Classificador Vision Transformer para detecção de pólipos em imagens de colonoscopia

Vision Transformer classifier for polyp detection in colonoscopy images

Monique Borges Seixas, João Pedro Busnardo de Souza, Murilo Moreira Mello, Jader Tavares Lima, Maria Fernanda Oliveira de Figueiredo, Sthefanie Monica Premebida, Heron dos Santos Lima, Marcella Scoczynski [1] ,

Vinicius Sylvestre Simm, Mateus Dalla Costa, Wesley Pacheco, Paulo Victor dos Santos [2]

RESUMO

O câncer colorretal (CCR) é uma preocupação significativa que pode ser identificada de forma eficaz por meio da análise de pólipos, crescimentos anormais de tecido de várias formas e tamanhos encontrados em imagens de colonoscopia. A detecção desses pólipos desempenha um papel crucial no diagnóstico precoce do CCR, o que é fundamental para o tratamento bem-sucedido dos pacientes. Nesta pesquisa, apresentamos uma abordagem que utiliza um classificador baseado em *Transformer*, especificamente projetado para detectar e caracterizar pólipos em imagens de colonoscopia. O conjunto de dados utilizado para este estudo é o banco de dados CP-CHILD. Nossa trabalho demonstrou um desempenho notável, alcançando uma precisão de 97,86%. Além disso, obtivemos valores de precisão média de 97,51% no conjunto de testes, uma taxa de recall notável de 91,87% e uma pontuação F1 de 94,56%, destacando sua eficácia tanto em sensibilidade quanto em precisão na detecção de pólipos. Esses resultados promissores potencializam o classificador para a detecção precoce e o diagnóstico do CCR, contribuindo para uma melhoria nos resultados dos pacientes e estratégias de tratamento mais eficazes.
PALAVRAS-CHAVE: câncer colorretal, colonoscopia, classificação de imagens, mecanismo de atenção, pólipo, pré-processamento, pós-precessamento

ABSTRACT

Colorectal cancer (CRC) is a significant health concern that can be effectively identified through the analysis of polyps, abnormal tissue growths of various shapes and sizes found in colonoscopy images. Detecting these polyps plays a pivotal role in early diagnosis of CRC, which is critical for successful patient treatment, as all CRC cases originate from these precancerous growths. In this research, we introduce an approach utilizing a Transformer-based classifier specifically designed to detect and characterize polyps in colonoscopy images. The dataset employed for this study is the CP-CHILD database, a valuable resource for training and validation. Our proposed approach has showcased remarkable performance, achieving an impressive accuracy of 97.86%. Additionally, we obtained high average precision values of 97.51% on the test set, illustrating the robustness of our method in correctly identifying polyps. Furthermore, our model demonstrated a notable recall rate of 91.87% and an F1 score of 94.56%, underscoring its effectiveness in both sensitivity and precision in polyp detection. These promising results highlight the potential of our Transformer-based classifier to significantly enhance early detection and diagnosis of CRC to improved patient outcomes and more effective treatment strategies.
KEYWORDS: colonoscopy images, colorectal cancer, image classification, polyp, post- processing, pre-processing, vision transformer

[1] UTFPR, Ponta Grossa, Paraná, Brasil E-mail: [moniqueseixas, joaopedrosouza, murilomello, jaderlima, mariafigueiredo, heronlima, marcella]@utfpr.edu.br, smpremebida@gmail.com.

[2] [UEM, Maringá, UNIDEP,Pato Branco;UFG, Goiania], Brasil E-mail: viniciussimm, mateus dc1998@hotmail.com, wesley.pacheco@ufg.br,paulo.analise@live.com.

INTRODUÇÃO

Atualmente, o câncer colorretal (CCR) é uma das principais causas de mortes relacionadas ao câncer em todo o mundo [1], [2]. Ele engloba cânceres do cólon, junção reto-sigmóide e reto [3], e a estimativa mais recente do Instituto Nacional de Câncer (INCa) do Brasil prevê 704.000 novos casos de câncer no Brasil entre 2023-2025, com 6,5% dos casos sendo de câncer de cólon e reto [4].

A colonoscopia é o método padrão-ouro para diagnóstico e remoção de pólipos [7], mas a qualidade da análise intestinal depende tanto do operador quanto do paciente. A taxa de detecção de adenomas (ADR) é um indicador-chave de qualidade para a colonoscopia, medindo a proporção de colonoscopias em que pelo menos um adenoma é encontrado. Taxas de ADR mais altas geralmente indicam um melhor desempenho dos endoscopistas, com uma faixa estimada de 20-30%. Quando a ADR cai abaixo de 20%, há um aumento no risco de câncer intervalar, que se refere ao CCR que se desenvolve após uma colonoscopia negativa, mas antes do próximo exame recomendado [8].

O diagnóstico precoce impacta significativamente a taxa de sobrevivência, com estágios avançados do CCR resultando em uma menor expectativa de vida [6], [9]. Além disso, o CCR representa uma significativa carga para a saúde pública, com custos esperados de alcançar um bilhão de reais brasileiros até 2030 [10]. Embora os endoscopistas tenham um alto desempenho, erros humanos, preparo inadequado do intestino e desafios na visualização de pólipos ainda são prevalentes. Métodos de Aprendizado Profundo (AP) podem melhorar a qualidade da colonoscopia auxiliando no diagnóstico e visualização de pólipos, aumentando as taxas de diagnóstico precoce e reduzindo os gastos com saúde pública. Um desses métodos é o *Vision Transformer* [10], que apresentou resultados impressionantes em tarefas de visão computacional.

Este artigo tem como objetivo utilizar um classificador baseado em *Transformer* para realizar a detecção de pólipos em dados de testes de colonoscopia.

METODOLOGIA

O modelo proposto de *Vision Transformer* consiste em uma pilha de 6 camadas de codificador, cada uma com 8 cabeças de atenção. A imagem de entrada é primeiro incorporada em uma sequência de patches de 16x16 pixels, que são então achatados em um vetor de embeddings. Os embeddings são posteriormente alimentados nas camadas do codificador, que realizam operações de autoatenção e "feedforward"para produzir uma representação final para cada patch. A representação final é passada por uma camada linear e uma função sigmoide, produzindo uma probabilidade entre 0 e 1.

Antes do treinamento, os dados foram pré-processados para eliminar artefatos que poderiam degradar a qualidade das imagens. Primeiramente, os conjuntos de dados CP-CHILD A e CP-CHILD B foram mesclados em um único conjunto de dados para aumentar os dados disponíveis. O próximo passo foi redimensionar cada imagem para o tamanho de 224x224 pixels e normalizá-las na faixa de [-1, 1] antes de aplicar etapas adaptadas [11] para eliminar realces especulares, que são pontos de alta intensidade na imagem que aparecem devido à iluminação de objetos brilhantes. As imagens foram primeiro convertidas para escala de cinza, em seguida, o valor mediano do tensor da imagem foi tomado e multiplicado por um peso para produzir um limiar. Depois, o valor do limiar foi subtraído da imagem, que foi posteriormente usada para criar uma máscara, selecionando os valores que

estavam acima de um valor arbitrário (0,6 neste caso). Finalmente, a máscara foi dilatada para garantir melhores resultados e usada para preencher a imagem original com a funcionalidade de inpaint do OpenCV. A Figura 1 ilustra a etapa de pré-processamento.

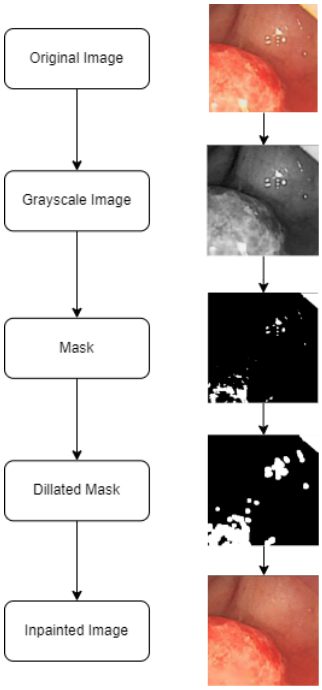

Figura 1 – Pipeline de remoção de realces especulares aplicada às imagens.

Para cada lote, técnicas de aumento de dados foram aplicadas, incluindo corte aleatório, espelhamento horizontal aleatório e rotação aleatória, para aumentar a variabilidade do conjunto de treinamento e evitar o overfitting. Os dados também são carregados usando uma estratégia de oversampling, na qual as imagens em cada lote foram amostradas aleatoriamente e carregadas com frequências correspondentes a um peso especificado, neste caso, um menos a porcentagem de ocorrências para cada classe no conjunto de treinamento. Essa estratégia tem como objetivo aumentar as ocorrências da classe minoritária com o objetivo de compensar os desequilíbrios no conjunto de dados original.

O modelo é treinado por um total de 500 épocas usando descida de gradiente estocástica, especificamente o algoritmo Adam [12], com uma taxa de aprendizagem inicial de 0,0001 e um tamanho de lote de 100, bem como uma penalidade L2 de 10^{-9} para os pesos. Para garantir a convergência, a taxa de aprendizagem foi reduzida por um fator de 10 se não houvesse melhora na perda de validação após 20 épocas. Para evitar o overfitting, o dropout é aplicado às camadas de autoatenção e densas com uma probabilidade de 0,2. O treinamento levou aproximadamente 10,5 horas para ser concluído com aceleração de GPU, sendo que o modelo que resultou na menor perda de validação foi salvo para avaliação.

RESULTADOS

O classificador foi treinado em um conjunto de dados com 8100 imagens e avaliado a cada época em um conjunto de validação separado contendo 700 imagens. No final do treinamento, o

modelo foi validado em um conjunto de teste contendo 700 imagens. As Figuras 2 e 3 exibem, respectivamente, as curvas de treinamento para a perda e a precisão nos conjuntos de treinamento e validação.

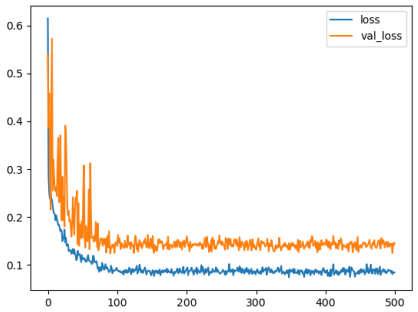

Figura 2 – Modelo de perda durante treinamento

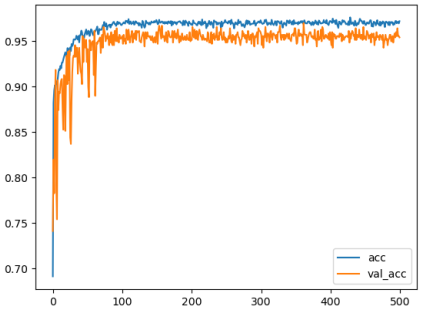

Figura 3 – Modelo de precisão durante treinamento

Embora as curvas indiquem um leve grau de overfitting, o classificador ainda demonstra uma capacidade satisfatória de generalização, como evidenciado pelas métricas (Tabela 1). Conforme mostrado na Tabela 1, os resultados indicam que o classificador teve um desempenho muito bom, com uma precisão de 97.86% no conjunto de dados de teste.

Tabela 1 – Métricas

Metrics	Test	Train	Val
Loss	0.0408	0.1255	0.0882
Accuracy	0.9786	0.9641	0.9700
Precision	0.9751	0.9920	0.9398
Recall	0.9187	0.9360	0.9107
F1 Score	0.9456	0.9628	0.9201
Area under ROC curve	0.9966	0.9953	0.9889

Além disso, os resultados também mostraram que o classificador teve um desempenho particularmente bom em termos de precisão, com valores médios de precisão de 97.51% no conjunto de teste. A precisão é uma métrica importante, pois mede o número de exemplos verdadeiros positivos classificados corretamente em relação ao total de exemplos positivos. O classificador foi capaz de obter pontuações altas em todas as outras métricas também, incluindo recall, pontuação

F1 e área sob a curva ROC (AUROC). Isso sugere que o modelo é altamente eficaz na identificação tanto de exemplos positivos quanto negativos, mantendo um equilíbrio geral entre precisão e recall.

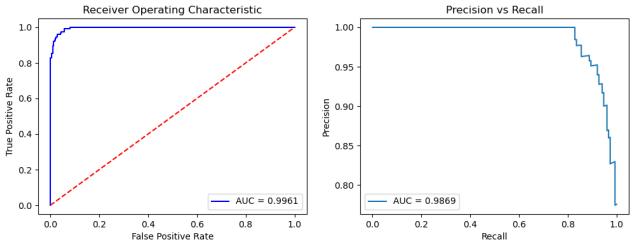

Figura 4 – Curvas ROC e de Precisão-Recall para o modelo

CONCLUSÃO

O câncer colorretal (CCR) é uma das principais causas de morte por câncer nos dias de hoje. Ele abrange cânceres do cólon, junção reto-sigmóide e reto. A detecção de pólipos, que são crescimentos anormais de tecido apresentados em várias formas e tamanhos, ajuda no diagnóstico precoce e no tratamento, já que todos os casos de CCR têm origem em pólipos.

Neste artigo, propusemos um classificador baseado em Transformer para detectar pólipos em imagens de colonoscopia. Nossos resultados indicaram que o classificador obteve uma precisão de 97,86% e valores de precisão média de 97,51% no conjunto de teste, superando redes em trabalhos similares. Nossa abordagem também demonstrou uma capacidade satisfatória de generalização. Como trabalhos futuros, pretendemos aplicar sintonia e otimização de hiperparâmetros, além de considerar arquiteturas pré-treinadas que os Vision Transformers funcionam melhor em tarefas downstream pequenas [10]. Outra possibilidade é usar uma sequência de entrada formada por mapas de características extraídos de uma CNN como alternativa aos patches de imagem brutos.

Agradecimentos

Nosso agradecimentos à Universidade Tecnológica Federal do Paraná (UTFPR) e à Fundação Araucária pelo apoio para a realização desta pesquisa.

Conflito de interesse

Os autores declaram que não há conflito de interesse.

BIBLIOGRAFIA

[1] World Health Organization, "Cancer statistics", WHO, 2020.

[2] H. J. Sung, S. H. Lee, and J. H. Park, "Colorectal cancer statistics and trends", J. Cancer, vol. 12, pp. 1405–1412, 2021.

[3] M. P. O'Leary, M. R. Kaufman, J. A. Smith, and S. M. Schlossberg, National Center for Biotechnology Information (US), "Peyronie's Disease: AUA Guideline,"in The AUA Guideline on Peyronie's Disease [Internet], American Urological Association, May 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK553197/ [Accessed: Apr. 28, 2023].

[4] Instituto Nacional de Câncer, "Estimativa 2023 - Incidência de câncer no Brasil", INCA, 2023.

[5] Universidade Federal do Rio Grande do Sul, "TelessaúdeRS - Pólipos Colorretais", UFRGS, 2022.

[6] N. C. Thanh, T. Q. Long, "CRF-EfficientUNet: An improved UNet framework for polyp segmentation in colonoscopy images", IEEE Access, vol. 9, pp. 156987–157001, 2021.

[7] R. C. Pinto, M. K. Seabra, A. A. Cunha, C. G. M. Pagano, H. G. Mussnich, "Assessment of Quality Indexes in Colonoscopy", J. Coloproctology, vol. 41, no. 1, pp. 23–29, 2021.

[8] M. V. Furlanetto, J. A. Zwierzikowski, C. F. Bertoldo, G. A. S. M. Wistuba, E. I. B. Tashima, A. H. B. G. Vieira, H. L. Invitti, A. S. Brenner, "Analysis of Patients Undergoing Colonoscopies", J. Coloproctology, vol. 42, no. 1, pp. 14–19, 2022.

[9] R. A. Rostirolla, J. C. Pereira-Lima, C. R. Teixeira, A. W. Schuch, C. Perazzoli, C. Saul, "Desenvolvimento de neoplasias/adenomas avançados colorretais", Arq. Gastroenterol., vol. 46, no. 3, pp. 167–172, 2009.

[10] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale."arXiv preprint arXiv:2010.11929 (2020).

[11] Sánchez-González, Alain, and Begoña García-Zapirain Soto. "Colonoscopy image pre-processing for the development of computer-aided diagnostic tools."Surgical Robotics. IntechOpen, 2017.

[12] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization."arXiv preprint arXiv:1412.6980 (2014).