

Metodologia de Geração de Dados aplicada à Geração de Imagens de Exames de Colonoscopia

André Cerqueira Castro¹, Lucas Lima Neves², Heitor Sardinha Gonçalves Paiva²,
Ricardo Augusto Pereira Franco¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

²Escola de Engenharia Elétrica, Mecânica e de Computação
Universidade Federal de Goiás (UFG) – Goiânia – GO – Brasil

andre.castro@discente.ufg.br, lucas.neves@discente.ufg.br

heitorsardinha@discente.ufg.br, ricardofranco@ufg.br

Abstract. *Colonoscopy exams can detect adenomatous polyps, which represent the early stage in the vast majority of colorectal cancer cases. However, their varying sizes and shapes make them difficult to detect, even for state-of-the-art machine learning models. One of the reasons for this challenge is the lack of available data for this task. To address this issue, this paper proposes a methodology for generating artificial data for colonoscopy images, in which three image generative models, Guided Diffusion, PFGM++, and Improved Diffusion, are analyzed, and the generated images are processed and evaluated. The results show satisfactory performance in generating colonoscopy exam images across all the generative models analyzed, achieving FID values of 33.89 and SSIM of 0.2573 for the best generative model among all evaluated models and those existing in the literature.*

Resumo. *Os exames de colonoscopia podem detectar pólipos adenomatosos, que representam o estado inicial da grande maioria dos casos de câncer colorretal. Todavia, seus diferentes tamanhos e formatos tornam a tarefa de detectá-los difícil até mesmo para modelos estado da arte em aprendizado de máquina. Um dos motivos se deve à escassez de dados disponíveis para esta tarefa. Visando contornar esta problemática, é proposta neste artigo uma metodologia de geração de dados artificiais para imagens de colonoscopia, na qual se analisam três modelos generativos de imagens (Guided Diffusion, PFGM++ e o Improved Diffusion) e se realiza o tratamento e a avaliação das imagens geradas. Os resultados alcançados demonstram desempenho satisfatório na geração de imagens de exames de colonoscopia para todos os modelos generativos analisados, obtendo-se valores de FID igual a 33,89 e 0,2573 de SSIM para o melhor modelo generativo dentre todos os modelos avaliados e existentes na literatura.*

1. Introdução

Câncer colorretal é a segunda causa mais comum de mortes por câncer tanto para o sexo masculino quanto para o feminino ao redor do mundo [Baidoun et al. 2021]. Em 2024, é previsto que sejam diagnosticados mais de dois milhões de casos de câncer nos

Estados Unidos e, dentre estes casos, o câncer colorretal é aquele com maior taxa de mortalidade entre homens e o segundo maior entre mulheres [Siegel et al. 2024].

O câncer colorretal se desenvolve em estágios, sendo o primeiro deles o surgimento de pólipos adenomatosos benignos no intestino delgado e intestino grosso [Wu et al. 2020], como o caso exibido na Figura 1. Dentre os métodos possíveis para detecção destas anomalias, o exame de colonoscopia tem sido considerado a melhor alternativa há alguns anos devido à sua capacidade de identificação precoce e remoção de lesões pré-cancerígenas. Contudo, os principais problemas deste método são: a dificuldade para a realização do exame e para a visualização do pólipo, e a susceptibilidade a erros de diagnóstico, que podem ocorrer quando pólipos não são identificados através da análise visual do profissional durante a realização do exame [Wu et al. 2020].



Figura 1. Exemplo de pólipo adenomatoso colorretal [Jha et al. 2020].

Tendo isso em vista, novas tecnologias têm sido amplamente aplicadas para auxiliar diagnósticos em exames médicos, com destaque para o uso de Inteligência Artificial (IA), que tem avançado significativamente nos últimos anos e beneficiado diversas aplicações, incluindo na área da saúde [Viscaino et al. 2021]. Especificamente, algoritmos de aprendizado de máquina (*machine learning*) têm mostrado grande eficácia na detecção de pólipos, reduzindo erros e identificando anomalias com maior precisão [Marques et al. 2023]. Dentro dessa área, o aprendizado profundo (*deep learning*), que utiliza redes neurais profundas com múltiplas camadas, tem se destacado por alcançar resultados superiores em comparação às redes neurais rasas, sendo treinado para realizar tarefas de diferentes complexidades a partir de dados representativos, sem necessidade de programação explícita [Mitchell 1997].

Almejando contornar os problemas citados, este artigo propõe uma metodologia de geração de dados na qual se analisam e avaliam três modelos de geração de imagens artificiais, do estado da arte, visando gerar novas imagens de exames de colonoscopia com pólipos. O objetivo é determinar o modelo que gere imagens mantendo as características fisiológicas das imagens originais, enquanto obtém incremento da diversidade do conjunto de dados. Para isto, foram utilizados os seguintes modelos de geração de imagens: *Poisson Flow Generative Models* (PFGM++) [Xu et al. 2023], *Guided Diffusion* (GD) [Dhariwal and Nichol 2021] e *Improved Denoising Diffusion Probabilistic Models* (IDDPM) [Nichol and Dhariwal 2021].

Essa abordagem tem como objetivo superar a escassez de dados neste domínio e evitar preocupações legais acerca da privacidade dos pacientes. Gerando, assim, imagens sintéticas que poderão aprimorar o conjunto de dados reais para treinar modelos de detecção, segmentação e classificação de pólipos. Esses modelos possuem um impacto clínico significativo, ao proporcionar diagnósticos mais rápidos e precisos, reduzindo erros humanos e possibilitando a detecção precoce de pólipos, fundamental para

a prevenção do câncer colorretal. Dessa forma, pode-se definir abaixo as principais contribuições deste artigo:

- Propor uma metodologia de geração de imagens de exames de colonoscopia contendo pólipos;
- Analisar e implementar três modelos generativos (PFGM++, GD e IDDPM) para a geração de imagens sintéticas de exames de colonoscopia com pólipos;
- Desenvolver um *pipeline* de geração automática para imagens de colonoscopia;
- Avaliar as imagens geradas e comparar os resultados obtidos com os modelos generativos de imagens de colonoscopia disponíveis na literatura.

A organização do trabalho segue da seguinte forma: na Seção 2 serão apresentados os trabalhos relacionados; a metodologia utilizada para a geração dos dados, bem como breve descrição dos modelos e métricas utilizadas, são apresentadas na Seção 3; a Seção 4 contém os resultados obtidos e a análise comparativa com relação aos demais trabalhos da literatura; por fim, serão apresentadas as conclusões e trabalhos futuros na Seção 5.

2. Trabalhos Relacionados

A geração de imagens sintéticas na área médica ainda é um problema desafiador, visto a pouca quantidade de imagens para determinadas tarefas e a alta demanda de processamento computacional exigida para a realização do treinamento de modelos generativos. Todavia, alguns trabalhos vêm sendo desenvolvidos nesta área de pesquisa.

Imagens sintéticas do trato gastrointestinal podem ser criadas por diversos modelos generativos, alguns dos quais foram explorados para embasar o desenvolvimento da proposta deste artigo, a exemplo do PolypConnect [Fagereng et al. 2022], que realiza o preenchimento de pólipos em imagens saudáveis utilizando modelos de geração de imagens. Segundo os autores deste trabalho, os modelos de melhor resultado utilizados foram o Edge-Connect [Nazeri et al. 2019] e o AOT-GAN [Zeng et al. 2021], ambos utilizando *Generative Adversarial Network* (GANs).

Em [Waisberg et al. 2024], os autores analisaram diferentes vertentes de participação de modelos generativos para o campo da oftalmologia, incluindo *upscale* de imagens, geração de imagens de um determinado tipo de exame e transcrição de imagens de um exame para outro. Uma análise similar foi realizada pelos autores em [Lan et al. 2020], porém para o campo da biomedicina, em que se observou, novamente, os benefícios da utilização de modelos generativos, mais especificamente das GANs tanto para geração de novos dados, quanto para a melhoria dos dados existentes.

Os autores em [Pishva et al. 2023] desenvolveram um *framework* para geração de imagens de pólipos chamado RePolyp. Esse *framework* utiliza a lógica de *inpainting* para combinar um pólipo gerado artificialmente com uma imagem saudável também gerada.

A outra abordagem de modelos generativos proposta na literatura consiste no modelo SinGAN-SEG [Thambawita et al. 2022], o qual foi projetado para gerar imagens sintéticas com máscaras correspondentes a partir de uma única imagem de treinamento. Este método se destaca de GANs tradicionais pois requer apenas uma imagem e seu respectivo rótulo para o treinamento.

De forma semelhante, esse trabalho possui como objetivo propor uma metodologia de geração de dados avaliando três modelos generativos de imagens (estado da arte)

para gerar imagens de pólipos em exames de colonoscopia.

3. Metodologia de Geração de Dados

A geração de imagens sintéticas desempenha um papel crucial ao suprir a escassez de dados médicos. Além disso, essa técnica permite o aumento de conjuntos de dados, o que contribui para melhorar a generalização e o desempenho de modelos de aprendizado de máquina em tarefas como classificação e segmentação de imagens médicas.

3.1. Modelo de Geração de Dados

Ao longo dos últimos anos, múltiplos estudos estabeleceram a existência de uma correlação positiva entre a quantidade de dados utilizados e o desempenho de modelos de IA, especialmente para redes neurais profundas [Elhmadany et al. 2024]. Assim, a geração de dados sintéticos pode contribuir tanto para o aumento de dados quanto para a melhora da generalização do *dataset* e esta técnica se apresenta como alternativa pertinente para contornar esta problemática.

A Figura 2 mostra o fluxograma proposto neste artigo. Primeiramente, foram analisados e definidos três modelos do estado da arte a serem utilizados (PFGM++, GD e IDDPM) considerando seus desempenhos na tarefa de geração de imagens sintéticas em aplicações em várias áreas.

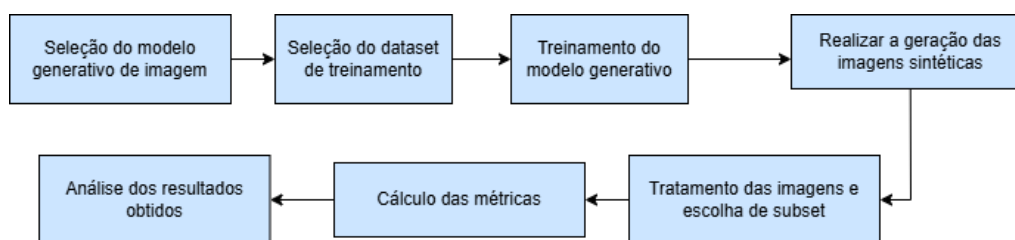


Figura 2. Fluxograma proposto de geração de imagens de exames de colonoscopia.

Foi definido, em seguida, o *dataset* a ser utilizado para treinamento dos modelos. Para este trabalho, o *dataset* escolhido foi o Kvasir-SEG [Jha et al. 2020], que consistem em um conjunto de 1.000 imagens de exames de colonoscopia com pólipos adenomatosos de diferentes tamanhos e formatos.

Posteriormente, foram escolhidos os hiperparâmetros de treinamento (resolução, número de épocas, *batch size*, dimensionalidade, entre outros) de treinamento, de acordo com as especificidades de cada modelo.

Após o treinamento do modelo de geração de imagens, realizou-se a geração das imagens de exames de colonoscopia. Observou-se que haviam imagens sintéticas repetidas. Assim, na etapa de tratamento das imagens, foi necessário desenvolver um processo de eliminação de imagens repetidas, comparando a similaridade entre uma imagem gerada e as imagens do *dataset* sintético e entre as imagens do *dataset* de imagens reais.

Por fim, as métricas foram calculadas e os resultados para cada modelo foram obtidos e comparados com os seguintes trabalhos: [Fagereng et al. 2022], [Pishva et al. 2023] e [Thambawita et al. 2022].

3.2. Modelo Guided Diffusion

O modelo *Modelo Guided Diffusion* [Dhariwal and Nichol 2021] integra o processo de difusão com orientações adicionais para guiar a geração de imagens. No processo de difusão, é aplicado gradualmente um ruído gaussiano à imagem, transformando-a, assim, progressivamente de uma imagem real em puro ruído. O processo inverso, nomeado como *denoising*, permite que o modelo aprenda a remover esse ruído, gerando imagens a partir de uma condição inicial aleatória.

O diferencial do *Modelo Guided Diffusion* está na utilização de guias que influenciam o processo de reversão do ruído. Essas orientações podem incluir informações externas, como rótulos de classes ou dados condicionais, que ajustam as previsões feitas pelo modelo, resultando em amostras geradas de maneira direcionada. Esse recurso melhora substancialmente a qualidade das amostras, permitindo, dessa forma, a aplicação do modelo em tarefas como geração de imagens condicionadas ou restauração de imagens [Dhariwal and Nichol 2021].

3.3. Modelo PFGM++

O modelo *Poisson Flow Generative Models* [Xu et al. 2023] utiliza conceitos físicos de campos elétricos para encontrar uma função capaz de mapear a distribuição real dos dados para uma distribuição em um espaço latente, o qual poderá ser amostrado posteriormente, percorrendo o caminho inverso para gerar uma imagem que poderia fazer parte da distribuição real. Esta versão do modelo utiliza um treinamento baseado no mesmo princípio que os modelos de difusão, com a adição e remoção de ruído gaussiano em etapas.

O espaço latente mapeado pelo modelo pode ter sua dimensionalidade definida pelo parâmetro D e, quando D tende ao infinito, o modelo se comporta de forma semelhante a um modelo de difusão convencional. Quanto maior o valor de D , mais o modelo produzirá imagens fiéis ao *dataset* original, porém como consequência, menor será a variabilidade das imagens produzidas. Em contrapartida, conforme o valor de D diminui, maior a variabilidade e menor a especificidade das imagens [Xu et al. 2023]. De acordo com os testes realizados e com os resultados obtidos pelos próprios autores do modelo para a resolução das imagens utilizadas, foi definido o valor de 4.096 para este parâmetro.

3.4. Modelo Improved Diffusion

O modelo *Modelo Improved Diffusion* [Nichol and Dhariwal 2021] obteve resultados relevantes quando comparados com os modelos do estado da arte por meio de modificações nos *Denoising Diffusion Probabilistic Models* (DDPM). Essas modificações incluíam ajustes na arquitetura do modelo e no processo de treinamento, permitindo uma melhor representação das distribuições de dados.

Uma das principais inovações apresentadas pelos autores foi a introdução de técnicas que possibilitaram ao modelo aprender as variâncias do processo de difusão reversa. Isso não apenas aprimorou a eficiência na amostragem, reduzindo o número de etapas necessários para gerar as amostras, mas também garantiu boa qualidade das amostras geradas [Nichol and Dhariwal 2021].

3.5. Métricas

No contexto da geração de imagens, é fundamental avaliar objetivamente a qualidade das gerações, considerando a fidelidade visual e a diversidade. Para isso, indicadores amplamente utilizados são o *Fréchet Inception Distance* (FID) [Heusel et al. 2017] e o *Structural Similarity Index Measure* (SSIM) [Wang et al. 2004].

O **FID** mede a distância entre as distribuições de características das imagens reais (r) e sintéticas (s), avaliando o quão próximas as geradas estão das reais. Sua fórmula é definida conforme apresentada abaixo:

$$FID = \|\mu_r - \mu_s\|^2 + \text{Tr}(\Sigma_r + \Sigma_s - 2(\Sigma_r \Sigma_s)^{\frac{1}{2}}) \quad (1)$$

onde μ_r e μ_s são as médias das características extraídas das imagens reais e sintéticas, respectivamente. Já Σ_r e Σ_s representam as matrizes de covariância das características das imagens reais e sintéticas, capturando a dispersão das distribuições e refletindo a diversidade. Valores de FID próximos de 0 indicam maior similaridade entre imagens reais e geradas [Heusel et al. 2017].

O **SSIM** Avalia a similaridade estrutural entre duas imagens, sendo, no contexto da geração de imagens, uma real e outra sintética. A métrica avalia os aspectos de luminância, contraste e estrutura.. Sua fórmula é dada por:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

onde x e y são as imagens a ser comparadas, μ_x e μ_y são as médias locais, σ_x^2 e σ_y^2 representam as variâncias locais, e σ_{xy} é a covariância local que reflete a similaridade estrutural entre as imagens. Valores de SSIM próximos de 1 indicam alta similaridade visual, enquanto valores próximos de 0 apontam diferenças significativas [Wang et al. 2004].

4. Resultados e Discussão

Esta seção apresenta os experimentos realizados para aplicação da metodologia de geração de imagens de exames de colonoscopia. Os treinamentos foram todos realizados utilizando uma *Graphics Processing Unit* (GPU) Nvidia V-100.

Realizou-se uma divisão aleatória do *dataset* Kvasir-SEG alocando 80% (800 imagens) para treinamento e 20% (200 imagens) para validação. Todos os três modelos de geração de imagens foram treinados utilizando o mesmo *set* de imagens de treinamento e de validação. As imagens do *dataset* foram redimensionadas para a resolução 128x128 *pixels* por questões relacionadas à fidelidade ao comparar com os resultados de outros trabalhos, recursos computacionais utilizados e compatibilidade com os modelos escolhidos.

Foi gerado, por cada modelo, um total de 1.000 imagens. Observou-se que algumas imagens geradas eram iguais a outras imagens geradas ou iguais às imagens do *dataset* original. Em seguida, foi realizada a remoção de imagens repetidas. As imagens restantes foram submetidas à escolha aleatória de um *subset* de 800 imagens, de forma a obter quantidade de imagens sintéticas igual à quantidade de imagens de treinamento do *dataset* de imagens reais (Kvasir-SEG).

Os resultados obtidos podem ser observados na Tabela 4. Os melhores resultados estão em negrito e os segundo melhores resultados estão em itálico.

Dentre os modelos avaliados, o modelo proposto para gerar imagens sintéticas de exames de colonoscopia *Improved Diffusion* apresenta o melhor resultado na métrica de FID, com um valor de 33,89. Isso indica que as imagens geradas por esse modelo possuem características estatísticas mais próximas das imagens reais, o que reflete, desse modo, uma maior similaridade na distribuição de texturas, formas e cores entre as imagens sintéticas e reais.

Modelo	SSIM	FID
Guided Diffusion	0,255	44,52
PFGM++	0,239	<u>40,39</u>
Improved Diffusion	<u>0,257</u>	33,89
Repolyp [Pishva et al. 2023]	-	93,43
PolypConnect [Fagereng et al. 2022]	0,527	76,31
SinGAN-Seg [Thambawita et al. 2022]	-	43,46

Tabela 1. Comparação das métricas FID e SSIM de avaliação dos modelos de geração proposto e os modelos disponíveis na literatura.

O modelo PFGM++ também se destaca, alcançando um FID de 40,39, posicionando-se, assim, como a segunda melhor alternativa para a geração de imagens de exames de colonoscopia. O modelo *Guided Diffusion* obteve FID igual a 44,52, um pouco superior quando comparado com os resultados do modelo SinGAN-Seg [Thambawita et al. 2022], embora ainda apresente resultados satisfatórios quando comparados aos demais.

Pode-se observar que os três modelos propostos neste artigo para a geração de imagens de exames de colonoscopia com pólipos obtiveram os melhores valores de FID quando comparados com os modelos generativos presentes na literatura, exceto pelo modelo SinGAN-Seg, que obteve o terceiro melhor FID com valor de 43,46. Contudo, dentre os quatro melhores modelos, o modelo SinGAN-Seg realiza a geração de várias imagens baseada em apenas uma imagem de entrada, de forma com que as imagens geradas por esse modelo possuam padrão bem repetitivo [Thambawita et al. 2022]. Assim, pode-se considerar que o modelo *Guided Diffusion*, apesar do valor de FID maior, realiza a geração de imagens distintas de exames de colonoscopia, incrementando a generalização de imagens no *dataset*.

Na métrica de SSIM, o modelo PolypConnect obteve o maior valor, com um SSIM de 0,527, apesar de possuir o segundo maior FID dentre os modelos. Pode-se deduzir que apesar de ter o maior valor de SSIM, as imagens geradas são similares ao conjunto de dados real, contudo, a distribuição de probabilidade das imagens geradas não está próxima da distribuição de probabilidade dos dados reais. Ou seja, as imagens, apesar de conter características similares, o modelo PolypConnect não possui boa representação do conjunto de dados reais quando comparados com os modelos analisados neste trabalho.

Para fins comparativos, a Tabela 4 mostra a métrica SSIM cujo valor foi obtido através da análise entre o *dataset* Kvasir-SEG e outros dois *datasets* reais de imagens de exames de colonoscopia contendo pólipos, sendo estes o CVC-ClinicDB

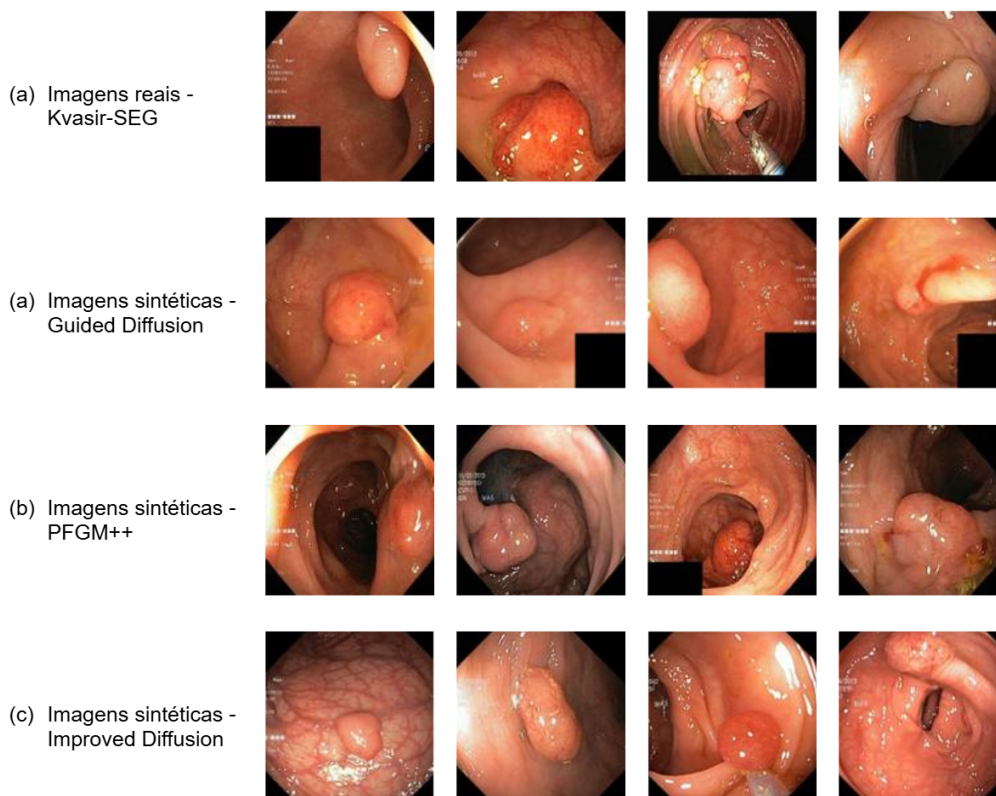


Figura 3. Comparação entre as imagens sintéticas e reais.

[Bernal et al. 2015] e o Etis-LaribPolypDB [Silva et al. 2014].

Dataset	SSIM
CVC-ClinicDB	0,205
Etis-LaribPolypDB	0,234

Tabela 2. Comparação dos valores da métrica SSIM entre os *datasets* Kvasir-SEG, CVC-ClinicDB [Bernal et al. 2015] e Etis-LaribPolypDB [Silva et al. 2014]

Os valores exibidos na Tabela 4, entre *datasets* de imagens reais, demonstram que não necessariamente um valor de SSIM muito elevado é o melhor para o domínio desse problema, uma vez que, com um valor muito elevado, a variabilidade dos dados gerados pode ser comprometida, indicando uma aproximação em excesso com as imagens originais. Em contrapartida, os modelos propostos apresentam valores de SSIM que se assemelham aos dos dados reais (CVC-ClinicDB e Etis-LaribPolypDB) quando comparados com o Kvasir-SEG. Dessa forma, observa-se maior aproximação dos dados artificiais produzidos pelo PFGM++, GD e IDDPM aos dados reais do *dataset* Kvasir-SEG em termos estruturais.

Logo, nota-se que os modelos PFGM++, *Guided Diffusion* e *Improved Diffusion* apresentaram valores de SSIM menores em comparação ao PolypConnect, que pode ser explicado por uma geração de imagens mais diversa por parte desses modelos quando comparados com o valor obtido de FID para os três modelos. Estes resultados indicam que, embora a similaridade estrutural das imagens geradas seja inferior, quando com-

parados ao *dataset* Kvasir-SEG, a variabilidade e a qualidade geral das imagens foram superiores aos demais métodos disponíveis na literatura.

Por fim, os resultados visuais observados na Figura 3 demonstram que as imagens geradas pelos modelos PFGM++, *Guided Diffusion* e *Improved Diffusion* apresentam uma aproximação significativa ao padrão dos dados originais do Kvasir-SEG. Nota-se que essas imagens geradas de pólipos mantêm uma coerência anatômica marcante, além de exibirem coloração, brilho e contraste bastante similares aos dados originais. Assim, isso reforça a capacidade desses modelos em gerar amostras realistas e com alta fidelidade visual em relação ao conjunto de referência.

5. Conclusões e Trabalhos Futuros

Neste artigo, foi proposto uma metodologia de geração de dados sintéticos para exames de colonoscopia com pólipos, utilizando modelos generativos de propósito geral. A partir dos resultados obtidos pelos modelos analisados, observou-se que o modelo *Improved Diffusion* gerou imagens sintéticas de alta qualidade, proporcionando aumento na diversidade dos conjuntos de dados. Os resultados obtidos demonstraram o potencial das técnicas utilizadas para contornar limitações associadas à escassez de *datasets* para aplicações médicas relacionadas à exames de colonoscopia.

Por fim, como trabalhos futuros, pode-se citar a avaliação de novos modelos de geração de imagens e a inclusão de imagens sintéticas no treinamento para aprimorar a detecção, segmentação ou classificação de pólipos. Essas ações podem ampliar o diagnóstico clínico, oferecendo tomadas de decisão mais precisas, que são fundamentais para a detecção precoce de pólipos e a prevenção de câncer colorretal.

Referências

- Baidoun, F., Elshiwiy, K., Elkerai, Y., Merjaneh, Z., Khoudari, G., Sarmini, M. T., Gad, M., Al-Husseini, M., and Saad, A. (2021). Colorectal cancer epidemiology: Recent trends and impact on outcomes. *Current Drug Targets*, 22(9):998–1009.
- Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., and Vilariño, F. (2015). Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111. Epub 2015 Mar 20.
- Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis.
- Elhmadany, M., Elmadah, I., and Abdelmunim, H. (2024). Instance segmentation on distributed deep learning big data cluster. *Journal of Big Data*, 11.
- Fagereng, J. A., Thambawita, V., Storås, A. M., Parasa, S., de Lange, T., Halvorsen, P., and Riegler, M. A. (2022). Polypconnect: Image inpainting for generating realistic gastrointestinal tract images with polyps.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Klambauer, G., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a nash equilibrium. *CoRR*, abs/1706.08500.
- Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., and Johansen, H. D. (2020). Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling*, pages 451–462. Springer.

- Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., and Zhou, X. (2020). Generative adversarial networks and its applications in biomedical informatics. *Frontiers in Public Health*, 8:164.
- Marques, A. F., Marques, K. F., Beraldo, M. N. M. d. S., Lima, T. B., Sasaki, L. Y., and Beraldo, R. F. (2023). Inteligência artificial na colonoscopia no rastreamento do câncer colorretal: revisão de literatura. *Brazilian Journal of Health Review*, 6(4):18764–18774.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA.
- Nazeri, K., Ng, E., Joseph, T., Qureshi, F. Z., and Ebrahimi, M. (2019). Edgeconnect: Generative image inpainting with adversarial edge learning.
- Nichol, A. and Dhariwal, P. (2021). Improved denoising diffusion probabilistic models.
- Pishva, A. K., Thambawita, V., Torresen, J., and Hicks, S. A. (2023). Repolyp: A framework for generating realistic colon polyps with corresponding segmentation masks using diffusion models. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 47–52.
- Siegel, R. L., Giaquinto, A. N., and Jemal, A. (2024). Cancer statistics, 2024. *CA: A Cancer Journal for Clinicians*, 74(1):12–49. Erratum in: *CA Cancer J Clin*. 2024 Mar-Apr;74(2):203. doi: 10.3322/caac.21830.
- Silva, J., Histace, A., Romain, O., Dray, X., and Granado, B. (2014). Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):283–293. Epub 2013 Sep 15.
- Thambawita, V., Salehi, P., Sheshkal, S. A., Hicks, S. A., Hammer, H. L., Parasa, S., Lange, T. d., Halvorsen, P., and Riegler, M. A. (2022). Singan-seg: Synthetic training data generation for medical image segmentation. *PLOS ONE*, 17(5):e0267976.
- Viscaino, M., Torres Bustos, J., Muñoz, P., Auat Cheein, C., and Cheein, F. A. (2021). Artificial intelligence for the early detection of colorectal cancer: A comprehensive review of its advantages and misconceptions. *World Journal of Gastroenterology*, 27(38):6399–6414.
- Waisberg, E., Ong, J., Kamran, S. A., Masalkhi, M., Paladugu, P., Zaman, N., Lee, A. G., and Tavakkoli, A. (2024). Generative artificial intelligence in ophthalmology. *Survey of Ophthalmology*, Epub ahead of print. S0039-6257(24)00044-4.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- Wu, Z., Li, Y., Zhang, Y., Hu, H., Wu, T., Liu, S., Chen, W., Xie, S., and Lu, Z. (2020). Colorectal cancer screening methods and molecular markers for early detection. *Technology in Cancer Research Treatment*, 19:1533033820980426. Jan-Dec.
- Xu, Y., Liu, Z., Tian, Y., Tong, S., Tegmark, M., and Jaakkola, T. (2023). Pfgm++: Unlocking the potential of physics-inspired generative models.
- Zeng, Y., Fu, J., Chao, H., and Guo, B. (2021). Aggregated contextual transformations for high-resolution image inpainting.