento na 40ª época.

Os experimentos foram conduzidos em ambiente Google Colab, com suporte de GPU NVIDIA Tesla T4, aproveitando aceleração por hardware para otimização do tempo de treinamento.

#### 4.4. Função de Custo

A função de custo adotada na VAE foi composta por dois termos: a Binary Cross-Entropy (BCE) e a Kullback–Leibler Divergence (KL). O termo de BCE [Ramos et al., 2018] foi utilizado para quantificar a diferença entre a imagem original e a reconstruída, atuando como medida de erro de reconstrução. Esse termo avalia a similaridade entre os pixels correspondentes das duas imagens, considerando a natureza binária ou normalizada dos valores de entrada no intervalo  $[0, 1]$ .

A incorporação da KL [Kullback, 1959] tem como objetivo a regularização do espaço latente, forçando a distribuição aprendida  $q(z|x)$  a se aproximar de uma distribuição normal padrão  $\mathcal{N}(0, 1)$ . Essa regularização é essencial para garantir a continuidade e estruturação do espaço latente, permitindo amostragem eficiente e interpolação coerente entre representações latentes.

A função de perda total  $\mathcal{L}_{VAE}$  é, portanto, definida como:

$$\mathcal{L}_{VAE} = \mathcal{L}_{BCE}(x, \hat{x}) + \beta \cdot \mathcal{D}_{KL}(q(z|x) \parallel p(z)) \quad (1)$$

onde  $x$  representa a imagem original,  $\hat{x}$  a imagem reconstruída,  $z$  o vetor latente,  $p(z) = \mathcal{N}(0, I)$  a distribuição alvo, e  $\beta$  um hiperparâmetro de ponderação.

Para evitar o colapso precoce do espaço latente e facilitar uma transição gradual para a regularização probabilística, foi aplicado um esquema de KL annealing, no qual o peso do termo de divergência KL foi aumentado progressivamente a cada 5 épocas.

#### 4.5. Extração e Análises

Após o treinamento do modelo proposto, procedeu-se à análise das representações geradas no espaço latente. Especificamente, foram extraídos os vetores de média ( $\mu$ ) da distribuição latente para cada amostra do conjunto de dados. Essas representações foram utilizadas como base para uma análise exploratória com o objetivo de investigar a organização, separabilidade e estrutura interna do espaço latente aprendido pelo modelo. Inicialmente, aplicaram-se técnicas de redução de dimensionalidade para projetar os vetores latentes de alta dimensão em espaços de duas e três dimensões. Foram utilizadas duas abordagens complementares:

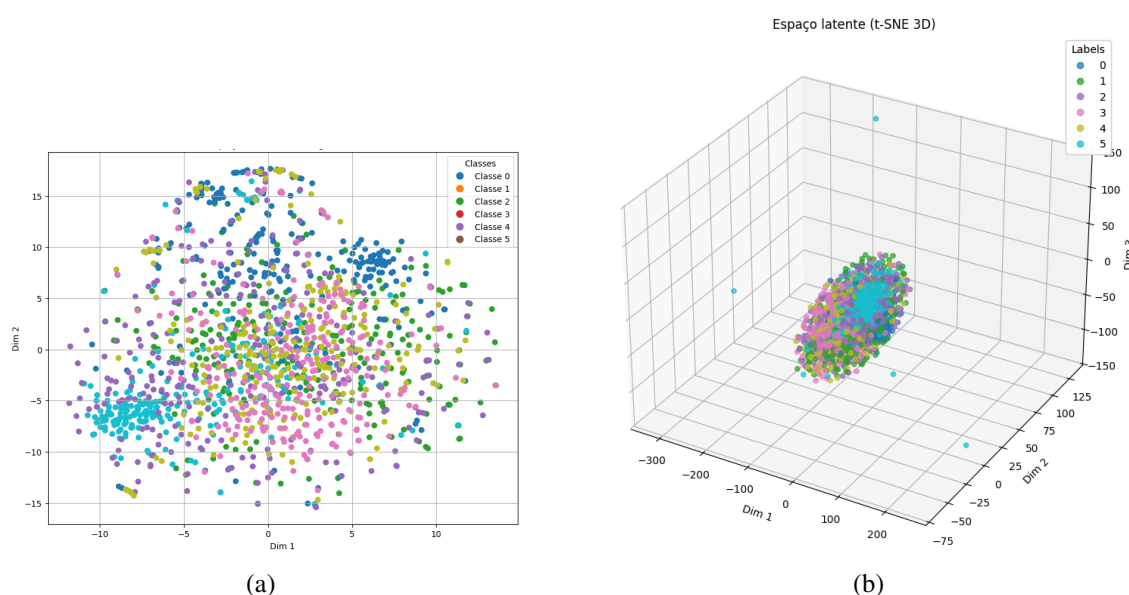
- Principal Component Analysis (PCA): permite preservar a variância global dos dados. O PCA foi empregado tanto em duas (2D) quanto em três dimensões (3D), visando compreender a distribuição geral dos pontos latentes e avaliar a presença de direções de máxima variação.
- t-Distributed Stochastic Neighbor Embedding (t-SNE): orientada à preservação de vizinhanças locais. O t-SNE foi utilizado em 2D e 3D para revelar possíveis estruturas locais, agrupamentos e padrões que não seriam facilmente identificáveis por métodos lineares.

Em seguida, com o intuito de investigar a presença de agrupamentos latentes nos dados codificados, foram aplicadas técnicas de clusterização não supervisionada:

- K-Means: algoritmo baseado na minimização da distância intra-cluster. Diversos valores de  $k$  foram testados.
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): método baseado em densidade que permite identificar clusters de forma arbitrária e detectar pontos considerados ruído.

## 5. Resultados

### 5.1. Redução de Dimensionalidade



**Figura 3. (a) Projeção bidimensional direta dos vetores latentes, onde observa-se uma distribuição aproximadamente normal centrada na origem, com alta sobreposição entre as classes (tecido saudável e padrões de Gleason 1 a 5), indicando ausência de estrutura semântica discriminativa. Em (b) projeção tridimensional via t-SNE, que preserva relações locais, mas também revela forte sobreposição entre os rótulos, sem formação de agrupamentos distintos. Ambas as visualizações sugerem que o espaço latente, apesar de contínuo e coerente, não organiza os dados de forma alinhada aos padrões histológicos.**

A projeção do espaço latente utilizando t-SNE, apresentada na Figura 3(a), evidencia uma significativa sobreposição entre os diferentes grupos de padrões histológicos, com todos os rótulos concentrados de forma densa e misturada em uma única região central. Essa distribuição indica a ausência de agrupamentos claros ou separações definidas, o que sugere que o espaço latente não conseguiu capturar de maneira eficaz as variações morfológicas necessárias para discriminar entre os graus de Gleason.

Diversos fatores podem contribuir para esse resultado. Características visuais semelhantes entre padrões próximos, como Gleason 3 e 4, dificultam a distinção. Além disso, a heterogeneidade dos dados de entrada pode confundir o modelo. A utilização

de um modelo não supervisionado, treinado sem orientações explícitas para organizar semanticamente o espaço latente, também limita sua capacidade discriminativa. Por fim, as limitações inerentes ao t-SNE, que preserva preferencialmente relações locais e pode distorcer as relações globais entre classes, devem ser consideradas.

A presença de outliers distantes na projeção sugere amostras atípicas (Figura 3(b)), possivelmente decorrentes de erros de rotulagem, imagens ruidosas ou exemplos raros, que não são bem representados no espaço latente.

Observa-se que apenas uma pequena região da projeção, localizada no canto inferior esquerdo, apresenta uma formação mais coesa e aparentemente dominada por uma única classe (Classe 0). As demais classes estão distribuídas de maneira difusa e intercalada, evidenciando a alta sobreposição. Essa sobreposição intensa pode refletir fronteiras morfológicas pouco definidas entre os graus de Gleason, bem como a extração de características que não são suficientemente discriminativas, como texturas e formas. Além disso, a elevada variabilidade interna em classes específicas, como a Classe 3, dificulta ainda mais a separação clara dos grupos na redução de dimensionalidade.

## 5.2. Interpolação do Espaço Latente

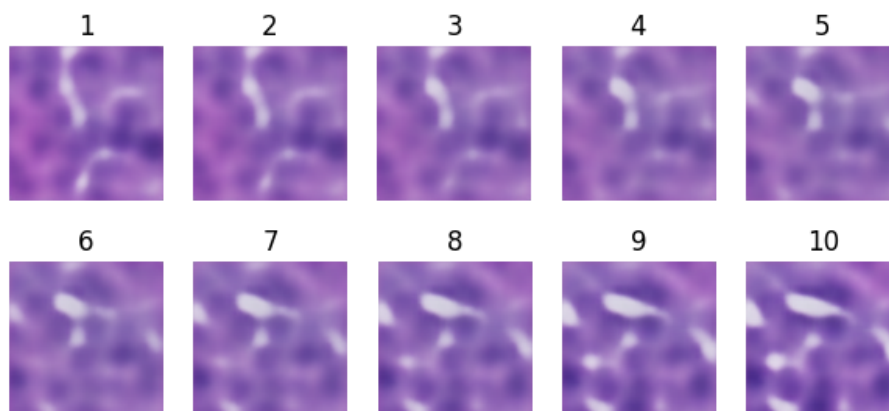
Ao realizarmos interpolações no espaço latente (Figura 4), observamos uma progressão contínua e gradual de padrões morfológicos presentes nas imagens geradas, destacando-se especialmente a distribuição das áreas claras, que provavelmente correspondem a lúmens glandulares, e a textura do fundo, que pode representar o estroma ou as células epiteliais circundantes. Nas primeiras imagens da sequência (1 a 4), as formas são pouco definidas, apresentando regiões claras isoladas, de contornos difusos e dispersos, o que sugere menor organização estrutural. A partir da imagem 5, nota-se que essas regiões claras começam a se alongar e a se orientar em uma direção comum, indicando o início de uma transição morfológica coerente e contínua. Nas imagens finais da interpolação (8 a 10), os lúmens parecem estar mais alongados e organizados em padrões alinhados, sugerindo a presença de estruturas glandulares mais desenvolvidas e bem definidas.

Esse processo de interpolação é uma ferramenta valiosa para avaliar a coerência e a continuidade do espaço latente aprendido pelo modelo. No entanto, é importante ressaltar que a interpolação por si só não confirma a capacidade do modelo em distinguir ou separar classes distintas sem a realização de análises quantitativas ou inspeção visual criteriosa de amostras representativas de cada categoria. O que fica evidenciado é que o espaço latente é contínuo, sem apresentar descontinuidades abruptas ou saltos repentinos entre diferentes padrões — uma característica fundamental em modelos generativos, como os VAEs, pois indica que o modelo aprendeu uma representação latente suave e consistente. Essa transição gradual entre os diferentes padrões morfológicos mostra que o modelo aprendeu a mapear as variações observadas nas imagens de forma coerente, evitando saltos bruscos que indicariam descontinuidades no espaço latente.

Ainda assim, as variações morfológicas apresentadas nas imagens interpoladas são sutis. De maneira geral, as imagens mantêm uma aparência visual semelhante ao longo da interpolação, o que sugere algumas hipóteses importantes sobre a representação latente aprendida pelo modelo:

- O espaço latente parece capturar principalmente variações relacionadas à textura e coloração das imagens, mais do que diferenças estruturais morfológicas profun-





**Figura 4.** Cada imagem representa um ponto intermediário obtido por interpolação linear entre dois vetores latentes correspondentes a imagens reais. A transição gradual ao longo da sequência reflete mudanças suaves nas características morfológicas codificadas pelo modelo. A continuidade visual sugere que o espaço latente possui uma estrutura semântica coerente, capaz de gerar imagens plausíveis mesmo para combinações intermediárias.

das ou qualitativas, que seriam indicativas de padrões histológicos distintamente diferentes.

- Essa limitação pode indicar que o modelo enfrenta dificuldades para representar diferenças qualitativas significativas entre padrões histológicos distintos. Tal limitação pode decorrer tanto da capacidade limitada do modelo em aprender representações altamente discriminativas quanto da possível homogeneidade ou insuficiência da diversidade dos dados de entrada utilizados para treinamento.

Portanto, apesar da continuidade e suavidade no espaço latente serem indicativos positivos de uma boa modelagem das variações internas dos dados, a ausência de mudanças morfológicas mais marcantes sugere que outras abordagens, ajustes no modelo ou aumento da diversidade dos dados podem ser necessários para melhorar a capacidade discriminativa entre classes histológicas distintas.

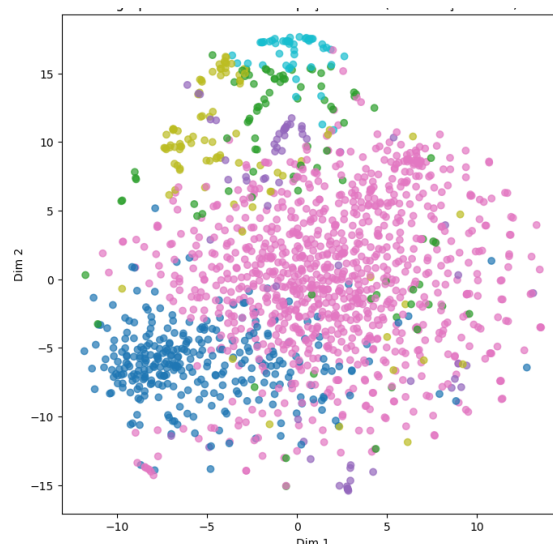
### 5.3. Agrupamento do Espaço Latente

A análise dos agrupamentos no espaço latente revelou uma separação clara em dois grandes grupos predominantes. Um grupo, composto majoritariamente por pontos na cor azul, localiza-se na porção inferior esquerda do mapa. O outro grupo, de cor rosa, ocupa uma região central mais dispersa. As demais categorias, representadas por cores como verde, ciano, roxo e amarelo, aparecem em regiões pequenas e esparsas, predominantemente nas extremidades superiores do espaço.

O algoritmo KMeans foi aplicado com  $n_{clusters} = 6$ , resultando na formação de seis agrupamentos. No entanto, observou-se que apenas dois clusters — azul e rosa — dominaram a maior parte da amostragem, enquanto os demais se concentraram em subconjuntos menores. Esse comportamento sugere que o espaço latente aprendido pela rede está capturando duas macroestruturas principais nos dados.

Uma possível explicação para essa divisão é que o espaço latente esteja representando categorias amplas, como tecido saudável (Classe 0) versus tecido tumoral (Classes

1 a 5), ou ainda uma separação entre padrões de baixa e alta agressividade. Nesse cenário, o KMeans tende a alocar os dois maiores clusters para essas macroclasses, forçando os demais a subdividirem variações morfológicas menores entre as classes patológicas.



**Figura 5. Projeção bidimensional dos vetores latentes extraídos, com coloração indicando os rótulos atribuídos pelos clusters do K-Means. Observa-se a formação de dois grupos principais: um grupo compacto à esquerda (azul) e outro mais disperso no centro-direita (rosa), enquanto os demais clusters aparecem como regiões menores e esparsas nas extremidades. A separação entre os grupos ainda não corresponda claramente às classes morfológicas esperadas.**

A elevada concentração de pontos rosa no centro do espaço pode indicar uma sobreposição entre as classes patológicas (1 a 5), sugerindo que o espaço latente apresenta baixa granularidade entre os diferentes padrões de Gleason. Dado que o KMeans pressupõe clusters esféricos, de tamanho semelhante e variância homogênea, essa limitação estrutural dificulta a separação eficiente entre as classes tumorais, levando à formação de um cluster único e dominante (rosa), que abrange múltiplas categorias com características morfológicas semelhantes ou confundidas.

Foi também realizada uma tentativa de agrupamento utilizando o algoritmo DBSCAN no mesmo espaço latente. Contudo, o DBSCAN não conseguiu identificar clusters significativos, retornando apenas uma única classe para todos os pontos. Esse resultado pode ser explicado pela densidade e distribuição dos dados no espaço latente: como os pontos estão muito próximos uns dos outros e apresentam pouca variação local de densidade, o DBSCAN, que depende de regiões densas e bem definidas para formar agrupamentos, acaba classificando a maioria dos pontos como pertencentes a um único grupo. Além disso, parâmetros como o valor de epsilon e o número mínimo de pontos podem não estar adequadamente ajustados para o formato e densidade do espaço latente, dificultando a separação em múltiplos clusters. Essa situação reforça a hipótese de que o espaço latente aprendido não apresenta fronteiras claras ou densidades diferenciadas suficientes para a segmentação eficaz por métodos baseados em densidade.

## 6. Considerações Finais

Os resultados indicam que a incorporação de informações morfológicas adicionais ao input, como texturas, mapas de segmentação glandular ou embeddings provenientes de redes pré-treinadas, pode constituir uma estratégia promissora para enriquecer a representação no espaço latente. Ademais, a adoção de modelos supervisionados ou VAE condicionais, que alinham o aprendizado do espaço latente com os rótulos clínicos, surge como um caminho relevante para aprimorar a capacidade discriminativa do modelo.

Embora o espaço latente apresente uma organização coerente, com características de suavidade e continuidade, não foi observada evidência clara de discriminação entre os diferentes padrões de Gleason na configuração atual. Essa limitação restringe sua aplicabilidade direta em diagnósticos, indicando a necessidade de intervenções supervisionadas ou da incorporação de segmentações específicas para potencializar sua utilidade na análise de padrões complexos, como os de Gleason. Em consequência, as deep features extraídas do espaço latente não foram utilizadas em classificadores, devido à ausência dessa discriminação evidente entre os padrões.

## Disponibilidade do Material

Os códigos e resultados desenvolvidos neste trabalho podem ser obtidos no repositório do projeto, acessando: [https://github.com/eduardoparaíso/I2DL\\_visao](https://github.com/eduardoparaíso/I2DL_visao). O conjunto Diagset pode ser obtido em: <https://ai-econsilio.diag.pl/>.

## Referências

- D. S. Academy. *Deep Learning Book*. 2025. URL <https://www.deeplearningbook.com.br/>.
- R. Bhattacharyya, P. Roy, S. Banerji, e S. Mitra. Efficient grading of prostate cancer WSI with deep learning. In *Medical Imaging 2024: Digital and Computational Pathology*, page 4, San Diego, United States, Apr. 2024. SPIE. doi: 10.1117/12.3006835.
- M. A. Butt, M. F. Kaleem, M. Bilal, e M. S. Hanif. Using multi-label ensemble CNN classifiers to mitigate labelling inconsistencies in patch-level Gleason grading. *PLOS ONE*, 19(7):e0304847, July 2024. doi: 10.1371/journal.pone.0304847.
- A. Chatrian, K. Sirinukunwattana, C. Verrill, e J. Rittscher. Towards the Identification of Histology Based Subtypes in Prostate Cancer. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 948–952, Venice, Italy, Apr. 2019. IEEE. doi: 10.1109/isbi.2019.8759199.
- M. M. Dalmedico, C. K. L. D. Carvalho, P. P. D. Oliveira, P. K. Hembecker, J. L. S. Ávila, e S. O. Ioshii. Patologia digital e computacional: um panorama atual e perspectivas futuras. *CONTRIBUCIONES A LAS CIENCIAS SOCIALES*, 17(4):e6027, Apr. 2024. doi: 10.55905/revconv.17n.4-103.
- K. Hammouda, F. Khalifa, N. S. Alghamdi, H. Darwish, e A. El-Baz. Multi-Stage Classification-Based Deep Learning for Gleason System Grading Using Histopathological Images. *Cancers*, 14(23):5897, Nov. 2022. doi: 10.3390/cancers14235897.
- R. Karthik, R. Menaka, M. V. Siddharth, S. Hussain, P. Siddharth, e D. Won. HMAR-NET—A Hierarchical Multi-Attention Residual Network for Gleason scoring of prostate cancer. *International Journal of Imaging Systems and Technology*, 34(1), Jan. 2024. doi: 10.1002/ima.22976.

- U. Khalid, J. Gurung, M. Doykov, G. Kostov, B. Hristov, P. Uchikov, M. Kraeva, K. Kraev, D. Doykov, K. Doykova, S. Valova, L. Chervenkov, E. Tilkiyan, e K. Eneva. Artificial Intelligence Algorithms and Their Current Role in the Identification and Comparison of Gleason Patterns in Prostate Cancer Histopathology: A Comprehensive Review. *Diagnostics*, 14(19):2127, Sept. 2024. doi: 10.3390/diagnostics14192127.
- D. P. Kingma e M. Welling. Auto-Encoding Variational Bayes, 2013.
- A. H. KiRaz, F. O. Djibrillah, e M. E. Yüksel. Deep feature extraction, dimensionality reduction, and classification of medical images using combined deep learning architectures, autoencoder, and multiple machine learning models. *Turkish Journal of Electrical Engineering and Computer Sciences*, 31(6):1113–1128, Oct. 2023. doi: 10.55730/1300-0632.4037.
- M. Koziarski, B. Cyganek, P. Niedziela, B. Olborski, Z. Antosz, M. Żydak, B. Kwolek, P. Wasowicz, A. Buwała, J. Swadźba, e P. Sitkowski. DiagSet: a dataset for prostate cancer histopathological image classification. *Scientific Reports*, 14(1), Mar. 2024. doi: 10.1038/s41598-024-52183-4.
- S. Kullback. *Information theory and statistics*. 1959.
- D. Mahapatra, S. Kuanar, B. Bozorgtabar, e Z. Ge. Self-supervised Learning of Inter-label Geometric Relationships for Gleason Grade Segmentation. In *Lecture Notes in Computer Science*, pages 57–67. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-87722-4\_6.
- A. Malekmohammadi, A. Badiezadeh, S. M. Mirhassani, P. Gifani, e M. Vafaezadeh. Classification of Gleason Grading in Prostate Cancer Histopathology Images Using Deep Learning Techniques: YOLO, Vision Transformers, and Vision Mamba, 2024.
- G. Mittmann, S. Laiouar-Pedari, H. A. Mehrstens, S. Haggemüller, T.-C. Bucher, T. Chanda, N. T. Gaisa, M. Wagner, G. G. Klamming, T. T. Rau, C. Nepl, E. M. Compérat, A. Gocht, M. Hämmerle, N. J. Rupp, J. Westhoff, I. Krücken, M. Seidl, C. M. Schürch, M. Bauer, W. Solass, Y. C. Tam, F. Weber, R. Grobholz, J. Augustyniak, T. Kalinski, C. Hörner, K. D. Mertz, C. Döring, A. Erbersdobler, G. Deubler, F. Bremmer, U. Sommer, M. Brodhun, J. Griffin, M. S. L. Lenon, K. Trpkov, L. Cheng, F. Chen, A. Levi, G. Cai, T. Q. Nguyen, A. Amin, A. Cimadamore, A. Shabaik, V. Manucha, N. Ahmad, N. Messias, F. Sanguedolce, D. Taheri, E. Baraban, L. Jia, R. B. Shah, F. Siadat, N. Swarbrick, K. Park, O. Hassan, S. Sakhaie, M. R. Downes, H. Miyamoto, S. R. Williamson, T. Holland-Letz, C. V. Schneider, J. N. Kather, Y. Tolkach, e T. J. Brinker. Pathologist-like explainable AI for interpretable Gleason grading in prostate cancer, 2024.
- D. Müller, P. Meyer, L. Rentschler, R. Manz, J. Bäcker, S. Cramer, C. Wengenmayr, B. Märkl, R. Huss, I. Soto-Rey, e J. Raffler. DeepGleason: a System for Automated Gleason Grading of Prostate Cancer using Deep Neural Networks, 2024.
- K. Rais, M. Amroune, A. Benmachiche, e M. Y. Haouam. Exploring Variational Autoencoders for Medical Image Generation: A Comprehensive Study, 2024.
- D. Ramos, J. Franco-Pedroso, A. Lozano-Diez, e J. Gonzalez-Rodriguez. Deconstructing cross-entropy for probabilistic binary classifiers. *Entropy*, 20(3):208, Mar. 2018. doi: 10.3390/e20030208.
- N. M. Rodrigues, J. G. D. Almeida, A. Rodrigues, L. Vanneschi, C. Matos, M. V. Litsitskaya, A. Uysal, S. Silva, e N. Papanikolaou. Deep Learning Features Can Improve Radiomics-Based Prostate Cancer Aggressiveness Prediction. *JCO Clinical Cancer Informatics*, (8), Sept. 2024. doi: 10.1200/cci.23.00180.

- J. Silva-Rodriguez, A. Colomer, J. Dolz, e V. Naranjo. Self-Learning for Weakly Supervised Gleason Grading of Local Patterns. *IEEE Journal of Biomedical and Health Informatics*, 25(8):3094–3104, Aug. 2021. doi: 10.1109/jbhi.2021.3061457.
- J. Silva-Rodríguez, A. Colomer, e V. Naranjo. WeGleNet: A weakly-supervised convolutional neural network for the semantic segmentation of Gleason grades in prostate histology images. *Computerized Medical Imaging and Graphics*, 88:101846, Mar. 2021. doi: 10.1016/j.compmedimag.2020.101846.
- Z. Tabatabaei, A. Colomer, K. Engan, J. Oliver, e V. Naranjo. Self-supervised learning of a tailored Convolutional Auto Encoder for histopathological prostate grading. In *2023 31st European Signal Processing Conference (EUSIPCO)*, pages 980–984, Helsinki, Finland, Sept. 2023. IEEE. doi: 10.23919/eusipco58844.2023.10289741.
- S. Vats, S. M. A. Al-Heejawi, T. Kondejkar, A. Breggia, B. Ahmad, R. Christman, R. S. T, e S. Amal. Segmenting Tumor Gleason Pattern using Generative AI and Digital Pathology: Use Case of Prostate Cancer on MICCAI Dataset, June 2024.
- W. Yang. Autoencoder in Machine Learning. *Science and Technology of Engineering, Chemistry and Environmental Protection*, 1(8), Aug. 2024. doi: 10.61173/p7hxhp58.