KvasirCAM: Diagnóstico Explicável de Doenças Gastrointestinais com Atenção Visual e Otimização Bayesiana

Marcos R. A. Amorim[1], **Neilson P. Ribeiro**[12], **Luana B. da Cruz**[3], **João O. B. Diniz**[12], **Geraldo B. Júnior**[1], **João Dallyson S. Almeida**[1]

[1]Núcleo de Computação Aplicada — Universidade Federal do Maranhão (UFMA)
Caixa Postal 65.085–580 — São Luís — MA — Brasil

[2]Fábrica de Inovação - Instituto Federal do Maranhão (IFMA)
Caixa Postal 65.940–000 — Grajaú — MA — Brasil

[3]Laboratório de Inteligência Computacional Aplicada (LICA) -
Universidade Federal do Cariri (UFCA)
Caixa Postal 63.048-080 — Juazeiro do Norte — CE — Brasil

marcos.raffael@discente.ufma.br

Abstract. *Gastrointestinal cancer represents a significant portion of global oncological diseases, with colorectal cancer being the third most diagnosed neoplasm and stomach cancer the fifth. Early detection increases survival rates, and endoscopy is the primary examination for identifying these pathologies. Detection studies using Deep Learning have assisted in early identification; however, hyperparameter search is not a trivial task. This work proposes an automatic method for detecting gastrointestinal pathologies by integrating region of interest extraction, visual attention, hyperparameter optimization, and explainability into CNN architectures. The results demonstrate improvements across architectures, with ResNet50 achieving an F1-Score of 94.33% and an AUC of 98.22%, confirming that the approach is effective for gastrointestinal diagnosis.*

Resumo. *O câncer gastrointestinal representa uma parcela das doenças oncológicas mundiais, sendo o câncer colorretal a terceira neoplasia mais diagnosticada e o câncer de estômago a quinta. A detecção precoce aumenta a sobrevida, sendo a endoscopia o principal exame para identificação dessas patologias. Trabalhos de detecção utilizando Deep Learning têm auxiliado na identificação precoce, contudo a busca de hiperparâmetros não é uma tarefa trivial. Este trabalho propõe um método automático de detecção de patologias gastrointestinais integrando extração de região de interesse, atenção visual, otimização de hiperparâmetros e explicabilidade em arquiteturas CNN. Os resultados demonstram melhora nas arquiteturas, com a ResNet50 alcançando F1-Score de 94,33% e AUC de 98,22%, evidenciando que a abordagem é eficaz no diagnóstico gastrointestinal.*

1. Introdução

O câncer gastrointestinal (GI) representa uma parcela substancial da carga global de doenças oncológicas. Segundo o GLOBOCAN 2022, o câncer colorretal figura como

a terceira neoplasia mais diagnosticada mundialmente, com mais de 1,9 milhão de novos casos (9,6% do total), e o câncer de estômago como o quinto, com mais de 968 mil diagnósticos [Bray et al. 2024]. Corroborando a gravidade deste cenário, um estudo populacional publicado na *The Lancet* estima que o risco global de um indivíduo desenvolver GI ao longo da vida é de aproximadamente 8,20%, o que significa que cerca de uma em cada 12 pessoas será afetada pela doença [Wang et al. 2024].

A endoscopia permanece como padrão-ouro para o rastreamento e prevenção. No entanto, sua eficácia diagnóstica apresenta alta variabilidade interobservador, com aproximadamente um quarto das neoplasias não detectadas durante exames convencionais, constituindo a principal causa do câncer de intervalo. Uma meta-análise recente de ensaios clínicos randomizados demonstrou que o uso de endoscopia assistida por computador reduz a taxa de perda de adenoma em 54% e a taxa de perda de pólipos em 56% quando comparada aos procedimentos manuais [Maida et al. 2025].

O desafio de escalabilidade é ainda mais crítico na Cápsula Endoscópica. Embora seja uma alternativa não invasiva, essa técnica gera um volume massivo de dados, capturando entre 2 a 35 quadros por segundo. Um único exame pode produzir cerca de 50.000 imagens, exigindo até duas horas de dedicação exclusiva de um especialista para revisão, o que cria um gargalo operacional significativo nos fluxos clínicos [Su et al. 2025].

Diante disso, a área evolui rapidamente para a integração de sistemas de *Artificial Intelligence* (AI) para auxiliar em doenças gastrointestinais [Aguiar et al. 2024]. Iniciativas recentes, como o desafio Medico 2025, destacam a urgência de desenvolver modelos de *Explainable AI* (XAI) e suporte à decisão baseados na base de dados Kvasir [Gautam et al. 2025]. Contudo, a simples aplicação de *Convolutional Neural Networks* (CNNs) pré-treinadas não garante desempenho ótimo. A literatura indica que o ajuste manual de hiperparâmetros é ineficiente e propenso a vieses, exigindo métodos de otimização para extrair o máximo potencial dessas arquiteturas [El-Bouzaidi et al. 2025].

Neste contexto, o objetivo principal deste trabalho é desenvolver um método automático e otimizado para a classificação de doenças gastrointestinais, através da integração de *Region of Interest* (ROI), *Convolutional Block Attention Module* (CBAM), Otimização Bayesiana de Hiperparâmetros e técnicas de XAI. Com isso, acredita-se que o método proposto apresenta as seguintes contribuições:

- Integrar extração de ROI e CBAM em modelos otimizados para reduzir ruído e aprimorar o foco em regiões discriminantes;
- Validar a confiabilidade por meio de XAI, demonstrando que a otimização direciona corretamente a atenção da rede para as lesões.

Dessa forma, esta pesquisa propõe uma solução eficaz para auxiliar o diagnóstico médico, garantindo alto desempenho aliado à interpretabilidade das decisões do modelo.

2. Trabalhos Relacionados

As doenças GI exigem diagnósticos rápidos e precisos. Nesse cenário, o uso de CNN tornou-se essencial, evoluindo de arquiteturas simples para modelos híbridos. Apesar do avanço, a área ainda busca mais eficiência e explicabilidade. A seguir, são discutidos os trabalhos voltados a essas patologias.

[Borgli et al. 2019] enfatizam que o *Transfer Learning* (TL) isolado é insuficiente para a detecção robusta de doenças gastrointestinais na base de dados Kvasir, defendendo a *Hyperparameter Optimization* (HPO). Os autores confirmam que a Otimização Bayesiana supera métodos manuais, elevando o desempenho em 10% e atingindo um *F1-Score* de 88,0% ao explorar o espaço de configuração de forma sistemática.

O uso de TL mostrou-se crucial para mitigar a escassez de dados, estratégia corroborada por [Nouman Noor et al. 2023], que alcançaram um *F1-Score* de 95,24% utilizando a arquitetura MobileNetV2, ainda que baseado em uma versão reduzida das bases de dados Kvasir V-2 e Hyper-Kvasir. Para lidar com a sobreposição visual, [Malik et al. 2024] propuseram um *ensemble* híbrido de VGG-19 com uma CNN personalizada que atingiu *F1-Score* de 98,84%, mas também em um cenário simplificado, construído a partir de bases híbridas, ignorando marcos anatômicos desafiadores presentes na base de dados Kvasir.

No cenário de aprimoramento visual, [Elmagzoub et al. 2024] validaram a integração de mecanismos de atenção a uma ResNet101 otimizada via *grid search*, atingindo *F1-Score* de 93,0% na classificação de doenças gástricas. A evolução para arquiteturas que priorizam regiões informativas também foi explorada pelo modelo *Spatial-Attention ConvMixer*, proposto por [Demirbaş et al. 2024]. O estudo demonstra que atribuir pesos de importância espacial nos mapas de características permite ao modelo alcançar *F1-Score* de 93,42%, focando estritamente em áreas relevantes da lesão em um contexto reduzido da base de dados Kvasir.

Recentemente, a demanda por transparência no diagnóstico foi reforçada por [Kamble et al. 2025], que propuseram o uso da EfficientNetB3 com técnicas de XAI. O modelo alcançou *F1-Score* de 94,29% na Kvasir, utilizando validação por divisão simples (*hold-out*). Simultaneamente, o método EndoNet inovou ao propor uma estrutura híbrida de múltiplos estágios, fundindo características de redes como ResNet101, Xception e Inception [Attallah et al. 2025]. Esta abordagem robusta, verificada via validação cruzada (*5-folds*), atingiu um *F1-Score* de 97,79%, estabelecendo um alto padrão de desempenho mediante a fusão complexa de arquiteturas profundas.

Apesar dos avanços reportados, os trabalhos analisados apresentam algumas limitações metodológicas. A maioria dos estudos adota validação *hold-out*, protocolo que pode superestimar o desempenho ao não capturar adequadamente a variabilidade estatística e a capacidade de generalização do modelo em cenários clínicos reais. Além disso, observa-se que técnicas relevantes, como HPO, mecanismos de atenção visual e métodos de explicabilidade, são frequentemente empregadas de forma isolada. Essa utilização fragmentada dificulta a análise do impacto combinado dessas técnicas e limita o desenvolvimento de *pipelines* mais robustos e clinicamente confiáveis.

Diante desse contexto, o presente trabalho propõe um método que integra estratégias complementares. Inicialmente, realiza-se a extração de ROI com o objetivo de reduzir artefatos visuais e padronizar o campo de análise. Em seguida, incorpora-se o CBAM [Woo et al. 2018], permitindo que a rede enfatize regiões espacialmente e semanticamente relevantes. A Otimização Bayesiana via Optuna [Akiba et al. 2019] é empregada para automatizar a busca por hiperparâmetros, reduzindo a dependência de ajustes manuais e aumentando a eficiência do treinamento. A robustez experimental é

assegurada por validação cruzada de *5-folds*, enquanto a interpretabilidade das decisões é garantida pelo AblationCAM [Ramaswamy et al. 2020], fortalecendo a confiabilidade do modelo para aplicações clínicas.

3. Materiais e Método

Nesta seção, são apresentados os materiais e o método proposto, divididos em cinco etapas principais. Primeiramente, descreve-se a base de dados que foi usada para validar o método proposto. Em seguida, detalha-se o pré-processamento focado na extração de ROI. Posteriormente, aborda-se a classificação empregando CNNs integradas ao módulo CBAM e otimizadas via Optuna. Por fim, calculam-se as métricas de validação e aplicam-se técnicas de XAI para interpretar o modelo. A Figura 1 ilustra estas etapas, que são detalhadas a seguir.

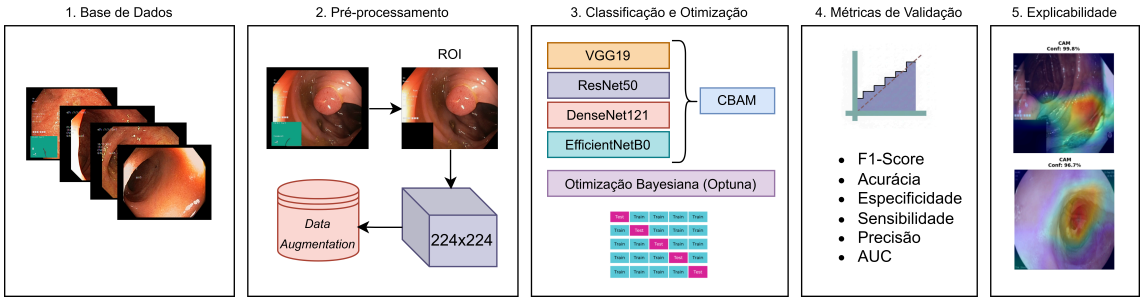


Figura 1. Ilustração do método proposto.

3.1. Base de Dados

Para o desenvolvimento do método, utilizou-se a base de dados Kvasir, que contém 4.000 imagens de exames endoscópicos do trato gastrointestinal [Pogorelov et al. 2017]. A base é composta originalmente por oito classes distintas: pólipos tingidos e elevados (*dyed-lifted-polyps*), margens de ressecção tingidas (*dyed-resection-margins*), esofagite (*esophagitis*), pólipos (*polyps*) e colite ulcerativa (*ulcerative-colitis*) como classes Patológicas; e ceco normal (*normal-cecum*), piloro normal (*normal-pylorus*) e linha Z normal (*normal-z-line*) como classes Saudáveis. As imagens originais possuem resoluções variadas e foram redimensionadas para 224 $\times$ 224 *pixels*, tamanho padrão comumente adotado como entrada para as CNNs [Júnior et al. 2021, Diniz et al. 2024b].

Para o método proposto, utilizaram-se as macro-classes, dividindo-as em duas classes: Patológico e Saudável. Essa estratégia favorece uma triagem clínica objetiva, permitindo a distinção entre tecidos normais e anômalos, aspecto particularmente relevante em cenários de rastreamento. Para validação do modelo, foi adotada a validação cruzada estratificada de *5-folds*, assegurando a preservação da proporção entre classes em todos os *folds* e proporcionando uma estimativa da capacidade de generalização. A Figura 2 apresenta exemplos das imagens Patológicas e Saudáveis.

3.2. Pré-processamento

A literatura mostra que extrair ROI de imagens melhora significativamente os resultados [Diniz et al. 2024a]. O pré-processamento é definido para remover ruídos inerentes aos exames endoscópicos e direcionar o foco do modelo estritamente para as áreas relevantes.

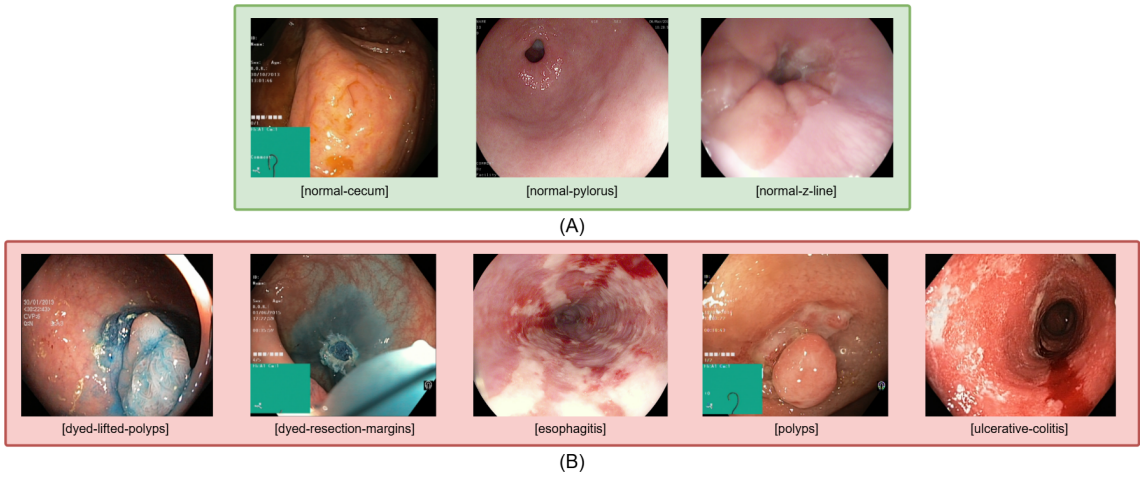


Figura 2. (A) Saudável e (B) Patológico.

Inicialmente, aplica-se um algoritmo para a extração da ROI. Muitas imagens da base de dados contêm bordas pretas e artefatos visuais do equipamento, como caixas de navegação verdes, que não agregam valor preditivo e podem direcionar o modelo para sobreajuste (*overfitting*).

Para suavizar esse ruído, o método converte a imagem para o espaço de cor *Hue, Saturation, Value* a fim de identificar e mascarar os artefatos verdes por meio de limiarização com matiz entre 70 e 95, saturação entre 100 e 255 e valor entre 50 e 255, seguida de detecção de contornos. Após a oclusão dessas marcações, a imagem é convertida para tons de cinza, onde *pixels* com intensidade superior a 15 são considerados região de interesse. O maior contorno detectado é selecionado e a imagem é recortada pelo seu retângulo delimitador (*bounding box*), eliminando as bordas pretas. A Figura 3 ilustra o resultado desse processo, demonstrando a remoção dos artefatos e o recorte da região relevante.

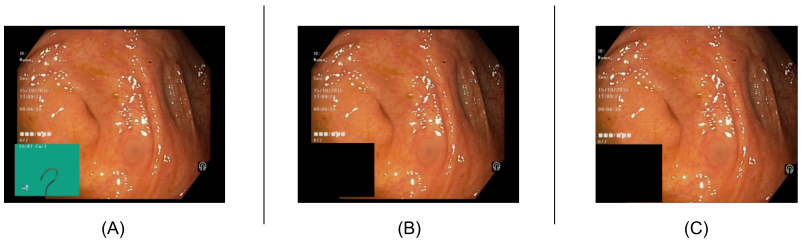


Figura 3. (A) Imagem original, (B) Imagem sem caixa verde e (C) Imagem processada.

Para o conjunto de treinamento, aplicam-se técnicas de aumento de dados (*Data Augmentation*), incluindo rotações aleatórias de até 10 graus e espelhamentos horizontais e verticais. Esse processo visa aumentar a variabilidade das amostras e prevenir o sobreajuste durante o aprendizado. Para lidar com o desbalanceamento entre as classes, emprega-se a amostragem ponderada (*Weighted Random Sampler*), atribuindo a cada amostra um peso inversamente proporcional à frequência de sua classe, de forma que ambas as classes sejam igualmente representadas durante o treinamento.

3.3. Classificação e Otimização

Nesta seção, descreve-se o processo de classificação das imagens e a otimização do modelo. Primeiramente, apresentam-se as arquiteturas CNNs avaliadas, seguido da estratégia de HPO e, por fim, o protocolo de validação adotado.

3.3.1. Arquiteturas Utilizadas

Foram avaliadas quatro arquiteturas consagradas na literatura: VGG19, ResNet50, DenseNet121 e EfficientNetB0. Todas foram inicializadas utilizando TL com pesos pré-treinados, mantendo os parâmetros do *backbone* congelados durante o treinamento. Na abordagem proposta, o módulo CBAM é inserido após o último bloco convolucional de cada *backbone*, adaptando-se ao número de canais de saída de cada arquitetura: 512 (VGG19) [Simonyan and Zisserman 2014], 2048 (ResNet50) [He et al. 2016], 1024 (DenseNet121) [Huang et al. 2017] e 1280 (EfficientNetB0) [Tan and Le 2019]. O CBAM atua sequencialmente inferindo mapas de atenção nos domínios de canal e espacial, permitindo que a rede enfatize características clinicamente relevantes antes de repassar os dados para as camadas de classificação. Apenas os parâmetros do CBAM e da cabeça classificadora (*classifier head*) são atualizados durante o treinamento.

3.3.2. Otimização de Hiperparâmetros

A configuração dos hiperparâmetros de uma CNN influencia diretamente o seu poder de generalização e a eficácia no aprendizado de características complexas. Para superar as limitações inerentes ao ajuste manual empírico, empregou-se o *framework* Optuna [Akiba et al. 2019] para realizar a busca de hiperparâmetros por meio da Otimização Bayesiana. Este método explora o espaço de busca de forma probabilística, avaliando o histórico das tentativas anteriores para sugerir combinações cada vez mais promissoras.

Durante o processo, configurou-se um total de 60 tentativas (*trials*) com o objetivo primário de maximizar a métrica *F1-Score* do modelo. O espaço de busca avaliado incluiu: taxa de aprendizado (*learning rate* - LR), tamanho de lote (*batch size*), taxa de regularização (*dropout*), decaimento de peso (*weight decay*) e a escolha do algoritmo otimizador (Adam e AdamW). Após a identificação dos melhores hiperparâmetros, o treinamento final foi conduzido em todos os *folds*, com ajuste dinâmico de LR via *ReduceLROnPlateau* e parada antecipada (*Early Stopping*) monitorando o *F1-Score* de validação, salvando o *checkpoint* do melhor modelo de cada *fold*.

Na avaliação, 10% das imagens foram reservadas para teste cego, a fim de simular um cenário clínico com imagens que o modelo nunca viu. Os 90% restantes foram submetidos à validação cruzada estratificada de *5-folds*, com cada *fold* dividido em 80% para treino e 20% para validação. O Optuna foi aplicado no primeiro *fold* para reduzir custos.

3.4. Métricas de Validação

Para avaliar o desempenho do método proposto, foram utilizadas métricas de validação amplamente empregadas na literatura para problemas de classificação. São elas: acurácia (ACC), especificidade (ESP), sensibilidade (SEN) ou *recall*, precisão (PRE), *F1-Score* e a área sob a curva ROC (*Area Under the Curve* - AUC).

3.5. Explicabilidade

A explicabilidade é essencial para a aceitação clínica de modelos de AI. Para interpretar as predições da rede, empregou-se o método AblationCAM (*Ablation Class Activation Mapping*) [Ramaswamy et al. 2020]. Diferente de abordagens tradicionais baseadas em gradiente, esta técnica avalia a importância das características removendo ou ocultando (*ablating*) os mapas de ativação individuais da última camada convolucional e medindo o impacto na saída do modelo. O resultado é um mapa de calor que destaca visualmente as regiões da imagem que mais contribuíram para a classificação, fornecendo uma justificativa transparente para a tomada de decisão.

4. Resultados e Discussão

Nesta seção, são descritos os experimentos realizados para validar o método proposto, apresentando as métricas de validação obtidas para cada arquitetura avaliada. Adicionalmente, são realizadas comparações com trabalhos relacionados e estudos de casos com análise qualitativa por meio de mapas de explicabilidade visual.

4.1. Ambiente de Treinamento

Os experimentos foram executados no ambiente de computação em nuvem Google Colab, utilizando GPU Nvidia A100. A implementação foi realizada em linguagem Python, com uso das bibliotecas PyTorch, Torchvision, OpenCV, Scikit-Learn, Optuna, Pandas, NumPy, Matplotlib e WandB.

4.2. Hiperparâmetros e Regularização

Para o *baseline*, os hiperparâmetros foram fixados manualmente, adotando valores padrões utilizados na literatura: LR de 1×10^{-4}, *batch size* de 32, taxa de *dropout* de 0,3, *weight decay* de 1×10^{-4} e otimizador Adam.

Para o método proposto, o Optuna conduziu 60 tentativas (*trials*) de Otimização Bayesiana com 15 épocas cada, maximizando o *F1-Score* médio de validação. No treinamento final foram definidas 40 épocas em *5-folds* que exigiram um tempo computacional de aproximadamente 1 hora por arquitetura. A Tabela 1 apresenta os melhores hiperparâmetros encontrados para cada arquitetura.

Tabela 1. Melhores hiperparâmetros encontrados pelo Optuna.

Arquitetura	*Batch Size*	LR	*Dropout*	Otim.	*Weight Decay*
VGG19	32	$1{,}69 \times 10^{-4}$	0,49	Adam	$2{,}76 \times 10^{-4}$
ResNet50	64	$2{,}96 \times 10^{-4}$	0,40	AdamW	$2{,}82 \times 10^{-3}$
DenseNet121	32	$2{,}57 \times 10^{-4}$	0,51	AdamW	$1{,}72 \times 10^{-4}$
EfficientNetB0	64	$2{,}96 \times 10^{-4}$	0,50	Adam	$1{,}10 \times 10^{-4}$

Observa-se que a Otimização Bayesiana convergiu para LRs superiores ao valor fixo do *baseline*, sugerindo que as arquiteturas se beneficiam de uma exploração mais agressiva do espaço de busca quando combinadas ao CBAM e à extração de ROI. Ademais, ResNet50 e DenseNet121 convergiram para o otimizador AdamW, indicando que o *Weight Decay* desacoplado favorece a regularização nessas arquiteturas de maior capacidade representacional.

4.3. Resultados Experimentais

Para validar o método proposto, foram conduzidos dois experimentos. No primeiro, avalia-se o desempenho das arquiteturas sem as contribuições propostas (*baseline*), utilizando as CNNs em seu estado original. No segundo, avalia-se o método proposto, integrando extração de ROI, CBAM e Otimização Bayesiana de hiperparâmetros via Optuna. Em ambas as abordagens, adotou-se a função de perda *Binary Cross-Entropy with Logits Loss* (BCEWithLogitsLoss), adequada para classificação binária com saída escalar.

4.3.1. Resultados do Baseline

A Tabela 2 apresenta os resultados do *baseline* para cada arquitetura, sem a aplicação de extração de ROI, CBAM ou HPO.

Tabela 2. Resultados do *baseline* (média ± desvio padrão).

Arquitetura	*F1-Score*	ACC	ESP	SEN	PRE	AUC
VGG19	**93,60 ± 1,24**	**93,56 ± 1,27**	92,67 ± 2,03	**93,56 ± 1,27**	**93,82 ± 1,12**	97,95 ± 0,72
ResNet50	91,28 ± 1,39	91,25 ± 1,42	92,00 ± 2,15	91,25 ± 1,42	91,38 ± 1,32	96,75 ± 0,80
DenseNet121	92,94 ± 1,10	92,94 ± 1,10	**94,44 ± 1,84**	92,94 ± 1,10	93,01 ± 1,12	**98,15 ± 0,53**
EfficientNetB0	93,03 ± 1,28	93,00 ± 1,29	93,07 ± 1,14	93,00 ± 1,29	93,12 ± 1,27	97,99 ± 0,64

Os resultados indicam que, sem as contribuições propostas, as arquiteturas apresentam desempenho variável entre si. A VGG19 destacou-se nas métricas de *F1-Score*, acurácia, sensibilidade e precisão, enquanto a DenseNet121 obteve os melhores resultados em especificidade e AUC. A ResNet50 apresentou o menor desempenho geral, com *F1-Score* de 91,28%, indicando maior sensibilidade à ausência de otimização e de mecanismos de atenção visual.

4.3.2. Resultados do Método Proposto

A Tabela 3 apresenta os resultados obtidos pelo método, integrando extração de ROI, CBAM e Otimização Bayesiana de hiperparâmetros via Optuna.

Tabela 3. Resultados do método proposto (média ± desvio padrão).

Arquitetura	*F1-Score*	ACC	ESP	SEN	PRE	AUC
VGG19	93,59 ± 0,83	93,56 ± 0,85	93,02 ± 1,65	93,56 ± 0,85	93,76 ± 0,76	98,10 ± 0,31
ResNet50	**94,33 ± 0,77**	**94,31 ± 0,78**	**93,82 ± 1,09**	**94,31 ± 0,78**	**94,45 ± 0,72**	**98,22 ± 0,36**
DenseNet121	93,95 ± 1,03	93,92 ± 1,05	93,24 ± 1,42	93,92 ± 1,05	94,11 ± 0,96	98,16 ± 0,42
EfficientNetB0	92,75 ± 1,12	92,72 ± 1,14	93,11 ± 2,04	92,72 ± 1,14	92,85 ± 1,01	97,69 ± 0,64

O método proposto demonstrou ganhos expressivos em arquiteturas de maior capacidade representacional. A ResNet50 obteve o melhor desempenho geral, alcançando *F1-Score* de 94,33% e AUC de 98,22%, representando um ganho de 3,05 pontos em *F1-Score* em relação ao seu respectivo *baseline*. A DenseNet121 também apresentou melhora consistente, com ganho de 1,01 pontos em *F1-Score*. A VGG19 manteve desempenho equivalente ao *baseline*, porém com redução expressiva do desvio padrão, de 1,24 para

0,83. A EfficientNetB0 apresentou redução de 0,28 pontos em *F1-Score*, comportamento atribuído à sua arquitetura compacta, que limita o ganho do CBAM em cenários de compressão de canais já otimizada. Notavelmente, todas as arquiteturas avaliadas apresentaram redução do desvio padrão em relação ao *baseline*, evidenciando maior estabilidade e consistência entre os *folds*, o que é especialmente relevante em contextos clínicos onde a confiabilidade do modelo é fundamental.

4.4. Comparação com a Literatura

Conforme mencionado na Seção 2, diversos estudos abordam a classificação de doenças gastrointestinais. Um resumo das técnicas e os resultados são apresentados na Tabela 4.

Tabela 4. Comparação com os trabalhos relacionados.

Trabalho	**Técnica**	**Validação**	***F1-Score***	**ACC**
[Borgli et al. 2019]	HPO + TL	*Hold-out*	88,00%	97,00%
[Elmagzoub et al. 2024]	ResNet101 + Atenção	*Hold-out*	93,00%	93,50%
[Demirbaş et al. 2024]	ConvMixer + Atenção	*Hold-out*	93,42%	93,37%
[Kamble et al. 2025]	EfficientNetB3 + XAI	*Hold-out*	94,29%	94,25%
[Nouman Noor et al. 2023]	CLAHE + MobileNetV2	*10-folds*	95,24%	96,40%
[Attallah et al. 2025]	Multi-CNN + mRMR	*5-folds*	97,79%	97,80%
[Malik et al. 2024]	VGG19 + CNN	*Hold-out*	98,84%	99,45%
Método Proposto	**ROI + CBAM + Optuna**	***5-folds***	**94,33%**	**94,31%**

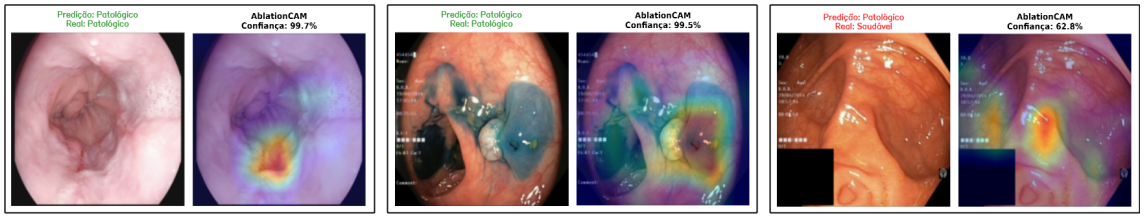
A análise da Tabela 4 evidencia que a maioria dos trabalhos relacionados adota validação *hold-out*, protocolo que tende a produzir estimativas otimistas do desempenho real. Trabalhos como [Malik et al. 2024], que reportam *F1-Score* de 98,84%, utilizam base de dados híbridas com número reduzido de classes, o que pode favorecer artificialmente métricas elevadas. Além disso, em nenhum dos métodos anteriores fica evidenciada a extração de artefatos visuais do equipamento endoscópico, que em muitas vezes direcionam o resultado do modelo gerando *overfitting*, uma vez que trazem padrões diretos de cada classe.

O EndoNet [Attallah et al. 2025], único trabalho com protocolo equivalente ao proposto, atingiu 97,79% mediante fusão de três arquiteturas profundas e *Minimum Redundancy Maximum Relevance* (mRMR), gerando elevada complexidade computacional. O método proposto alcança *F1-Score* de 94,33% com validação cruzada de *5-folds*, arquitetura única otimizada e explicabilidade visual via AblationCAM, constituindo uma alternativa eficiente e interpretável para aplicação clínica.

4.5. Estudo de Casos

Nesta seção, são apresentados três estudos de caso com as predições da ResNet50 obtidas pelo resultado do melhor dos *5-folds*. A Figura 4 ilustra dois casos corretos e um caso de erro, acompanhados dos respectivos mapas de ativação gerados pelo AblationCAM.

Em (A) e (B), o modelo classificou corretamente com confiança de 99,7% e 99,5%. Os mapas de ativação mostram a atenção para as regiões de lesão presentes nas imagens, demonstrando que o método é capaz de identificar corretamente as características discriminantes das patologias. No exemplo (C), o modelo classificou erroneamente uma imagem Saudável como Patológica, com confiança reduzida de 62,8%.



(A) (B) (C)

Figura 4. (A) e (B) Acertos com alta confiança e (C) Erro com baixa confiança.

O mapa revela ativação difusa e sem foco definido, indicando que a rede não encontrou padrões discriminantes claros, o que resultou na incerteza da predição. A baixa confiança associada ao erro indica que o método possui potencial para integração com limiares de decisão clínica, nos quais predições de baixa confiança poderiam ser encaminhadas para revisão especializada.

5. Conclusão

Este trabalho apresentou um método para classificação de doenças gastrointestinais na base de dados Kvasir, integrando extração de ROI, CBAM e Otimização Bayesiana de hiperparâmetros via Optuna. Os resultados demonstram a eficácia do método, com a ResNet50 alcançando *F1-Score* de 94,33% e AUC de 98,22% com validação cruzada de *5-folds*, representando um ganho de 3,05 pontos em relação ao *baseline*. A redução do desvio padrão em todas as arquiteturas mostra maior estabilidade diagnóstica, e a aplicação do AblationCAM fornece transparência às decisões do modelo.

Como trabalhos futuros, propõem-se abordagens de *ensemble* combinando as arquiteturas para aprimorar o desempenho, embora seja necessário avaliar o impacto na viabilidade em dispositivos de baixo custo. Paralelamente, planeja-se incorporar novos filtros de pré-processamento, classificação em cascata das classes patológicas e um estudo de ablação aprofundado. Sugere-se ainda a avaliação do método em outras bases de dados para verificar sua capacidade de generalização.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, Fundação de Amparo a Pesquisa do Maranhão (FAPEMA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Ainda, reconhecemos o uso do LLM para revisão ortográfica, correção gramatical e assistência na tradução de termos específicos.

Referências

Aguiar, R. M., Scheeren, M. H., de Araujo Junior, S. L., Mendes, E., de Paula Filho, P. L., and Franco, R. A. (2024). Aplicaçao de modelos de aprendizado profundo para a segmentaçao semântica de imagens de colonoscopia. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 389–399. SBC.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2623–2631.

Attallah, O., Aslan, M. F., and Sabanci, K. (2025). Endonet: A multiscale deep learning framework for multiple gastrointestinal disease classification via endoscopic images. *Diagnostics*, 15(16):2009.

Borgli, R. J., Stensland, H. K., Riegler, M. A., and Halvorsen, P. (2019). Automatic hyperparameter optimization for transfer learning on medical image datasets using bayesian optimization. In *2019 13th international symposium on medical information and communication technology (ISMICT)*, pages 1–6. IEEE.

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., and Jemal, A. (2024). Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263.

Demirbaş, A. A., Üzen, H., and Fırat, H. (2024). Spatial-attention convmixer architecture for classification and detection of gastrointestinal diseases using the kvasir dataset. *Health Information Science and Systems*, 12(1):32.

Diniz, J. O., Dias Jr, D. A., da Cruz, L. B., Marques, R. C., Gomes Jr, D. L., Cortês, O. A., de Carvalho Filho, A. O., and Quintanilha, D. B. (2024a). Efficientensemble: Diagnóstico de câncer de mama em imagens de ultrassom utilizando processamento de imagens e ensemble de efficientnets. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 202–213. SBC.

Diniz, J. O., Ribeiro, N. P., Junior, D. A. D., da Cruz, L. B., de Carvalho Filho, A. O., Gomes Jr, D. L., Silva, A. C., and de Paiva, A. C. (2024b). Efficientxyz-deepfeatures: seleção de esquema de cor e arquitetura deep features na classificação de câncer de cólon em imagens histopatológicas. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 82–93. SBC.

El-Bouzaidi, Y. E. I., Hibbi, F. Z., and Abdoun, O. (2025). Optimizing convolutional neural network impact of hyperparameter tuning and transfer learning. In *Innovations in Optimization and Machine Learning*, pages 301–326. IGI Global Scientific Publishing.

Elmagzoub, M. A., Kaur, S., Gupta, S., Rajab, A., Rajab, K. D., Al Reshan, M. S., Alshahrani, H., and Shaikh, A. (2024). Improving endoscopic image analysis: Attention mechanism integration in grid search fine-tuned transfer learning model for multi-class gastrointestinal disease classification. *IEEE Access*, 12:80345–80358.

Gautam, S., Thambawita, V., Riegler, M., Halvorsen, P., and Hicks, S. (2025). Medico 2025: Visual question answering for gastrointestinal imaging. *arXiv preprint arXiv:2508.10869*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.

Júnior, D. A. D., da Cruz, L. B., Diniz, J. O. B., da Silva, G. L. F., Junior, G. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., and Gattass, M. (2021). Automatic method

for classifying covid-19 patients based on chest x-ray images, using deep features and pso-optimized xgboost. *Expert Systems with Applications*, 183:115452.

Kamble, A., Bandodkar, V., Dharmadhikary, S., Anand, V., Sanki, P. K., Wu, M. X., and Jana, B. (2025). Enhanced multi-class classification of gastrointestinal endoscopic images with interpretable deep learning model. *arXiv preprint arXiv:2503.00780*.

Maida, M., Marasco, G., Maas, M., Ramai, D., Spadaccini, M., Sinagra, E., Facciorusso, A., Siersema, P., and Hassan, C. (2025). Effectiveness of artificial intelligence assisted colonoscopy on adenoma and polyp miss rate: A meta-analysis of tandem rcts. *Digestive and Liver Disease*, 57(1):169–175.

Malik, H., Naeem, A., Sadeghi-Niaraki, A., Naqvi, R. A., and Lee, S.-W. (2024). Multi-classification deep learning models for detection of ulcerative colitis, polyps, and dyed-lifted polyps using wireless capsule endoscopy images. *Complex & Intelligent Systems*, 10(2):2477–2497.

Nouman Noor, M., Nazir, M., Khan, S. A., Song, O.-Y., and Ashraf, I. (2023). Efficient gastrointestinal disease classification using pretrained deep convolutional neural network. *Electronics*, 12(7):1557.

Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Schmidt, P. T., et al. (2017). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 164–169.

Ramaswamy, H. G. et al. (2020). Ablation-cam: Visual explanations for deep convolutional network via gradient-free localization. In *proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 983–991.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Su, C.-C., Chou, C.-K., Mukundan, A., Karmakar, R., Sanbatcha, B. F., Huang, C.-W., Weng, W.-C., and Wang, H.-C. (2025). Capsule endoscopy: Current trends, technological advancements, and future perspectives in gastrointestinal diagnostics. *Bioengineering*, 12(6):613.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Wang, S., Zheng, R., Li, J., Zeng, H., Li, L., Chen, R., Sun, K., Han, B., Bray, F., Wei, W., et al. (2024). Global, regional, and national lifetime risks of developing and dying from gastrointestinal cancers in 185 countries: a population-based systematic analysis of globocan. *The Lancet Gastroenterology & Hepatology*, 9(3):229–237.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19.