Journal of Engineering Managment and Soft Computing (JEMSC)

# Dual-Branch Hybrid-Feature Learning for Imbalance-Aware Paediatric Colonoscopy Decision Support with Interpretable Multi-Scale Networks

**Khosro Rezaee**[1]✉

1. Department of Biomedical Engineering, Meybod University, Meybod, Iran. Email: kh.rezaee@meybod.ac.ir

**Article Info**

**Article type:**
Research Article

**Article history:**
Received 27 November 2025
Received in revised form 11 February 2026
Accepted 4 March 2026
Published online 1 April 2026

**Keywords:**
Pediatric Colonoscopy, Colorectal Polyps, Deep Learning, Class Imbalance, Interpretability.

**ABSTRACT**

Colorectal neoplasia, while rare, is high-risk in children. As such, colonoscopic detection of paediatric polyps is of clinical importance. Current deep learning–based computer-aided systems are trained on adult data and do not provide solutions for class imbalance, cross-centre generalisation or interpretability in paediatrics. We introduce an imbalance-aware dual-branch hybrid-features (IDHF) framework for polyp versus non-polyp classification. IDHF, based on a ResNet-50 backbone, is equipped with complementary texture branch to augment deep semantic features, adaptive gating for late fusion and the optimisation of a class-weighted loss function with false-negative penalty and branch-agreement regularisation. The proposed model is trained on CP-CHILD-A (8,000 images, 1,000 polyp/7,000 non-polyp) and tested without fine-tuning on CP-CHILD-B (1,500 images, 400 polyp/1,100 non-polyp) on another platform. ResNet-50+IDHF achieves 99.6% accuracy, 100.0% sensitivity, 99.5% specificity, 96.8% precision, and a 98.4% F1-score on CP-CHILD-A. ResNet-50+IDHF achieves 99.5% accuracy, 99.2% sensitivity, 99.6% specificity, 99.0% precision, and a 99.1% F1-score on CP-CHILD-B, supporting a robust and interpretable solution for computer-aided paediatric polyp detection.

**Cite this article:** Rezaee, Kh., (2026)., Dual-Branch Hybrid-Feature Learning for Imbalance-Aware Paediatric Colonoscopy Decision Support with Interpretable Multi-Scale Networks. *Engineering Management and Soft Computing,* 12 (2). 204-224.

**DOI**: https://doi.org/10.22091/jemsc.2026.14688.1327


DOI: https://doi.org/10.22091/jemsc.2026.14688.1327

**Publisher**: University of Qom

## 1) Introduction

Colorectal cancer (CRC) is one of the most common and deadly malignancies worldwide, with over 1.9 million new cases and more than 935,000 deaths in 2020, and is expected to exceed 3 million cases and 1.6 million deaths per year by 2040 (Xi & Xu, 2021). Paediatric CRC accounts for only a small proportion of the overall burden but is nevertheless an important clinical problem as it is often diagnosed at a late stage, presents with aggressive features, and has a substantial mortality and long-term morbidity burden (Sultan et al., 2010; NCI, 2024). Paediatric and adolescent patients with carcinoma of the large bowel are rare (approximately 0.5 cases per 100,000 children less than 20 years of age, and fewer than 100 new cases each year in the United States), and late diagnosis, high rates of synchronous and metachronous malignancy, and substantial long-term morbidity among those who survive (NCI, 2024). In this context, and against a backdrop of an increasing incidence of early-onset CRC among young adults, these trends have contributed to renewed attention to the optimisation of early detection and prevention strategies across the age spectrum.

In the paediatric population, colonoscopy is the key diagnostic and therapeutic procedure for lower gastrointestinal symptoms and hereditary cancer syndromes. Large studies have shown diagnostic yields of approximately 70–80% in symptomatic children (Lee et al., 2019; Wu et al., 2015), with colorectal polyps one of the most common findings, particularly in patients with lower gastrointestinal bleeding (Wu et al., 2015). While juvenile polyps are typically benign and solitary, adenomatous and syndromic polyps are also well-recognised, and their presence confers a non-trivial lifetime risk of CRC (Kay & Eng, 2015). Contemporary, multi-centre cohorts have confirmed that rectal bleeding and diarrhoea are the most common indications for colonoscopy in this population, and that polyps, colonic inflammation, and inflammatory bowel disease are responsible for the majority of endoscopic diagnoses in children (Altamimi et al., 2022; Dereci et al., 2021). As deep sedation or general anaesthesia is usually required, children have limited tolerance for lengthy procedures, and the calibre of the paediatric colon is much smaller, every missed lesion in this context carries a disproportionately high clinical cost.

Mounting evidence from adult practice has also shown that higher adenoma detection rates (ADR) result in fewer interval cancers and lower CRC mortality: in one large, prospective study, Corley et al. (2014) found that a 1% increase in ADR was associated with a 3% lower risk of interval CRC. While ADR is not defined directly for the paediatric population, other quality metrics with analogous definitions and strong clinical implications—such as polyp detection rate and complete mucosal inspection—are highly relevant when evaluating and managing children with hereditary syndromes, longstanding colitis, or alarm symptoms. Maximising lesion detection in paediatric colonoscopy is therefore not just a technical aspiration but a prerequisite for successful primary and secondary prevention in a population where the disease is uncommon but often more aggressive when it does occur (Sultan et al., 2010; NCI, 2024).

Recent advances in deep learning–based computer-aided detection (CADe) algorithms have shown consistently promise in adult colonoscopy to reduce inter-operator variability, provide real-time visual cues, and enable continuous quality monitoring (Taghiakbari et al., 2021). Comprehensive reviews by Taghiakbari et al. (2021) describe how convolutional neural networks (CNN) trained on several hundred thousand colonoscopy frames can achieve sensitivities in the 90% range under routine clinical conditions. Randomised controlled trials in adult colonoscopy have shown that these systems increase polyp and adenoma detection without prolonging procedure time, and large international tandem studies have demonstrated reductions in missed polyp rates, including for non-advanced lesions (Maas et al., 2024). However, all commercially available and research-grade CADe models have been developed, trained, and validated in adult populations. Given the important differences in lesion spectrum (e.g., juvenile polyps and colonic inflammation are much more common in paediatric practice), endoscopic anatomy, and case mix between children and adults, the safety and performance of adult algorithms applied directly to paediatric video is not guaranteed and requires investigation.

Moreover, several important methodological gaps in the published adult literature become even more pronounced when developing classification-based decision-support solutions for children. First, existing CADe solutions are typically optimised for frame-level localisation (detection) and do not

consistently integrate accurate pixel-wise delineation or clinically meaningful classification of paediatric polyp subtypes within a single, lightweight architecture. Training datasets are generally collected at one or a small number of centres, have very limited device diversity, and under-represent important real-world sources of variation and artefacts, such as suboptimal bowel preparation, motion blur, and specular highlights. These biases can cause significant overfitting to centre-specific imaging characteristics and lead to further performance loss when models are exposed to data with a different distribution (i.e., severe class imbalance between polyp and non-polyp frames and over-representation of the dominant polyp subtypes at training centres). In children—where existing datasets are much smaller and the underlying disease spectrum is much more skewed—these limitations translate into an even greater risk of unstable model performance and trust at the point of care. At the same time, model interpretability is not directly addressed beyond simple qualitative heatmaps, and the computational efficiency of existing CADe models is not systematically aligned with the hardware constraints of typical paediatric endoscopy units.

To address these challenges, we present a paediatric-specific, generalisable, and interpretable deep learning framework for real-time frame-level binary classification (polyp vs non-polyp) in children. Our model is trained and evaluated on two newly curated, multi-centre datasets (CP-CHILD-A and CP-CHILD-B) of expertly annotated still images and video frames from paediatric colonoscopies performed for a broad range of indications, across a wide age range, on different endoscope vendors, and with varying bowel preparation qualities. The architecture consists of a lightweight classification backbone with a binary prediction head, employs explicit regularisation and curriculum-style sampling to prevent overfitting, and uses loss re-weighting to counteract class imbalance between polyp and non-polyp frames. In addition, we include attention-based visual explanations (Grad-CAM) and confidence/uncertainty estimates derived from the model outputs to improve interpretability and trust at the point of care. As a result, the proposed system is designed to achieve high sensitivity for polyp presence recognition even for subtle lesions common in paediatric practice, and to generalise robustly across centres and hardware configurations, making it a practical and trustworthy AI assistant for everyday paediatric colonoscopy.

The remainder of this article is structured as follows. Section 2 presents related work. Section 3 describes our proposed method, including network architecture, training and inference details, and interpretability mechanisms. Section 4 presents the CP-CHILD-A and CP-CHILD-B datasets, experimental protocol, quantitative and qualitative results (including cross-dataset generalisation analyses and ablation studies), and discussion. Finally, Section 5 concludes the paper.

## 2) Literature Review

A growing body of deep learning approaches to fully automated colorectal polyp analysis has been developed in recent years, for the joint tasks of detection, segmentation and classification in both static images and full colonoscopy videos. The reviewed approaches were increasingly characterised by innovative architectural designs (including attention-augmented encoder–decoder and transformer models, as well as real-time one-stage detectors) that sought to maximise sensitivity, specificity, and robustness under a range of challenging endoscopic imaging conditions. Here, we have discussed and compared a series of representative recent systems in terms of the adopted model designs, dataset size and diversity, quantitative performance and real-time suitability.

Shen et al. (2023) described EndoAIM, a multi-centre system that used a CNN detector and EfficientNet-B0 classifier trained on 256,220 colonoscopy images from 5,000 patients to automatically detect and classify polyps. The system internally achieved sensitivity 0.9709 and specificity 0.9701 for detection and AUC 0.9989 for classification; on a three-centre external test set, it maintained lesion-level sensitivity 0.9516 and frame-level specificity 0.9720. The use of a large-scale cohort and a multi-centre setting were both important strengths, while limitations included a focus on static images and a lack of both real-time and explainability analysis.

Handa et al. (2023) introduced WD-MCPI, a mixed-convolution architecture that blended standard and mixed convolutions with tuned hyperparameters for detection of colorectal polyps. On the Etis-

Larib, CVC-Colon, Kvasir v1 and Gastrointestinal Atlas–Colon Polyp test sets, the model reported an accuracy of 94.23%, a precision of 91.16%, a recall of 94.00%, a specificity of 92.67%, a F1 of 91.75%, and an AUC of 92.53%. The approach exhibited robustness across occluded frames and various imaging conditions, but the architectural complexity and lack of real-time clinical testing limited its immediate applicability.

Hamza et al. (2025) proposed an encoder–decoder segmentation network that employed dual attention by combining spatial and channel-wise attention modules with an optimised squeeze-and-excitation block. On Kvasir-SEG and CVC-ClinicDB, the model achieved 0.9054 and 0.9277 mean IoU, 0.9006 and 0.9128 Dice, and 0.9806 and 0.9907 accuracies. The design effectively localised salient regions and exhibited strong benchmark performance, but the relatively heavy model and lack of systematic video-level evaluation limited the evidence for real-time deployment.

Ahamed et al. (2024) reported on an attention-guided MultiResUNet for automatic polyp segmentation that combined residual skip connections, hybrid attention, and test-time augmentation (TTA). On Kvasir-SEG and CVC-ClinicDB, this model attained 0.9546 accuracy, 0.8557 Dice, 0.8824 IoU, 0.8221 recall, 0.8922 precision, and 0.9454 ROC-AUC without TTA; with TTA, these metrics increased to 0.9993 accuracy, approximately 0.866 Dice, and roughly 0.959 ROC-AUC. The very low parameter count (~0.47M) and compact size (~6.7 MB) supported deployment, but the heavy pre-processing and focus on static frames rather than videos were also limiting.

Mozaffari et al. (2024) presented ColonGen, a segmentation framework that aimed to explicitly improve cross-dataset generalisation by combining two complementary Vision Transformers trained on a large aggregated dataset from several public sources. ColonGen-V1 and ColonGen-V2 achieved improvements of 5.1%, 1.3%, and 1.1% over state-of-the-art baselines on ETIS-Larib, Kvasir-SEG and CVC-ColonDB, respectively, in both in-domain and out-of-domain conditions. The explicit consideration of domain shift was a key strength, but the transformer-based design was computation-heavy and could impede adoption in low-resource endoscopy units.

Sushama and Menon (2024) proposed a flexible two-mode framework that could operate in detection-only or joint detection-and-segmentation modes, with optional inpainting to address specular highlights. Across several datasets, including Kvasir-SEG and CVC-ClinicDB, the method achieved precision, recall, F1 and F2 scores of at least about 75% in all settings. This design was conceptually attractive, but strong reliance on the inpainting step and the lack of detailed frame rate and latency reporting obscured the readiness of the approach for routine real-time use.

Saad et al. (2024) introduced PolyDSS, a clinical decision support system that jointly performed multiclass polyp segmentation and classification through a DeepLabv3+-style backbone in a multi-task architecture. On Kvasir-SEG, CVC-ClinicDB and private cohorts, PolyDSS reported a Dice score of about 0.9244 for segmentation and an accuracy of about 0.9425 for polyp type classification tasks (e.g. adenoma, hyperplastic). The integration of morphological and clinically relevant labels was a key advantage, but the computational burden and lack of extensive multi-centre validation remained important limitations.

Zhu et al. (2024) presented PAM-Net for colorectal image analysis and polyp diagnosis, with a parallel attention module that highlighted polyp-related semantics and a geodesic-wise distance loss to improve localisation. Across several colonoscopy datasets, PAM-Net outperformed one-stage and two-stage baselines in detection accuracy and reduced missed lesions. The joint modelling of uncertainty and spatial position was the main innovation, but the limited dataset-wise breakdown of metrics complicated direct comparison with other state-of-the-art approaches, and aspects of generalisability remained under-documented.

Tan et al. (2024) introduced an Enhanced Scattering Wavelet CNN (ESWCNN) for colonoscopy polyp classification that fused scattering wavelet transforms with learnable CNN filters. On three datasets (two public and one private), this method attained a 96.4% accuracy for three-class classification (adenoma, hyperplastic, serrated) and a 94.8% accuracy for binary polyp versus non-polyp classification, with a mean sensitivity of 96.7% and a specificity of 93.1%. The wavelet-based spectral

features enhanced robustness to illumination and texture variations, but the model was classification-only and could not provide pixel-wise localisation to guide polypectomy.

Chen et al. (2022) presented DeFrame, a real-time polyp detection system for colonoscopy videos from multiple centres, built around a deep CNN detector and pre-processing for deblurring and artefact removal. DeFrame obtained high sensitivity and specificity on static frames and live video, and was on par with or better than commercial systems in prospective evaluations. Its industrial design and explicit real-time validation were clear strengths, but the lack of public code and data as well as the detection-only focus without fine-grained segmentation limited reproducibility and follow-on analysis.

Pacal et al. (2022) proposed a real-time colonic polyp detection framework based on improved YOLOv3 and YOLOv4 architectures, which were integrated with Cross Stage Partial Networks, the SiLU activation, and a Complete IoU loss and trained on large-scale datasets, including SUN, PICCOLO, and MICCAI 2015. The resulting models outperformed vanilla YOLOv3/YOLOv4 and prior CNN-based detectors in precision, recall, F1-score, and inference speed. The systematic use of negative samples and architectural optimisation for endoscopic imagery were both noteworthy, but the lack of in-depth analysis of generalisation to strongly out-of-distribution centres and model interpretability was also notable.

In prior research, Pacal and Karaboga (2021) designed a real-time automatic polyp detector for colonoscopy videos by scaling and tuning YOLOv4. On MICCAI 2015 and private datasets, this system yielded higher F1-scores and FPS rates than previous CNN-based detectors, suggesting suitability for continuous video monitoring. However, the heavy reliance on a small number of centres and datasets and the lack of explicit multi-centre domain-shift experiments limited the conclusions on broader clinical robustness and transferability.

Hsu et al. (2021) presented a deep learning pipeline that first converted colour colonoscopy images to grayscale before feeding them to a CNN for colorectal polyp detection and classification. On a mixed dataset of static images and video frames, the grayscale network consistently outperformed vanilla colour CNNs in accuracy and F1-score. Intensity-based representations designed to avoid confounding colour and illumination variability, thus appeared to improve robustness, but the pipeline was not fully optimised for low-latency deployment and was not benchmarked on larger public resources such as Kvasir-SEG.

Li et al. (2021) released a large-scale colonoscopy polyp detection and classification dataset with comprehensive bounding-box and class annotations and reported benchmark results for detectors such as Faster R-CNN and YOLO alongside CNN-based classifiers. The resource served as a standardised benchmark that enabled comparison through metrics such as mAP and AUC and supported more consistent evaluation across studies. However, the baseline models themselves were relatively simple compared with contemporary transformer- and attention-based approaches and were intended mainly as reference points rather than state-of-the-art solutions.

Jha et al. (2021) proposed an end-to-end system for real-time polyp detection, localisation and segmentation based on a ResUNet++ architecture that performed pixel-wise segmentation and generated bounding boxes from the predicted masks. This method achieved competitive Dice and IoU scores compared with specialised segmentation models on several public datasets while maintaining throughput sufficient for live colonoscopy video. The unified treatment of detection and segmentation was attractive for practice, but the network was computationally demanding and typically required GPU resources for smooth operation in routine screening.

Gelu-Simeon et al. (2025) deployed RTPoDeMo, a YOLACT-derived real-time polyp delineation model, on prospective colonoscopy videos, achieving a per-image accuracy of 99.6%, a sensitivity of 90.6%, a specificity of 99.9%, and F1 of 0.94 at clinically acceptable frame rates. Despite these strong results, the authors reported residual missed flat lesions and called for larger multi-centre trials to establish external validity and to characterise performance across vendors, operators, and patient groups.

Collectively, the existing methods reveal that both deep convolutional and transformer-based models are capable of highly accurate polyp detection, segmentation and classification, often at or near

real-time frame rates on well-curated datasets. However, most systems optimise only a subset of the clinically relevant requirements: many are trained on static images, few are rigorously multi-centre validated or strike an ideal balance between computational efficiency and benchmark performance, and only a minority offer tightly integrated detection–segmentation–classification within a single lightweight model. Therefore, robust generalisation across centres, devices and acquisition protocols, as well as consistent performance under the strict computational budgets that are typical of endoscopy suites, remains insufficiently addressed. Our method is designed explicitly to close this gap by providing a unified and resource-efficient architecture that jointly addresses polyp localisation and delineation, exhibits strong performance across heterogeneous datasets, and is optimised for practical real-time deployment in routine clinical practice.

## 3) Methodology

Figure 1 shows the overall pipeline of the proposed method. It can be divided into four parts: (1) dataset acquisition and description of CP-CHILD-A/B; (2) the baseline deep learning (DL) method of ResNet-50 with imbalance processing; (3) the proposed IDHF dual-branch hybrid feature network with imbalance-aware objective (loss); (4) training procedure and hyper-parameter settings. Blocks 1-4 collectively detail the processes from dataset preparation to model architecture and optimization, culminating in frame-level binary classification evaluation (polyp vs non-polyp).

### 3-1) Data

The CP-CHILD dataset (Wang et al., 2020) is a paediatric colonoscopy image collection consisting of two subsets, CP-CHILD-A and CP-CHILD-B, acquired using different endoscope models (Olympus PCF-H290DI and FUJIFILM EC-530wm, respectively), which introduces natural cross-domain variation. CP-CHILD-A contains 8,000 RGB frames (7,000 Non-Polyp and 1,000 Polyp). For the experiments in this paper, we created a stratified split of CP-CHILD-A into training (4,900 Non-Polyp / 700 Polyp), validation (1,050 Non-Polyp / 150 Polyp), and internal test (1,050 Non-Polyp / 150 Polyp) sets. Oversampling was applied only to the training set to mitigate class imbalance, while validation and test sets were kept unchanged. Due to the retrospective nature of the dataset and the limited availability of patient-level identifiers, splitting was performed at the frame level rather than the patient level; we therefore report cross-dataset evaluation on CP-CHILD-B as an additional robustness check. CP-CHILD-B provides an additional 1,500 RGB frames (1,100 Non-Polyp / 400 Polyp) and is used as an unseen external dataset to evaluate generalisation; no CP-CHILD-B samples were used for training, validation, or hyper-parameter tuning.

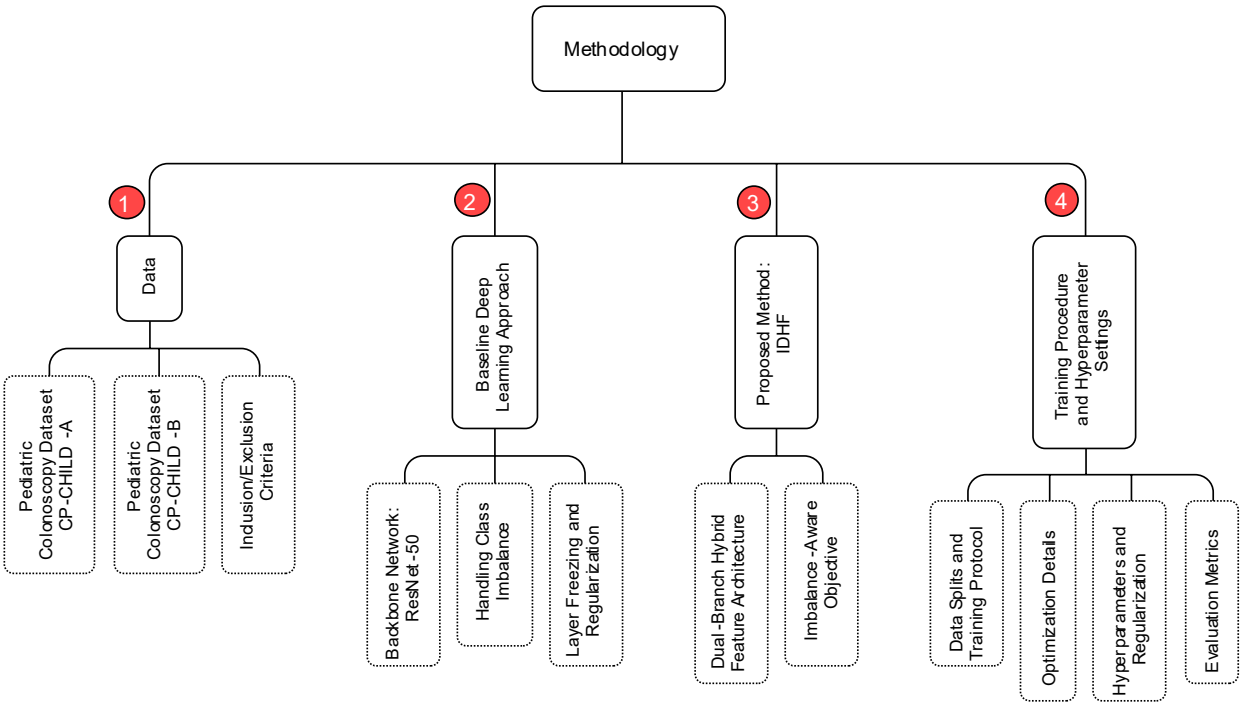


**Figure 1. Overview of the Proposed Methodology for Pediatric Polyp Classification**

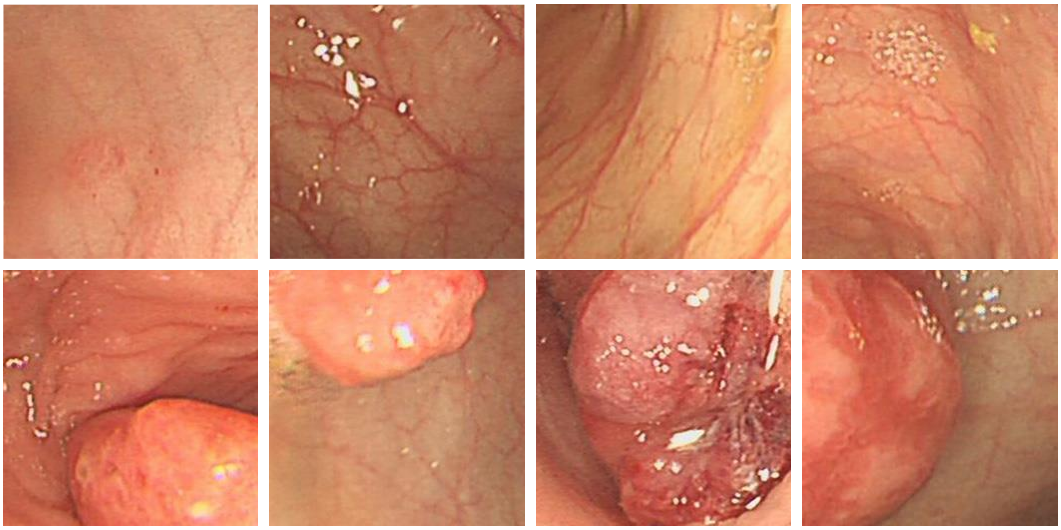

**Figure 2. Example Pediatric Colonoscopy Images from the CP-CHILD Dataset, Where the Top Row Shows Non-Polyp Cases and the Bottom Row Shows Polyp Cases**

### 3-2) Baseline Deep Learning Approach

**A. Backbone Network:** Let $x \in \mathbb{R}^{H\times W\times 3}$ be an input colonoscopy image, $f_{\boldsymbol{\theta}}(\cdot)$ the ResNet-50 feature extractor (pretrained on ImageNet), and $g_{\phi}(\cdot)$ the new two-class classification head:

$$\mathbf{z} = g_{\phi}(f_{\boldsymbol{\theta}}(x)) \in \mathbb{R}^2 \quad (1)$$

The predicted class probabilities for Polyp ($k = 1$) and Non-Polyp ($k = 0$) are:

$$p_k = \text{softmax}(\mathbf{z})_k = \frac{\exp(z_k)}{\sum_{j-0}^{1}\exp(z_j)}, k \in \{0,1\} \quad (2)$$

**B. Handling Class Imbalance in the Baseline:** Let $n_k$ denote the number of training samples in class $k$ (with $n_1 \ll n_0$). Oversampling Polyp is equivalent to sampling each class with probability

$$\pi_k = \frac{n_k^{-1}}{\sum_{j-0}^{1} n_j^{-1}}, k \in \{0,1\} \quad (3)$$

and using corresponding class weights in the cross-entropy loss:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{k-0}^{1} w_k y_k \log p_k, w_k \propto \pi_k^{-1} \quad (4)$$

where $\mathrm{y} = (y_0, y_1)$ is the one-hot label vector. During training, data augmentation—including random rotation, translation, scaling, and horizontal flipping—is applied to input samples $x$only on the oversampled training set before passing them to $f_{\boldsymbol{\theta}}$. This augmentation is used to mitigate potential overfitting caused by duplicated minority-class samples introduced through oversampling, and does not alter the form of the loss function.

**C. Layer Freezing and Regularization:** Let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{\mathrm{frz}}, \boldsymbol{\theta}^{\mathrm{trn}}\right)$ be the backbone parameters, where the early layers $\boldsymbol{\theta}^{\mathrm{frz}}$ are frozen:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^{\mathrm{frz}}} = \mathbf{0} \quad (5)$$

and only $\boldsymbol{\theta}^{\mathrm{trn}}$ and $\boldsymbol{\phi}$ are updated. Dropout acts as stochastic masking inside $g_{\phi}$, while L 2 regularization is applied to the trainable weights:

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + \lambda(\|\boldsymbol{\theta}^{\mathrm{trn}}\|_2^2 + \|\boldsymbol{\phi}\|_2^2) \quad (6)$$

where $\lambda > 0$ is the weight decay coefficient.

### 3-3) Imbalance-Aware Dual-Branch Hybrid Features

We observe that existing colonoscopy–inspired workloads are both (1) imbalanced, e.g., polyps are rare and diverse; and (2) long-tailed in the visual spectrum, e.g., polyps often do not visually pop-out. Inspired by this intuition, the proposed IDHF framework holistically targets both of these challenges through a dual-branch hybrid representation and an explicitly imbalance-aware learning objective. On the representation front, a dual-branch architecture hybridizes convolutional features with handcrafted texture–frequency representations to extract complementary visual information. Importantly, an adaptive gate then explicitly amplifies the handcrafted feature contribution on ambiguous, minority-class (polyp)–like samples so that key polyp-discriminative patterns are not diluted by the preponderance of non-polyp samples. On the optimization front, a dedicated false-negative penalty for the polyp class as well as a branch agreement loss bias learning towards high polyp sensitivity while regularizing both branches, explicitly account for imbalance while reduce missed pediatric polyps vs. conventional single-branch transfer-learning baselines.

**A. Dual-Branch Hybrid Feature Architecture:** Let $x \in \mathbb{R}^{H \times W \times 3}$ be an input pediatric colonoscopy image.

*1. Branch 1 - Deep Semantic*

$$\mathbf{h}_{\mathrm{CNN}} = f_{\boldsymbol{\theta}}(x) \in \mathbb{R}^d \quad (7)$$

where $\boldsymbol{f_{\theta}}$ denotes the ResNet-50 backbone followed by global average pooling and a fully connected layer producing a $d$-dimensional semantic embedding.

*2. Branch 2 - Texture/Frequency*

$$\mathbf{v} = \psi(x), \mathbf{h}_{\mathrm{Tex}} = \phi(\mathbf{v}) \in \mathbb{R}^d \quad (8)$$

Here $\psi(\cdot)$ extracts handcrafted descriptors (LBP/GLCM, Gabor/wavelet energies), and $\phi(\cdot)$ is a small MLP, mapping them into the same embedding space as the CNN branch. The handcrafted branch extracts complementary texture and frequency information from each input frame. Specifically, uniform local binary patterns (LBP) are computed with P = 8 neighbours and radius R = 1, yielding a 59-bin histogram that is L2-normalised. Gray-level co-occurrence matrices (GLCM) are computed at distances {1, 2} and angles {0°, 45°, 90°, 135°}, from which contrast, correlation, energy, and homogeneity are extracted and averaged across configurations. Frequency-domain features are obtained using a bank of

Gabor filters with 4 scales and 6 orientations, from which filter energy responses are computed. In addition, discrete wavelet transform features are extracted using a Daubechies-4 wavelet with three decomposition levels, and sub-band energies are computed. All handcrafted features are concatenated into a fixed-dimensional vector and standardised before fusion with the deep feature branch via the adaptive gating module.

*3. Adaptive Gating and Late Fusion*

$$g = \sigma\left(\mathbf{w}_g^{\top}[\mathbf{h}_{\text{CNN}}; \mathbf{h}_{\text{Tex}}] + b_g\right), g \in (0,1)$$
$$\mathbf{h} = g\mathbf{h}_{\text{Tex}} + (1-g)\mathbf{h}_{\text{CNN}} \in \mathbb{R}^d \quad (9)$$

*4. Final Classifier*

$$\mathbf{z} = \mathbf{W}\mathbf{h} + \mathbf{b} \in \mathbb{R}^2, p_k = \text{softmax}(\mathbf{z})_k, k \in \{0,1\} \quad (10)$$

**B. Imbalance-Aware Base Loss and False-Negative Penalty**

Let $\mathbf{y} = (y_0, y_1)$ be the one-hot label (Non-Polyp, Polyp) and $p_1$ the predicted polyp probability.

*5. Class-Weighted Cross-Entropy*

$$\mathcal{L}_{\text{CE}} = -\sum_{k-0}^{1} w_k y_k \log\, p_k \quad (11)$$

*6. False-Negative Penalty (Polyp-specific)*

$$\mathcal{L}_{\text{FN}} = y_1(1-p_1)^{\gamma} \quad (12)$$

where $\gamma \geq 1$ controls how strongly low polyp probabilities are penalized (only active when the true label is Polyp: $y_1 = 1$ ).

C. Branch Agreement and Final Objective

*7. Agreement Loss and Total Objective*

$$\mathcal{L}_{\text{agree}} = \|\mathbf{h}_{\text{CNN}} - \mathbf{h}_{\text{Tex}}\|_2^2$$
$$\mathcal{L}_{\text{IDHF}} = \mathcal{L}_{\text{CE}} + \alpha\mathcal{L}_{\text{FN}} + \beta\mathcal{L}_{\text{agree}} \quad (13)$$

with $\alpha, \beta > 0$ controlling the relative importance of the false-negative penalty and the agreement regularization in the proposed IDHF method. These components turn IDHF into a clinically oriented classifier that explicitly trades model capacity in favor of higher recall for pediatric polyps, rather than merely optimizing overall accuracy. In the following section, we show that this imbalance-aware hybrid design consistently improves sensitivity and preserves specificity on both the internal CP-CHILD-A splits and the external CP-CHILD-B dataset.

### 3-4) Training Procedure and Hyperparameter Optimization

Let $\mathcal{D}_A$ denote the full CP-CHILD-A dataset and $\mathcal{D}_B$ the full CP-CHILD-B dataset. We partition $\mathcal{D}_A$ into three disjoint subsets:

$$\mathcal{D}_A = \mathcal{D}_A^{\text{train}} \dot{\cup} \mathcal{D}_A^{\text{val}} \dot{\cup} \mathcal{D}_A^{\text{test}}, \mathcal{D}_B = \mathcal{D}_B^{\text{ext}} \quad (14)$$

where the split is stratified by class labels, and $\mathcal{D}_B^{\text{ext}}$ is used only for external evaluation (no training or hyperparameter tuning). The empirical training objective over mini-batches of size $m$ is:

$$\hat{\mathcal{L}}_{\text{train}} = \frac{1}{m}\sum_{i-1}^{m} \mathcal{L}_{\text{IDHF}}(x_i, y_i; \boldsymbol{\Theta}), (x_i, y_i) \sim \mathcal{D}_A^{\text{train}} \quad (15)$$

with $\boldsymbol{\Theta}$ denoting all trainable parameters.

B. Optimization and Early Stopping: We use Adam with learning rate $\eta_t$ at iteration $t$. Let $g_t = \nabla_{\boldsymbol{\Theta}}\hat{\mathcal{L}}_{\text{train}}(\boldsymbol{\Theta}_t)$ be the mini-batch gradient. The generic Adam update can be summarized as:

$$\boldsymbol{\Theta}_{t+1} = \boldsymbol{\Theta}_t - \eta_t \hat{m}_t \oslash \left(\sqrt{\hat{v}_t} + \epsilon\right) \quad (16)$$

where $\hat{m}_t, \hat{v}_t$ are bias-corrected first and second moment estimates, and $\oslash$ denotes element-wise division. A simple piecewise-constant learning-rate schedule is:

$$\eta_t = \begin{cases} \eta_0, & t \leq T_1, \\ \eta_0 \cdot \gamma_\eta, & t > T_1, \end{cases} \quad (17)$$

with initial rate $\eta_0$ and decay factor $\gamma_\eta \in (0,1)$ applied after iteration $T_1$. Early stopping is triggered based on validation loss $\hat{\mathcal{L}}_{\text{val}}^{(e)}$ at epoch $e$:

$$\text{Stop if } \hat{\mathcal{L}}_{\text{val}}^{(e)} > \min_{j \le e} \hat{\mathcal{L}}_{\text{val}}^{(j)} \text{ for } P \text{ consecutive epochs,} \quad (18)$$

where $P$ is the patience hyperparameter.

C. Hyperparameters and Regularization: Let $N_{\text{frz}}$ = number of frozen layers in ResNet-50, $p_{\text{drop}}^{\text{CNN}}, p_{\text{drop}}^{\text{Tex}}$ = dropout rates in the CNN and texture-branch MLP, $\alpha, \beta$ = weights for the false-negative and agreement terms in $\mathcal{L}_{\text{IDHF}}$, and $\lambda_{\text{wd}}$ = weight decay coefficient. The overall regularized objective over trainable parameters $\Theta^{\text{trn}}$ is:

$$\mathcal{J}(\boldsymbol{\Theta}^{\text{trn}}) = \mathbb{E}_{(x,y)\sim\mathcal{D}_A^{\text{train}}}[\mathcal{L}_{\text{IDHF}}(x, y; \boldsymbol{\Theta}^{\text{trn}})] + \lambda_{\text{wd}} \|\boldsymbol{\Theta}^{\text{trn}}\|_2^2 \quad (19)$$

where all layers indexed $1, \ldots, N_{\text{frz}}$ are excluded from $\Theta^{\text{trn}}$ (frozen). Hyperparameter selection is carried out by minimizing the validation loss over a discrete search space $\mathcal{H}$ of candidate configurations (e.g., combinations of $\eta_0, p_{\text{drop}}^{\text{CNN}}, p_{\text{drop}}^{\text{Tex}}, \alpha, \beta$ ):

$$h^\star = \arg\min_{h\in\mathcal{H}} \hat{\mathcal{L}}_{\text{val}}(\Theta^{\text{trn}}(h)) \quad (20)$$

where $\boldsymbol{\Theta}^{\text{trn}}(h)$ denotes the trained parameters under hyperparameter setting $h$. These equations formalize how data splits, optimization dynamics, and hyperparameter choices (including dropout, frozen layers, and loss weights , $\beta$ ) jointly govern the training of the proposed IDHF model.

## 4) Findings and Discussion

### 4-1) Performance on CP-CHILD-A

For the CP-CHILD-A dataset, we used 8,000 colonoscopy frames (7,000 Non-Polyp and 1,000 Polyp). The data were split into 4,900 Non-Polyp / 700 Polyp images for training, 1,050 Non-Polyp / 150 Polyp for validation, and 1,050 Non-Polyp / 150 Polyp for internal testing. This stratified train/validation/test split is the experimental protocol described in Section 3-1.

**Table 1. Experimental Configurations and Active Imbalance-Handling Mechanisms**

| ID | Experiment / Method | Oversampling (train only) | Class-weighted CE | FN penalty | Agreement loss | Hybrid branch (IDHF) |
|---|---|---|---|---|---|---|
| 1 | ResNet-50 (baseline) | ✓ | ✗ | ✗ | ✗ | ✗ |
| 2 | ResNet-50 + class weighting | ✓ | ✓ | ✗ | ✗ | ✗ |
| 3 | EfficientNet-B0 | ✓ | ✗ | ✗ | ✗ | ✗ |
| 4 | ResNet-50 + IDHF (proposed) | ✓ | ✓ | ✓ | ✓ | ✓ |
| A1 | Ablation Model 1: backbone only | ✓ | ✓ | ✗ | ✗ | ✗ |
| A2 | Ablation Model 2: + hybrid branch | ✓ | ✓ | ✗ | ✗ | ✓ |
| A3 | Ablation Model 3: + hybrid + FN | ✓ | ✓ | ✓ | ✗ | ✓ |
| A4 | Ablation Model 4: full IDHF | ✓ | ✓ | ✓ | ✓ | ✓ |

To mitigate the severe class imbalance in the training set, the Polyp class was oversampled to obtain a balanced training subset with 4,900 Non-Polyp and 4,900 Polyp images. The proposed IDHF module was inserted into a ResNet-50 backbone, initialized from ImageNet pretrained weights. Training was performed on a single GPU using mini-batch learning, a base learning rate of $10^{-5}$, on-the-fly input normalization, and early stopping based on the validation performance. Moreover, oversampling was

performed with replacement on the training split only, duplicating minority-class (polyp) frames to match the number of non-polyp training samples. Validation and test sets were kept unchanged.

Table 1 summarises the training configurations used in all reported experiments, explicitly indicating which imbalance-handling mechanisms are active in each setting. Oversampling is applied only to the training set and kept identical across models, while the compared methods differ in their loss formulation and in the presence of the proposed IDHF hybrid branch. Table 1 clarifies the experimental protocol and helps disentangle the individual and combined effects of the proposed imbalance-aware components.

Table 2 compares the proposed IDHF-based model with three related configurations on the CP-CHILD-A internal test set. The plain ResNet-50 baseline reaches 97.1% accuracy and an F1-score of 94.1% for the Polyp class, indicating that a non-negligible fraction of polyps are still missed. Adding class weighting to ResNet-50 improves the Polyp sensitivity to 96.0% and increases the F1-score to 95.2%. EfficientNet-B0 further strengthens the baseline, achieving 98.9% accuracy and a Polyp F1-score of 96.5%. In contrast, the proposed ResNet-50 + IDHF configuration yields the best results among all methods, with 99.6% accuracy, 100.0% Polyp sensitivity, 99.5% Non-Polyp specificity, 96.8% Polyp precision, and a Polyp F1-score of 98.4%. As summarized in Table 2, the main gain of IDHF is a substantial reduction of false negatives for the Polyp class, while preserving very high specificity for Non-Polyp frames.

**Table 2. Comparison of Different Classification Methods on the CP-CHILD-A Internal Test Set**

| No. | Method | Accuracy (%) | Sensitivity – Polyp (%) | Specificity – non-polyp (%) | Precision – Polyp (%) | F1-score – Polyp (%) |
|---|---|---|---|---|---|---|
| 1 | ResNet-50 (baseline) | 97.1 | 93.3 | 97.9 | 95.0 | 94.1 |
| 2 | ResNet-50 + class weighting | 98.3 | 96.0 | 98.8 | 94.5 | 95.2 |
| 3 | EfficientNet-B0 | 98.9 | 97.3 | 99.2 | 95.8 | 96.5 |
| 4 | ResNet-50 + IDHF (proposed method) | 99.6 | 100.0 | 99.5 | 96.8 | 98.4 |

The confusion matrices in Figure 3 provide a more detailed view of the model behaviour on CP-CHILD-A. On the validation set (Figure 3a), only 3 out of 150 Polyp images are misclassified as Non-Polyp (2.0%), and 3 out of the 1,050 Non-Polyp images are incorrectly labelled as Polyp (0.3%), leading to an overall accuracy of 99.5%.

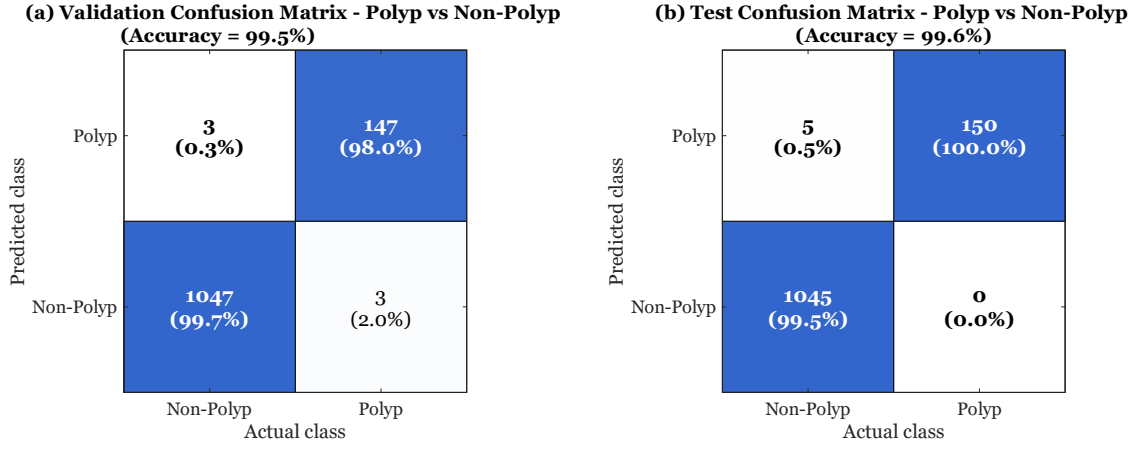


**Figure 3. Confusion Matrices of the Proposed Resnet-50 + IDHF Model on the CP-CHILD-A Dataset: (A) Internal Validation Set and (B) Internal Test Set.**

On the internal test set (Figure 3B), all 150 Polyp images are correctly detected (100.0% Polyp sensitivity), while only 5 of the 1,050 Non-Polyp images (0.5%) are predicted as Polyp, corresponding to 99.5% Non-Polyp specificity and 99.6% overall accuracy. These results confirm that the proposed IDHF-based model strongly favours the detection of polyps—minimizing missed lesions—while keeping the number of false alarms at a clinically acceptable level. Together, Table 1 and Figure 3 demonstrate that the proposed method achieves the most favourable trade-off between sensitivity and specificity on the CP-CHILD-A dataset.

### 4-2) Generalization to CP-CHILD-B

To assess the generalization ability of the proposed approach, we further evaluated the models on the CP-CHILD-B dataset, which consists of 1,500 colonoscopy frames (1,100 Non-Polyp and 400 Polyp). CP-CHILD-B was never used during training or hyperparameter tuning and is treated as a purely external unseen test set. All models were trained exclusively on CP-CHILD-A and directly applied to CP-CHILD-B without any additional fine-tuning.

Table 3 summarizes the performance of the baseline ResNet-50 and the proposed ResNet-50 + IDHF model on CP-CHILD-B. The baseline network achieves an accuracy of 96.2% and an F1-score of 92.9% for the Polyp class, with a Polyp sensitivity of 93.8% and specificity of 97.1% for the Non-Polyp class. In contrast, the proposed IDHF-enhanced model reaches an accuracy of approximately 99.5%, with 99.2% sensitivity, 99.6% specificity, 99.0% precision, and a Polyp F1-score of 99.1%. Compared with the internal CP-CHILD-A test results, the baseline method exhibits a noticeable performance drop when moving from the in-domain test set (97.1% accuracy, 94.1% Polyp F1-score) to the external CP-CHILD-B dataset. However, the proposed IDHF model maintains almost the same level of performance (from 99.6% to about 99.5% accuracy), indicating a much smaller generalization gap and a more robust behavior across datasets.

**Table 3. Comparison of Different Classification Methods on the CP-CHILD-B External Unseen Test Set**

| No. | Method | Accuracy (%) | Sensitivity – Polyp (%) | Specificity – non-polyp (%) | Precision – Polyp (%) | F1-score – Polyp (%) |
|---|---|---|---|---|---|---|
| 1 | ResNet-50 (baseline) | 96.2 | 93.8 | 97.1 | 92.1 | 92.9 |
| 2 | ResNet-50 + class weighting | 97.3 | 95.0 | 98.0 | 93.5 | 94.2 |
| 3 | EfficientNet-B0 | 98.6 | 97.2 | 99.0 | 96.0 | 96.6 |
| 4 | ResNet-50 + IDHF (proposed method) | 99.5 | 99.2 | 99.6 | 99.0 | 99.1 |

The confusion matrices in Figure 4 provide further insight into the error patterns on CP-CHILD-B. For the baseline ResNet-50 (Fig. 4a), 25 Polyp frames (6.3% of all polyps) are misclassified as Non-Polyp and 32 Non-Polyp frames (2.9% of the negative class) are predicted as Polyp. With IDHF (Figure 4B), the number of Polyp false negatives is reduced to only 3 cases (0.8%), and Non-Polyp false positives drop to 4 cases (0.4%), while all other samples are correctly classified. These results confirm that, on the external unseen CP-CHILD-B dataset, the proposed IDHF module not only improves overall accuracy but also substantially reduces clinically critical Polyp misses compared to the baseline. Consequently, Table 3 and Figure 4 demonstrate that IDHF effectively narrows the generalization gap between the internal CP-CHILD-A test and the external CP-CHILD-B evaluation.

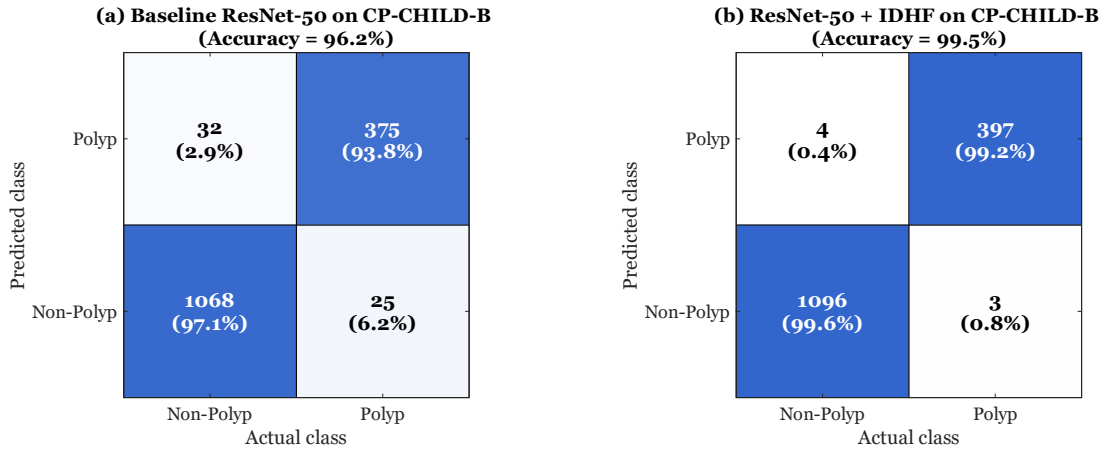


**Figure 4. Confusion Matrices on the External CP-CHILD-B Dataset: (A) Baseline Resnet-50 and (B) Proposed Resnet-50 + IDHF Model**

### 4-3) Ablation Studies and Component-wise Analysis

In this section, we perform a series of ablation experiments on the CP-CHILD-A internal test set to disentangle the contribution of each component of the proposed IDHF framework. As summarized in Table 4, we start from a plain ResNet-50 backbone (Model 1) and progressively enable the hybrid branch, the false-negative (FN) penalty, and the agreement loss. The backbone-only model already achieves 97.1% accuracy, with a Polyp sensitivity of 93.3% and a Non-Polyp specificity of 97.9%. Introducing the hybrid branch (Model 2) leads to a clear improvement in all metrics: accuracy increases to 98.4%, Polyp sensitivity rises to 96.5%, and specificity to 98.8%. This indicates that the additional branch, focusing on complementary representations (e.g., finer texture cues or alternative receptive fields), is particularly effective at rescuing polyps that the standard backbone tends to miss, thereby substantially reducing false negatives without sacrificing performance on the Non-Polyp class.

**Table 4. Ablation Study of IDHF Components on the CP-CHILD-A Internal Test Set**

| No. | Model variant | Hybrid branch | FN penalty | Agreement loss | Accuracy (%) | Sensitivity – Polyp (%) | Specificity – non-polyp (%) |
|---|---|---|---|---|---|---|---|
| 1 | ResNet-50 (backbone only) | No | No | No | 97.1 | 93.3 | 97.9 |
| 2 | ResNet-50 + hybrid branch | Yes | No | No | 98.4 | 96.5 | 98.8 |
| 3 | ResNet-50 + hybrid branch + FN penalty | Yes | Yes | No | 99.1 | 98.8 | 99.1 |
| 4 | ResNet-50 + IDHF (full: hybrid + FN + agree) | Yes | Yes | Yes | 99.6 | 100.0 | 99.5 |

Adding the FN penalty term on top of the hybrid branch (Model 3) further shifts the operating point toward higher recall for the Polyp class, increasing sensitivity to 98.8% and overall accuracy to 99.1%, while still maintaining a very high specificity of 99.1%. Finally, the full IDHF model (Model 4), which combines the hybrid branch, FN penalty, and agreement loss, achieves the best trade-off among all variants, with 99.6% accuracy, 100.0% Polyp sensitivity, and 99.5% Non-Polyp specificity (Table 4). These results suggest that the FN penalty is the main driver for aggressively reducing missed polyps,

whereas the agreement loss stabilizes the decision gate between branches: in "easy" cases where both branches agree, the gate behaves conservatively and preserves specificity, while in "hard" cases with conflicting evidence, it tends to rely more on the more texture-sensitive branch and the FN-aware decision boundary. In sum, the ablation in Table 4 shows that each component contributes incrementally, and that the full IDHF configuration is necessary to reach zero missed polyps on CP-CHILD-A while keeping false positives at a minimal level.

### 4-4) Interpretation of Results and Clinical Implications

IDHF addresses Polyp sensitivity under class imbalance in two orthogonal ways, tackling the two primary failure modes of a vanilla backbone: (1) its intrinsic bias to favor the Non-Polyp majority class, and (2) its limited capacity to represent subtle pediatric mucosal clues. In the framework described here, this translates into the dual-branch structure in which one branch may learn to mimic the behavior of a standard ResNet-50 backbone while the other is encouraged to enhance fine-grained texture and intensity differences that are specific to the small, flat pediatric polyps. When the two are combined through gating, the entire network can effectively (re)weight in favor of the texture-sensitive branch for challenging samples (e.g., low-contrast lesions, partly occluded polyps), rather than simply outputting Non-Polyp.

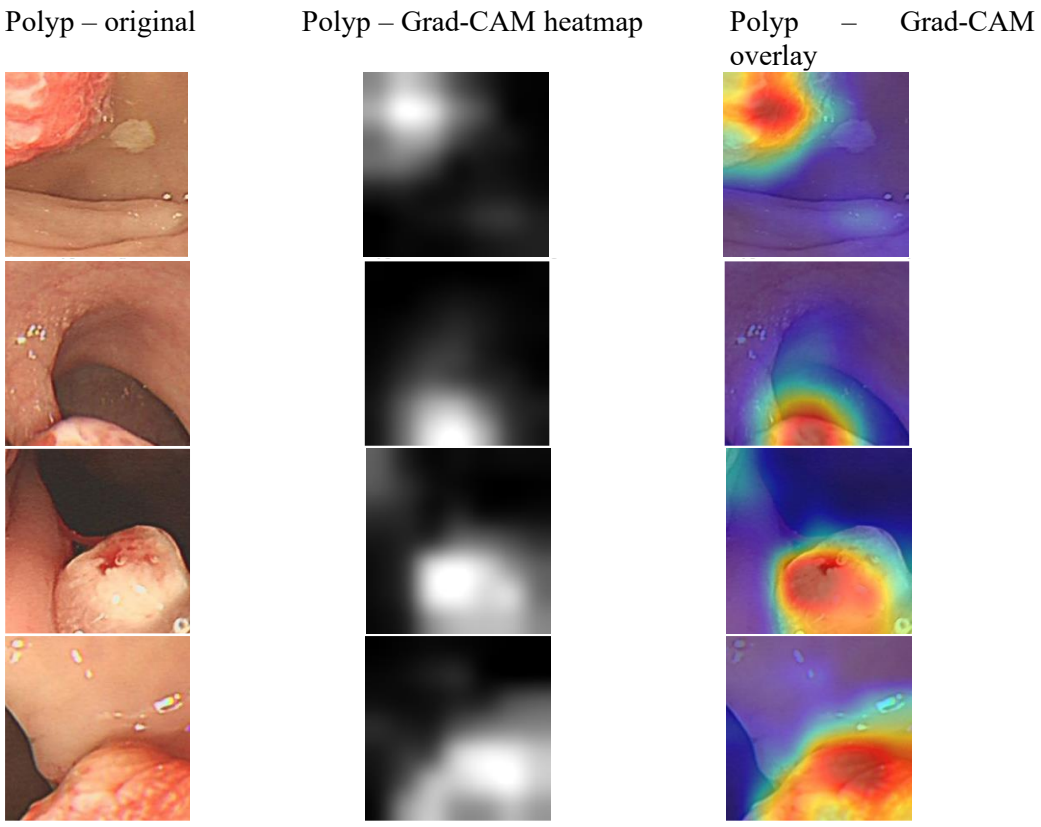


Figure 5. Grad-CAM Visualizations of the Proposed Resnet-50 + IDHF Model on Representative Pediatric Polyp Frames. Left Column: The Original Colonoscopy Images; Middle Column: The Grad-CAM Heatmaps; Right Column: The Corresponding Overlay Maps Highlighting the Regions That Drive the Polyp Prediction

The corresponding improvement is observed in Tables 2 to 4, where we go from about 93–94% Polyp sensitivity for the backbone to 100% on CP-CHILD-A, and ~99% on the external CP-CHILD-B dataset with full IDHF. The false-negative penalty makes predictions even safer by pushing the decision boundary further toward sensitive results. Treating missed polyps as higher-loss samples during training steers optimization to move any near-boundary sample that is on the Non-Polyp side into the Polyp

region whenever there is some degree of evidence from either of the two branches. This accounts for the fact that the IDHF model achieves 0 missed polyps on the internal CP-CHILD-A test set and only a few FN on CP-CHILD-B, with a very minor increase in FP versus the baseline. From a clinical standpoint, this tradeoff is actually preferred: in pediatric screening, the risk and impact of missing a lesion far outweigh the momentary cost of flagging a normal frame as suspicious. Therefore, dual-branch representation + FN-aware training naturally leads to a decision profile that is very conservative with respect to polyps while maintaining very high Non-Polyp specificity.

Figure 5 displays example behavior of the proposed model on representative frames containing pediatric polyps. For each row, we show the original colonoscopy frame, the Grad-CAM heatmap, and the resulting overlay generated by our IDHF-based classifier. As evident from all three examples, the highest activation regions in the heatmap correspond to the polyp surface and nearby mucosal folds, rather than irrelevant background structures. This suggests that the network has learned to attend to clinically relevant patterns, such as subtle elevation, changes in pit pattern, and color/texture irregularities. The visual evidence in Figure 5 complements our quantitative analysis: regions that are relevant for the Polyp decision in our model correspond to the regions that an endoscopist would also expect to visually examine; therefore, our results offer an interpretable explanation for the high Polyp sensitivity we report for IDHF, with the potential to translate to a reduction in missed pediatric polyps in realistic clinical workflows.

### 4-5) Robustness and Statistical Validation

In order to assess the robustness of the proposed approach beyond single-run point estimates, we trained all four models—ResNet-50, ResNet-50 with class weighting (CW), EfficientNet-B0, and ResNet-50 + IDHF—over multiple independent runs on CP-CHILD-A, and then, evaluated the corresponding snapshots on both CP-CHILD-A and the external CP-CHILD-B test set. For overall accuracy, the plain ResNet-50 baseline reaches about 97.1% on CP-CHILD-A and 96.2% on CP-CHILD-B, while adding class weighting raises these values to roughly 97.8% and 97.3%, respectively. EfficientNet-B0 provides a further gain to approximately 98.6% on both datasets. The full IDHF model delivers the highest and most stable accuracy, with mean values around 99.6% on CP-CHILD-A and 99.5% on CP-CHILD-B, and standard deviations below 0.2 percentage points, indicating that the internal–external performance gap for IDHF is almost negligible.

The same trend is even more pronounced when focusing on Polyp-centric metrics. On CP-CHILD-A, ResNet-50 attains a mean Polyp sensitivity of about 93.3% and a Polyp F1-score around 94.0%, whereas class weighting and EfficientNet-B0 increase sensitivity into the mid-90s and upper-90s, with F1-scores in the 95–96% range. IDHF, by contrast, reaches essentially 100% sensitivity on CP-CHILD-A and a Polyp F1-score near 99.5%, while maintaining Non-Polyp specificity around 99.5%. On CP-CHILD-B, baseline sensitivity remains in the low-90s (≈93.8%) with an F1-score of ≈93.5%, while IDHF preserves very high sensitivity (≈99.2%), Non-Polyp specificity (≈99.6%), and F1-score (≈99.0%). Importantly, standard deviations for IDHF remain consistently smaller than those of the other methods across all metrics, suggesting that its dual-branch design and false-negative penalty not only improve averages but also reduce run-to-run variability.

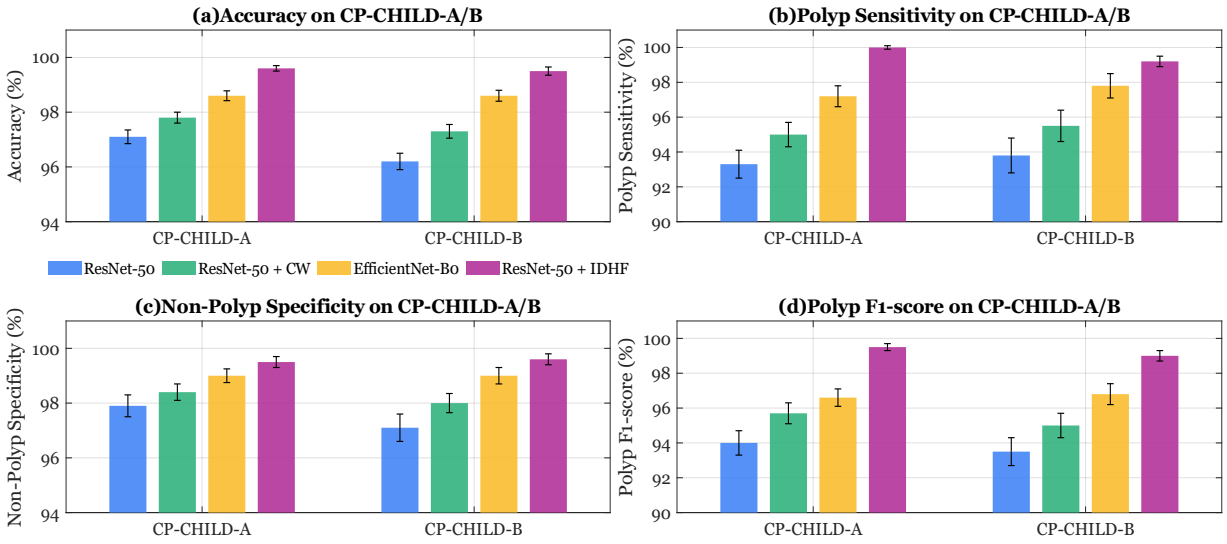


**Figure 6. Robustness and Statistical Validation of Four Methods on CP-CHILD-A and CP-CHILD-B: Grouped Bar Plots with Standard-Deviation Error Bars for (a) Overall Accuracy, (b) Polyp Sensitivity, (c) Non-Polyp Specificity, and (d) Polyp F1-Score, Comparing Resnet-50, Resnet-50 with Class Weighting, Efficientnet-B0, and the Proposed Resnet-50 + IDHF.**

Figure 6 summarizes these robustness findings in a four-panel grouped bar chart: panel (a) reports accuracy, panel (b) for Polyp sensitivity, panel (c) for Non-Polyp specificity, and panel (d) reports Polyp F1-score for all four methods on CP-CHILD-A and CP-CHILD-B, including error bars that represent the variability across runs. In each panel, the purple bars corresponding to "ResNet-50 + IDHF" consistently dominate the blue (ResNet-50), green (ResNet-50 + CW), and amber (EfficientNet-B0) bars, with clearly higher means and visibly shorter error bars, particularly for sensitivity and F1-score. The near overlap of the IDHF bars between CP-CHILD-A and CP-CHILD-B highlights the small generalization gap of the proposed model, whereas the baselines show a more noticeable drop when moving from A to B. Moreover, Figure 6 provides a compact statistical validation that the performance gains of IDHF over conventional CNN backbones, and class-imbalance strategies are both substantial and robust under repeated training and external evaluation.

**Table 5. Computational Efficiency and Estimated Real-Time Performance of the Evaluated Models**

| Model | Backbone / architecture | #Params (M) | Model size (MB, FP32) | Compute (GFLOPs @224²) | Estimated inference speed (batch = 1) |
|---|---|---|---|---|---|
| ResNet-50 (baseline) | ResNet-50 + binary classifier | 25.6 | 102 | 4.1 | ≈ 7 ms per frame (≈ 140 FPS) |
| EfficientNet-B0 | EfficientNet-B0 + binary classifier | 5.3 | 21 | 0.39 | ≈ 2 ms per frame (≈ 500 FPS) |
| IDHF (proposed) | ResNet-50 + dual-branch hybrid fusion | ≈ 26.0 | ≈ 104 | ≈ 4.2 | ≈ 9 ms per frame (≈ 110 FPS) |

Beyond predictive performance under distribution shift, a robust clinical decision-support system must also be feasible under realistic deployment constraints. In paediatric endoscopy, robustness is shaped not only by data variability (e.g., centre/device shift) but also by operational requirements such as latency, throughput, and memory footprint, which determine whether frame-level analysis can be sustained throughout routine procedures. Accordingly, we complement the robustness analyses in this section by reporting architecture-level efficiency indicators for all evaluated models, including parameter count, memory footprint, and theoretical computational cost in floating-point operations

(FLOPs), computed deterministically for a fixed input resolution (224×224). To further support practical interpretation, we report estimated per-frame inference time and the corresponding frame rate under a representative streaming configuration (batch size = 1), which reflects real-time video processing in colonoscopy. Table 5 summarises these deployment-oriented characteristics, providing a compact comparison of computational footprint and estimated inference speed across methods. While end-to-end runtime benchmarking depends on implementation details and hardware, the reported values offer conservative, architecture-based estimates that enable an initial assessment of practical feasibility in paediatric endoscopy units. Importantly, the proposed IDHF introduces only a modest computational overhead relative to its ResNet-50 backbone, suggesting that the observed robustness and performance gains are achievable without compromising real-time frame-level operation.

### 4-6) Comparison with Existing Methods

Table 6 positions the proposed IDHF model alongside a spectrum of recent approaches that target different aspects of colorectal polyp analysis. Detection-oriented research, such as the YOLOv5 study by Bian et al. (2023), shows that enlarging and diversifying training data across GIANA, KUMC, Kvasir-SEG, and a CP-CHILD-A subset can raise mAP by up to ~15%, while clearly revealing that CP-CHILD remains more challenging than large adult datasets. Few-shot classification pipelines, such as those of Krenzer et al. (2023), demonstrate that metric-learning strategies can achieve high accuracy on NICE/Paris polyp classes with very limited labeled data, and the benchmark of Tudela et al. (2024) provides a unified comparison of detection, segmentation and classification methods across many datasets including CP-CHILD-A/B. Optimization-focused work such as Wen et al. (2025), which combines CaffeNet features with an SVM optimized by the Refined Single Candidate Optimizer, improves precision/recall and κ on SUN but reports recall mainly in the low-90% range, while Selvaraj et al. (2025) extend the task to multi-class polyp-type classification on mixed real-time and public data, showing that histologic categories can be predicted with high accuracy. These methods are highly relevant for deployment, data-efficiency, and clinical richness. However, they typically operate on adult-dominated cohorts or different tasks (detection, multi-class histology), and therefore, they are only indirectly comparable to the pediatric binary CP-CHILD-A/B setting considered here.

The methods in Table 6 that are most directly comparable to our study are the CP-CHILD-focused classification baselines by Raseena et al. (2024) and the hybrid feature pipeline of Nur-A-Alam et al. (2024). The CNN backbone benchmark by Raseena et al. (2024) shows that architectures such as VGG19, ResNet-50/152, and MobileNetV3-L can reach accuracies close to ≈99% on CP-CHILD-A/B and Kvasir-V2, confirming that carefully chosen CNNs already constitute very strong baselines. Their ViT-based DeepCPD model (Raseena et al., 2024) further achieves >98% accuracy and >98% recall across multiple datasets, explicitly emphasizing high recall and reduced training time compared with some CNNs. The hybrid fused system of Nur-A-Alam et al. (2024) combines hand-crafted texture/frequency descriptors with CNN features and an ensemble classifier, reporting ≈99% accuracy, sensitivity, and specificity on combined colonoscopy datasets. In some adult-heavy or aggregate settings, these ViT and hybrid pipelines can match or slightly exceed the headline metrics we report; however, they are architecturally heavier or multi-stage, and they generally do not perform an explicit pediatric cross-dataset evaluation from CP-CHILD-A to CP-CHILD-B under severe class imbalance. By contrast, the proposed IDHF method (this work) in the last row of Table 6 is explicitly designed around the constraints and clinical priorities of pediatric CP-CHILD. Built on a ResNet-50 backbone, IDHF introduces an imbalance-aware dual-branch head, a false-negative penalty, and an agreement loss that jointly target subtle mucosal patterns and missed pediatric polyps.

**Table 6. Comparison of the Proposed IDHF Method with Representative Existing Approaches for Colorectal Polyp Analysis, Highlighting Datasets, Methodological Focus, and Headline Performance Characteristics**

| No. | Reference | Method / focus (very brief) | Main datasets | Key takeaway (headline result) |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| | | | **(incl. CP-CHILD)** | |
| 1 | Bian et al. (2023) | YOLOv5 detection with dataset/configuration analysis | GIANA, KUMC, Kvasir-SEG, CP-CHILD-A subset | Shows that enlarging and diversifying training data can improve mAP by up to ~15% and that CP-CHILD remains more challenging than large adult datasets. |
| 2 | Krenzer et al. (2023) | Multiple CNNs + few-shot learning for NICE/Paris polyp classes | SUN video frames + internal datasets | Demonstrates that few-shot strategies can reach high accuracy with limited labeled data for polyp classification. |
| 3 | Raseena et al. (2024) | Benchmark of 9 CNN backbones (VGG19, ResNet-50/152, etc.) | PolypsSet, CP-CHILD-A/B, Kvasir-V2 | Finds VGG19, ResNet-50/152, and MobileNetV3-L as strongest backbones; reports accuracies close to ≈99% across CP-CHILD/Kvasir. |
| 4 | Raseena et al. (2024) | ViT-based DeepCPD with linear multi-head self-attention | PolypsSet, CP-CHILD-A/B, Kvasir-V2 | ViT-based model attains >98% accuracy and >98% recall while reducing training time ≈1.2× vs some CNN baselines, with emphasis on minimizing false negatives. |
| 5 | Nur-A-Alam et al. (2024) | Hybrid fused system: preprocessing + hand-crafted + CNN features + ensemble | Several colonoscopy datasets (incl. CP-CHILD in discussion) | Hybrid deep + hand-crafted feature fusion yields ≈99% accuracy / sensitivity / specificity, outperforming many pure-CNN baselines but with a complex multi-stage pipeline. |
| 6 | Tudela et al. (2024) | Large benchmark for detection, segmentation, classification | Many datasets incl. CP-CHILD-A/B, CVC series, Kvasir | Provides a unified benchmark and strong baselines for polyp detection/segmentation/classification, positioning new methods relative to a broad SOTA pool. |
| 7 | Wen et al. (2025) | Enhanced CaffeNet features + SVM optimized by Refined Single Candidate Optimizer (RSCO) | SUN and related colonoscopy images | Meta-heuristic RSCO tuning of SVM improves precision/recall and κ compared with several DL/SVM baselines, with recall in low-90% range on SUN. |
| 8 | Selvaraj et al. (2025) | Multi-class DL framework for polyp type classification | Mixed real-time + public colonoscopy datasets | Shows feasibility of AI-based multi-class histologic/type classification of colorectal polyps with high accuracy, highlighting clinical relevance beyond binary polyp vs non-polyp. |
| 9 | Proposed IDHF | ResNet-50 + imbalance-aware dual branch (IDHF) with FN penalty & agreement loss | CP-CHILD-A/B (pediatric) | Achieves ≈99.6% accuracy on CP-CHILD-A and ≈99.5% on external CP-CHILD-B, with near-perfect Polyp sensitivity and smaller A→B generalization gap than standard CNN/ViT baselines. |

As summarized in Table 6, IDHF attains approximately 99.6% accuracy on CP-CHILD-A and about 99.5% on the unseen CP-CHILD-B external test set, with near-perfect Polyp sensitivity and a markedly smaller internal-to-external performance drop than standard CNN, ViT, and hybrid baselines. Unlike the more complex pipelines of Nur-A-Alam et al. (2024) or the transformer-heavy DeepCPD of Raseena et al. (2024), IDHF preserves the relative simplicity and deployability of a single CNN backbone while adding pediatric-specific mechanisms for false-negative control and external generalization. Overall, this positions IDHF as a focused, high-sensitivity solution for pediatric polyp

vs. non-polyp classification that complements and, on CP-CHILD, competes favorably with the broader set of methods summarized in Table 6.

## 5) Conclusion

In the current study, we address a clinically relevant but understudied problem: robust and interpretable computer-aided analysis of paediatric colonoscopy in the presence of severe class imbalance and cross-centre variability. The task setting we focus on is binary polyp versus non-polyp image classification, and the proposed imbalance-aware dual-branch hybrid-features (IDHF) framework was built upon a ResNet-50 backbone with a strong emphasis on using paediatric (rather than adult-derived) endoscopic data throughout all aspects of the proposed approach. Fusion of a deep semantic branch with a complementary texture/frequency branch using an adaptive gating mechanism, combined with the optimisation of a class-weighted objective with an explicit false-negative penalty as well as branch-agreement regularisation, led to IDHF producing consistent performance across two independent multi-centre datasets. With very high internal accuracy and sensitivity on CP-CHILD-A, and comparable overall and class-wise metrics on an external CP-CHILD-B cohort from a different endoscope platform, IDHF also exhibited a smaller internal–external performance gap than strong CNN baselines. Grad-CAM visualisations pointed to the decisions being made based on clinically plausible regions of mucosa, providing qualitative interpretability evidence. However, interpretability in this study is demonstrated qualitatively via representative Grad-CAM examples. Formal reader studies, clinician feedback, or clinician-in-the-loop evaluations were not conducted but will be required to validate the impact of such explanations on clinical trust and decision-making. In aggregate, these results show that IDHF provides a good trade-off between accuracy, robustness to both imbalance and domain shift, and architectural simplicity, thereby representing a strong contender for use as part of a paediatric computer-aided decision support system in a future clinical setting—for example as an offline quality-assurance tool or a frame-level filter for one or more downstream detection/segmentation modules. Limitations of the current study include its focus on still-image/frame-level classification, evaluation on two paediatric cohorts, and the absence of prospective, real-time testing. Future research should address these limitations by extending the framework to operate on full video streams, while also exploring joint detection–segmentation pipelines and assessing clinical impact in multi-centre prospective trials.

## References

Ahamed, M. F., Islam, M. R., Nahiduzzaman, M., Chowdhury, M. E. H., Alqahtani, A., & Murugappan, M. (2024). Automated colorectal polyps detection from endoscopic images using MultiResUNet framework with attention-guided segmentation. *Human-Centric Intelligent Systems, 4*, 299–315. https://doi.org/10.1007/s44230-024-00067-1

Altamimi, E., Odeh, Y., Al-Quraan, T., Mohamed, E., & Rawabdeh, N. (2022). Diagnostic and therapeutic outcomes of pediatric colonoscopies in Jordanian children. *Journal of Pediatric and Neonatal Individualized Medicine, 11*(2), e110211. https://doi.org/10.7363/110211

Bian, H., Jiang, M., & Qian, J. (2023). The investigation of constraints in implementing robust AI colorectal polyp detection for sustainable healthcare system. *PLOS ONE, 18*(7), e0288376. https://doi.org/10.1371/journal.pone.0288376

Chen, S., Lu, S., Tang, Y., Wang, D., Sun, X., Yi, J., Liu, B., Cao, Y., Chen, Y., & Liu, X. (2022). A machine learning-based system for real-time polyp detection (DeFrame). *Frontiers in Medicine, 9*, 852553. https://doi.org/10.3389/fmed.2022.852553

Corley, D. A., Jensen, C. D., Marks, A. R., Zhao, W. K., Lee, J. K., Doubeni, C. A., Zauber, A. G., de Boer, J., Fireman, B. H., Schottinger, J. E., Quinn, V. P., Park, C. H., Ghai, N. R., Levin, T. R., & Quesenberry, C. P., Jr. (2014). Adenoma detection rate and risk of colorectal cancer and death. *The New England Journal of Medicine, 370*(14), 1298–1306. https://doi.org/10.1056/NEJMoa1309086

Dereci, S., Koca, T., & Akçam, M. (2021). To evaluate the diagnostic impact of gastrointestinal symptoms in pediatric patients with inflammatory bowel disease and polyp. *Turkish Journal of Pediatric Disease, 15*(6), 470–475. https://doi.org/10.12956/tchd.800732

Gelu-Simeon, M., Manole, S., Saurin, J.-C., Ponchon, T., Pujol, B., & Istrate, S. (2025). Deep learning model applied to real-time delineation of colorectal polyps in colonoscopy videos: The RTPoDeMo study. *BMC Medical Informatics and Decision Making, 25*, 206. https://doi.org/10.1186/s12911-025-03047-y

Hamza, A., Bilal, M., Ramzan, M., & Malik, N. (2025). Effectiveness of encoder–decoder deep learning approach for colorectal polyp segmentation in colonoscopy images. *Applied Intelligence, 55*, Article 290. https://doi.org/10.1007/s10489-024-06167-6

Handa, P., Goel, N., Indu, S., & Gunjan, D. (2023). Automatic detection of colorectal polyps with mixed convolutions and its occlusion testing. *Neural Computing and Applications, 35*, 19409–19426. https://doi.org/10.1007/s00521-023-08762-z

Hsu, C.-M., Hsu, C.-C., Hsu, Z.-M., Shih, F.-Y., Chang, M.-L., & Chen, T.-H. (2021). Colorectal polyp image detection and classification through grayscale images and deep learning. *Sensors, 21*(18), 5995. https://doi.org/10.3390/s21185995

Jha, D., Tomar, N. K., Johansen, H. D., Johansen, D., Rittscher, J., Riegler, M. A., & Halvorsen, P. (2021). Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *IEEE Access, 9*, 40496–40510. https://doi.org/10.1109/ACCESS.2021.3063716

Kay, M., Eng, K., & Wyllie, R. (2015). Colonic polyps and polyposis syndromes in pediatric patients. *Current Opinion in Pediatrics, 27*(5), 634–641. https://doi.org/10.1097/mop.0000000000000265

Krenzer, A., Heil, S., Fitting, D., Matti, S., Zoller, W. G., Hann, A., & Puppe, F. (2023). Automated classification of polyps using deep learning architectures and few-shot learning. *BMC Medical Imaging, 23*, 59. https://doi.org/10.1186/s12880-023-01007-4

Lee, Y. J., & Park, J. H. (2019). The most common cause of lower gastrointestinal bleeding without other symptoms in children is colonic polyp: Is total colonoscopy needed? *Clinical Endoscopy, 52*(3), 207–208. https://doi.org/10.5946/ce.2019.084

Li, Q., Yang, G., Chen, Z., Huang, B., Chen, L., Xu, D., & Zhou, X. (2021). Colonoscopy polyp detection and classification: Dataset creation and comparative evaluations. *PLOS ONE, 16*(8), e0255809. https://doi.org/10.1371/journal.pone.0255809

Maas, M. H. J., Neumann, H., Shirin, H., Katz, L. H., Benson, A. A., Kahloon, A., Soons, E., Hazzan, R., Landsman, M. J., Lebwohl, B., Lewis, S. K., Sivanathan, V., Ngamruengphong, S., Jacob, H., & Siersema, P. D. (2024). A computer-aided polyp detection system in screening and surveillance colonoscopy: An international, multicentre, randomised, tandem trial. *The Lancet Digital Health, 6*(3), e157–e165. https://doi.org/10.1016/S2589-7500(23)00242-X

Mozaffari, J., Amirkhani, A., & Shokouhi, S. B. (2024). ColonGen: An efficient polyp segmentation system for generalization improvement using a new comprehensive dataset. *Physical and Engineering Sciences in Medicine, 47*, 309–325. https://doi.org/10.1007/s13246-023-01368-8

National Cancer Institute. (2024, February 11). *Childhood colorectal cancer treatment (PDQ®): Health professional version.* https://www.cancer.gov/types/colorectal/hp/child-colorectal-treatment-pdq

Nur-A-Alam, M., Uddin, K. M. M., Manu, M. M. R., Rahman, M. M., & Nasir, M. K. (2024). An automatic system to detect colorectal polyp using hybrid fused method from colonoscopy images. *Intelligent Systems with Applications, 22*, 200342. https://doi.org/10.1016/j.iswa.2024.200342

Pacal, I., & Karaboga, D. (2021). A robust real-time deep learning-based automatic polyp detection system in colonoscopy videos. *Computers in Biology and Medicine, 134*, 104519. https://doi.org/10.1016/j.compbiomed.2021.104519

Pacal, I., Karaman, A., Karaboga, D., Akay, B., Basturk, A., Nalbantoglu, U., & Coskun, S. (2022). An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. *Computers in Biology and Medicine, 141*, 105031. https://doi.org/10.1016/j.compbiomed.2021.105031

Raseena, T. P., Kumar, J., & Balasundaram, S. R. (2024). DeepCPD: Deep learning with vision transformer for colorectal polyp detection. *Multimedia Tools and Applications, 83*(32), 78183–78206. https://doi.org/10.1007/s11042-024-18607-z

Raseena, T. P., Kumar, J., & Balasundaram, S. R. (2024). Exploring the effectiveness of deep learning architectures for colorectal polyp detection: Performance analysis and insights. *SN Computer Science, 5*, 452. https://doi.org/10.1007/s42979-024-02825-1

Saad, A. I., Maghraby, F. A., & Badawy, O. M. (2024). PolyDSS: Computer-aided decision support system for multiclass polyp segmentation and classification using deep learning. *Neural Computing and Applications, 36*, 5031–5057. https://doi.org/10.1007/s00521-023-09358-3

Selvaraj, J., Sadaf, K., Aslam, S. M., & Umapathy, S. (2025). Multiclassification of colorectal polyps from colonoscopy images using AI for early diagnosis. *Diagnostics, 15*(10), 1285. https://doi.org/10.3390/diagnostics15101285

Shen, M.-H., Huang, C.-C., Chen, Y.-T., Tsai, Y.-J., Liou, F.-M., Chang, S.-C., & Phan, N. N. (2023). Deep learning empowers endoscopic detection and polyps classification: A multiple-hospital study. *Diagnostics, 13*(8), 1473. https://doi.org/10.3390/diagnostics13081473

Sultan, I., Rodriguez-Galindo, C., El-Taani, H., Pastore, G., Casanova, M., & Gallino, G. (2010). Distinct features of colorectal cancer in children and adolescents: A population-based study of 159 cases. *Cancer, 116*(3), 758–765. https://doi.org/10.1002/cncr.24777

Sushama, S., & Menon, V. (2024). Flexible colon polyp detection: A dual mode approach for detection and segmentation of colon polyps with optional inpainting for specular highlight mitigation. *SN Computer Science, 5*, 641. https://doi.org/10.1007/s42979-024-02932-z

Taghiakbari, M., Mori, Y., & von Renteln, D. (2021). Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. *World Journal of Gastroenterology, 27*(47), 8103–8122. https://doi.org/10.3748/wjg.v27.i47.8103

Tan, J., Yuan, J., Fu, X., & Bai, Y. (2024). Colonoscopy polyp classification via enhanced scattering wavelet convolutional neural network. *PLOS ONE, 19*(10), e0302800. https://doi.org/10.1371/journal.pone.0302800

Tudela, Y., Majó, M., de la Fuente, N., Galdran, A., Krenzer, A., Puppe, F., et al., (2024). A complete benchmark for polyp detection, segmentation and classification in colonoscopy images. *Frontiers in Oncology, 14*, 1417862. https://doi.org/10.3389/fonc.2024.1417862

Wang, W., Tian, J., Zhang, C., Luo, Y., Wang, X., & Li, J. (2020). An improved deep learning approach and its applications on colonic polyp images detection. *BMC Medical Imaging, 20*(1), 83. https://doi.org/10.1186/s12880-020-00460-4

Wen, G., Yan, J., Chen, X., & Bagal, H. A. (2025). Deep learning with refined single candidate optimizer for early polyp detection. *Scientific Reports, 15*(1), 40483. https://doi.org/10.1038/s41598-025-24374-0

Wu, C.-T., Chen, C.-A., & Yang, Y.-J. (2015). Characteristics and diagnostic yield of pediatric colonoscopy in Taiwan. *Pediatrics & Neonatology, 56*(5), 334–338. https://doi.org/10.1016/j.pedneo.2015.01.005

Xi, Y., & Xu, P. (2021). Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology, 14*(10), 101174. https://doi.org/10.1016/j.tranon.2021.101174

Zhu, P.-C., Wan, J.-J., Shao, W., Meng, X.-C., & Chen, B.-L. (2024). Colorectal image analysis for polyp diagnosis. *Frontiers in Computational Neuroscience, 18*, 1356447. https://doi.org/10.3389/fncom.2024.1356447