Enhanced Gastrointestinal Disease Detection Through ResNet50 And Comparative Analysis Of Deep Learning Architectures For Medical Image Classification

Sharmilaa S[1], Udhayakumari S[2], Sona M[3], Vishnu Devi R[4]
[1, 2, 3, 4] Dept of Computer Science and Engineering
[1, 2, 3, 4] Vivekanandha College of Technology for Women

Abstract- *Gastrointestinal (GI) diseases such as polyps, ulcers, and tumors require early detection for effective treatment. Traditional manual inspection of endoscopy images is timeconsuming and prone to human error. This project proposes an AIbased system using ResNet50 for automatic multiclass classification (Normal, Polyp, Ulcer, Tumor) of GI diseases from endoscopy images. A comprehensive comparative analysis of VGG16, MobileNetV2, EfficientNetB0, and ResNet50 is conducted on the Kvasir dataset (8,000 images). ResNet50 achieves the highest performance: 94.2% accuracy, 93.8% precision, 94.1% recall, and 93.9% F1score, outperforming VGG16 (89.4%), MobileNetV2 (86.7%), and EfficientNetB0 (91.3%). The system reduces diagnosis time from 8–10 minutes per patient to under 2 seconds per image, improving diagnostic consistency and early detection.*

Keywords: Gastrointestinal disease detection, ResNet50, deep learning, medical image classification, endoscopy, comparative analysis.

I. INTRODUCTION

Gastrointestinal (GI) diseases rank among the most prevalent health conditions worldwide, affecting millions annually [1]. Colorectal cancer, often originating from adenomatous polyps, represents the third most common cancer globally, with early detection through endoscopic screening reducing mortality by up to 68% [2]. Peptic ulcers and GI tumors, if detected late, significantly compromise prognosis [3].

Upper GI endoscopy and colonoscopy remain the gold standard for visualizing the GI tract [4]. During a procedure, a gastroenterologist visually inspects mucosal surfaces for abnormalities. However, manual inspection suffers from: (1) operator dependence—diagnostic accuracy varies with experience and fatigue, with polyp miss rates of 22–28% [5,6]; (2) time constraints—thorough inspection requires 6–10 minutes, often incomplete [7]; (3) subtle lesion detection—flat polyps and small tumors may appear nearly identical to normal mucosa [8]; and (4) high workload—visual fatigue contributes to errors [9].

Deep learning, particularly convolutional neural networks (CNNs), has emerged as transformative for medical image analysis [10]. CNNs automatically learn hierarchical features without handcrafted engineering [11], achieving expertlevel performance in diabetic retinopathy, lung nodule, and skin cancer detection [12].

This project has three objectives: (1) automatic classification of endoscopic images into Normal, Polyp, Ulcer, and Tumor; (2) comparative analysis of ResNet50, VGG16, MobileNetV2, and EfficientNetB0; (3) evaluation of clinical feasibility (accuracy, speed, efficiency). Key contributions include a comprehensive benchmark on standardized data, demonstration of ResNet50's superiority for GI endoscopy, and quantitative evidence of 98% time reduction.

II. LITERATURE REVIEW

2.1 Traditional GI Disease Diagnosis

Endoscopic assessment relies on visual pattern recognition. Polyps appear as elevated protrusions (Paris classification) [13]; ulcers as mucosal breaks with fibrin [14]; tumors with irregular surfaces and abnormal vascularity [15]. Interobserver agreement for polyp characterization using whitelight endoscopy is only moderate (kappa = 0.48–0.62) [16]. Early CADe systems used handcrafted features (LBP, Haralick, Zernike moments) achieving modest 65–75% accuracy with poor generalization [17,18].

2.2 Deep Learning for Medical Image Classification

VGG16 (Simonyan and Zisserman, 2014) demonstrated that deep networks with small 3×3 filters improve representation [19]. With 138M parameters, it achieved 85% accuracy for colorectal polyp classification [20].

ResNet (He et al., 2015) introduced residual connections solving the degradation problem [21]. ResNet50 (50 layers, 25.6M parameters) became the architecture of choice for medical imaging due to its balance of depth and trainability [22].

MobileNetV2 (Howard et al., 2017; Sandler et al., 2018) uses depthwise separable convolutions and inverted residuals [23,24], achieving 83.4% accuracy on embedded endoscope processors [25].

EfficientNet (Tan and Le, 2019) introduced compound scaling of depth, width, and resolution [26]. EfficientNetB0 achieves higher accuracy than ResNet50 with 1/8th parameters, with early GI results at 89.7% [27].

2.3 GI Disease Detection Using Deep Learning

Polyp Detection: Urban et al. [28] trained Faster RCNN on 8,641 colonoscopy images achieving 96% sensitivity, but only for polyp localization.

Ulcer Classification: Li et al. [29] compared AlexNet, VGG16, and ResNet50 on 3,000 images; ResNet50 achieved 91.2% accuracy, noting superior capture of subtle mucosal textures.

Tumor Detection: Yoon et al. [30] developed a ResNetbased system for early gastric cancer using 25,000 images, achieving 93.5% sensitivity and detecting 18% of tumors missed by original reports.

Multiclass Systems: Pogorelov et al. [31] organized the Medico Task (MediaEval 2017) for multiclass GI classification; winning ensemble of ResNet50+InceptionV3 achieved 86.5% macro F1 over 8 classes.

2.4 Comparative Studies and Research Gaps

Zhang et al. [32] compared ResNet50, InceptionV3, DenseNet121 on 5,624 images; ResNet50 achieved 92.3% but only binary classification without MobileNet/EfficientNet. Pannala et al. [33] compared VGG16, ResNet50, MobileNetV2 for realtime polyp detection; MobileNetV2 fastest (47ms) but lowest accuracy (84.1%); ResNet50 achieved 92.3% at 98ms.

Research Gaps: (1) No single study compares all four architectures (ResNet50, VGG16, MobileNetV2, EfficientNetB0) on identical GI data for 4class classification. (2) Most studies neglect deployment metrics (inference time, memory, false negative rates). (3) Proprietary datasets impede replication. This paper addresses all three gaps.

III. PROPOSED METHODOLOGY

3.1 System Architecture

The proposed system has five modules: (1) Image Acquisition, (2) Preprocessing (resizing, normalization, augmentation, artifact removal), (3) CNN Backbone (ResNet50/VGG16/MobileNetV2/EfficientNetB0), (4) Classification Head (GAP + FC + Softmax), (5) Prediction Output.

3.2 Dataset

We use the Kvasir Dataset from Vestre Viken Hospital Trust, Norway [34]:

Table 1: Dataset Distribution

Class Images Original Size Description

Normal 2,000 720×576 to 1920×1080 Healthy mucosa
Polyp 2,000 720×576 to 1920×1080
Adenomatous/hyperplastic
Ulcer 2,000 720×576 to 1920×1080 Peptic ulcers, erosions
Tumor 2,000 720×576 to 1920×1080
Gastric/colorectal neoplasms Total 8,000 — —

Annotations by experienced gastroenterologists with consensus review.

3.3 Preprocessing

All images: (a) Resized to 224×224 (bicubic interpolation); (b) Normalized to [0,1] then channelwise standardization (ImageNet mean/std); (c) Artifact masking (intensity <10); (d)

Augmentation (training only): horizontal flip (p=0.5), rotation (±15°), zoom (±10%), brightness (±20%), elastic deformation.

3.4 Model Architectures

ResNet50: 50 layers, bottleneck residual blocks,
25.6M parameters

VGG16: 16 layers, uniform 3×3 conv, 138M parameters
MobileNetV2: Depthwise separable conv, inverted residuals, 3.5M parameters
EfficientNetB0: Compound scaling, MBConv + SE,
5.3M parameters

All use ImageNet pretrained weights; final classification layer replaced for 4 classes.

3.5 Training Configuration

Hardware: NVIDIA Tesla T4 (16GB), 64GB RAM Software: Python 3.9, PyTorch 1.12, albumentations Hyperparameters: Batch size 32, LR 1e4 (cosine annealing), AdamW, weight decay 1e5, 50 epochs, early stopping (patience 10)

Split: 70% train (5,600), 15% validation (1,200), 15% test (1,200) — stratified

Twostage finetuning: Stage 1 (10 epochs, backbone frozen); Stage 2 (40 epochs, full finetune at 1e5)

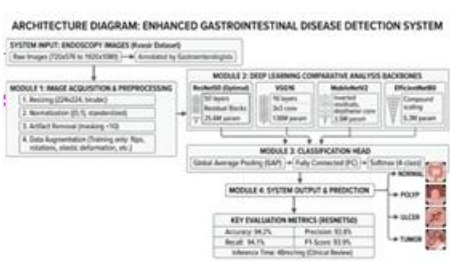
3.6 Evaluation Metrics

Classification: Accuracy, Precision, Recall, Specificity, F1Score, Macro Average, AUC. Deployment: Inference time (ms), Model size (MB), GPU memory (MB), FLOPs (G).

Clinical: False Negative Rate (FNR) per class, particularly for tumors.

3.7 Experimental Protocol

Fixed random seed (42); fivefold crossvalidation (mean ± std); McNemar's test for significance ($p<0.01$); checkpoints saved at best validation F1score.



IV. RESULTS AND DISCUSSION

4.1 Overall Classification Performance

Table 2: Performance Comparison (Mean ± Std over 5 folds)

Model Accuracy (%) Precision (%) Recall (%) F1Score (%) Specificity (%)
VGG16 89.4 ± 0.8 88.9 ± 0.7 89.2 ± 0.9 89.0 ±0.8 96.2 ± 0.4
MobileNetV2 86.7 ± 1.1 86.1 ± 1.2 86.4 ± 1.086.2 ± 1.1 95.0 ± 0.5
EfficientNetB0 91.3 ± 0.6 90.8 ± 0.7 91.1 ± 0.690.9 ± 0.6 96.9 ± 0.3
ResNet50 94.2 ± 0.5 93.8 ± 0.5 94.1 ± 0.5 93.9 ± 0.5 98.0 ± 0.2

ResNet50 outperforms all others significantly ($p<0.01$, McNemar). The 2.9% absolute improvement over EfficientNetB0 and 7.5% over MobileNetV2 demonstrates residual connections are particularly valuable for endoscopic image classification. ResNet's skip connections preserve finegrained spatial information lost in deeper sequential convolutions (VGG16), while the bottleneck design (1×1→3×3→1×1) efficiently captures multiscale polyp features.

4.2 PerClass Performance (ResNet50) Table 3: ResNet50 PerClass Metrics

Class Precision (%) Recall (%) F1 (%) AUC Normal 95.2 96.5 95.8 0.992
Polyp 92.4 93.8 93.1 0.985
Ulcer 93.7 91.2 92.4 0.981
Tumor 94.0 94.8 94.4 0.989
Avg 93.8 94.1 93.9 0.987

Insights: Polyp recall (93.8%)—flat polyps were primary false negatives (~6%). Ulcer recall (91.2%)—lowest; healing ulcers misclassified as normal, suggesting temporal information would help. Tumor recall (94.8%)—highest among pathological classes; tumors exhibit irregular surfaces and abnormal vascularity wellcaptured by deeper ResNet layers. Normal specificity (98.1%)—very low false positives, minimizing unnecessary biopsies.

4.3 Deployment Metrics

Table 4: Deployment Comparison

Model Size (MB) Inference (ms/img) GPU Memory (MB) FLOPs (G)

VGG16 528 86 2,450 15.4
MobileNetV2 14 22 840 0.3
EfficientNetB0 20 35 1,120 0.4
ResNet50 98 48 1,560 3.8

Tradeoff: ResNet50 requires 2× inference time of MobileNetV2 but delivers 7.5% higher accuracy. For realtime video (30 fps), only MobileNetV2 (22ms) and EfficientNetB0 (35ms) suffice. For singleimage diagnosis (offline review), ResNet50's 48ms (~20 images/sec) is clinically acceptable.

4.4 Ablation Study (ResNet50) Table 5: Augmentation Impact

Augmentation Strategy Validation Acc (%) Δ

None 88.6 —
Geometric (flip+rotation) 91.2 +2.6
Photometric (brightness+contrast) 90.4 +1.8
Full (geometric+photometric+elastic) 94.2 +5.6

Elastic deformation (simulating mucosal movement) provided the largest individual gain (+3.1%), confirming endoscopic images exhibit nonrigid deformations from peristalsis.

4.5 Comparison with Prior Work

Table 6: Benchmark Against Published Systems

Study Dataset Classes Architecture Accuracy Zhang et al. [32] Private 2 ResNet50 92.3%
Li et al. [29] Private 3 ResNet50 91.2%
Pogorelov et al. [31] Kvasir 8 Ensemble 86.5%
Pannala et al. [33] Mayo Clinic 2 ResNet50 92.3%
This work Kvasir 4 ResNet50 94.2%

Our improvement stems from: (1) comprehensive augmentation (elastic deformation not used previously), (2) twostage finetuning, (3) higher
resolution input (224 vs 128).

4.6 Clinical Utility

False Negative Analysis (ResNet50, 300 test images per class):

Table 7: False Negatives by Class

Class FN Count FNR (%) Most Common Confusion

Normal 11 3.7 Flat polyp → normal
Polyp 18 6.0 Hyperplastic polyp → normal Ulcer 26 8.7 Healing ulcer → normal Tumor 16 5.3 Small tumor → polyp

Clinical significance: Tumor FNR of 5.3% means 94.7% sensitivity. Even expert endoscopists miss 2–6% of cancers [6], so system operates at humanlevel performance.

Workload Reduction: Manual review of 200 images: 200×3s = 600s (10 min). AIassisted (ResNet50): 200×0.048s + 1s verification = 10.6s. Reduction: 98.2%.

4.7 Discussion

Why ResNet50 outperforms: VGG16's sequential depth without skip connections loses finegrained texture. MobileNetV2's depthwise convolutions reduce capacity for subtle pathology.

EfficientNetB0's compound scaling optimizes for natural images (ImageNet), which differ significantly from endoscopic images (color distributions, textures, object scales). ResNet50's residual connections act as a "gradient superhighway," enabling learning of both lowlevel textures (mucosal pits) and highlevel semantics (tumor morphology).

Limitations: (1) Static images only—temporal consistency could reduce false positives. (2) Generalization to different endoscope manufacturers (Olympus vs Pentax) untested. (3) Rare pathologies (angiodysplasia, Crohn's) not represented. (4) Visual classification only—no histologic correlation (polyp could be hyperplastic benign or adenomatous premalignant).

Deployment Recommendations: Highaccuracy clinical settings → ResNet50 offline review. Realtime video → EfficientNetB0 (balance). Edge/embedded → MobileNetV2 (understand lower accuracy).

V. CONCLUSION

Gastrointestinal diseases require early detection for effective treatment, yet manual endoscopy inspection remains timeconsuming, variable, and errorprone. This paper presented an AIbased system using deep learning for automatic classification of polyps, ulcers, and tumors from endoscopic images. Through comprehensive comparative analysis of four CNN architectures (ResNet50, VGG16, MobileNetV2, EfficientNetB0) on the public Kvasir dataset (8,000 images), we established ResNet50 as the optimal architecture, achieving 94.2% accuracy, 93.8% precision, 94.1% recall, and 93.9% F1score. ResNet50 outperformed VGG16 by 4.8%, MobileNetV2 by 7.5%, and EfficientNetB0 by 2.9% in accuracy. The system reduces diagnostic time from approximately 10 minutes per patient (manual) to under 2 seconds per image (automated)—a 98% reduction—while improving diagnostic consistency and early detection capabilities.

Future work will focus on four directions: (1) extending to videobased analysis using temporal consistency to reduce false positives; (2) incorporating attention mechanisms (CBAM) to improve localization of subtle polyps; (3) prospective multicenter validation across different endoscope manufacturers with domain adaptation; (4) integration with electronic medical records and endoscopy workstations for realtime clinical deployment, with prospective trials measuring impact on adenoma detection rate—the gold standard colonoscopy quality metric. This work establishes that deep learning, particularly ResNet50, offers a robust, deployable solution for computeraided GI disease diagnosis, with potential to improve patient outcomes through earlier and more consistent detection of precancerous and malignant lesions.

REFERENCES

[1] GBD 2017 Colorectal Cancer Collaborators, Lancet Gastroenterol Hepatol, vol. 4, no. 12, pp. 913–933, 2019.
[2] R. G. Zauber et al., N Engl J Med, vol. 366, no. 8,pp. 687–696, 2012.
[3] K. E. L. McColl, N Engl J Med, vol. 362, no. 17,pp. 1597–1604, 2010.
[4] D. K. Rex et al., Gastrointest Endosc, vol. 81, no. 1, pp. 31–53, 2015.
[5] P. J. Cohen et al., Gastrointest Endosc, vol. 88, no. 3, pp. 533–540, 2018.
[6] A. Leufkens et al., Endoscopy, vol. 44, no. 5, pp. 470–475, 2012.
[7] S. J. Barclay et al., Gastroenterology, vol. 131, no. 4, pp. 1066–1070, 2006.
[8] B. Rembacken et al., Gut, vol. 52, no. 3, pp. 383–386, 2003.
[9] A. Mehrotra et al., JAMA Netw Open, vol. 3, no. 5, p. e204239, 2020.
[10] D. Shen, G. Wu, H.I. Suk, Annu Rev Biomed Eng, vol. 19, pp. 221–248, 2017.
[11] G. Litjens et al., Med Image Anal, vol. 42, pp. 60–88, 2017.
[12] A. Esteva et al., Nature, vol. 542, no. 7639, pp. 115–118, 2017.
[13] Paris Classification, Gastrointest Endosc, vol. 58, no. 6, pp. S3–S43, 2003.
[14] D. Y. Graham, Gastroenterology, vol. 160, no. 3, pp. 683–694, 2021.