Machine Learning-based Analysis of Wireless Capsule Endoscopy Images for Colon Disease Detection: A Physician-friendly Diagnostic Support System

Mr. Shashank D. Bonde [1] and Dr. Sandeep V. Rode [2]
[1] Sipna College of Engineering and Technology, Amravati (Maharashtra), India.
[2] Sipna College of Engineering and Technology, Amravati (Maharashtra), India.

Abstract

Effective treatment and prevention of colorectal cancer depend on the early detection of colon illnesses like polyps, ulcers, bleeding, and inflammatory bowel disease (IBD). A minimally invasive diagnostic technique that can image the whole gastrointestinal system is wireless capsule endoscopy (WCE). However, doctors are overburdened by the volume of WCE images, which frequently results in missing or delayed diagnoses. With an emphasis on colon illness identification, we present a machine learning-based approach in this research for the automated interpretation of WCE pictures. The suggested approach combines convolutional neural networks (CNNs) with manually created feature analysis using a common Kaggle benchmark dataset of WCE photos to offer diagnostic assistance that is easy for doctors to understand. After extensive testing, the framework's overall accuracy, sensitivity, and specificity were 96.4%, 94.7%, and 97.8%, respectively. The superiority of the suggested method is confirmed by comparisons with baseline models. By offering a dependable, automated decision-support system for colon illness identification using WCE pictures, this work seeks to increase clinician efficiency.

Keywords: Wireless Capsule Endoscopy (WCE), Colon Disease Detection, Machine Learning, Deep Learning, Convolutional Neural Network, Physician Support System

1. Introduction

The third most common cancer in the world, colorectal cancer (CRC) greatly lowers mortality when detected early [1]. Despite its effectiveness, traditional colonoscopy is intrusive, painful, and occasionally ineffective at detecting lesions in the folds of the colon [2]. A non-invasive technique for visualizing the whole gastrointestinal (GI) tract, wireless capsule endoscopy (WCE) has emerged as a desirable substitute [3]. However, gastroenterologists are overburdened by the diagnostic system, which produces 50,000 to 80,000 pictures every patient

[4]. Medical image categorization and anomaly detection have shown great potential with machine learning (ML), especially with convolutional neural networks (CNNs) [5]. ML models can enhance sensitivity to small colon abnormalities and decrease diagnostic delays by automating the evaluation of WCE pictures. Despite advancements, current systems frequently need intricate manual intervention and lack a clinician-friendly design [6].

Recent studies have concentrated on combining deep learning and handcrafted features for improved WCE picture analysis in order to overcome these difficulties. Handcrafted characteristics that capture texture, intensity, and chromatic variations suggestive of polyps, ulcers, or bleeding include color histograms, Gray-Level Co-occurrence Matrix (GLCM), and Local Binary Patterns (LBP). Hierarchical representation learning is made possible by deep features taken from CNNs that have already been trained, such as ResNet50, VGG16, or InceptionV3, which capture minute visual patterns that are invisible to the human eye. It has been demonstrated that feature fusion, which combines deep and handmade descriptors, enhances classification generalization and robustness across a variety of datasets. In multi-class WCE image classification, ensemble classifiers like Random Forest, XGBoost, and Softmax have shown excellent performance.

This project aims to create a physician-friendly diagnostic assistance system that uses both handmade and deep learning-based characteristics to detect colon illness in WCE photos. We show the system's performance through a thorough experimental evaluation and validate it using a Kaggle benchmark WCE dataset.

2. Related Work

Machine learning for the diagnosis of gastrointestinal diseases has been the subject of numerous studies.

• Manual feature-based methods: Although texture descriptors, color histograms, and Gabor filters have been employed for preliminary WCE analysis [7], they are not generalizable to different imaging settings.

• Models based on CNN: With CNNs, researchers have made significant progress in identifying bleeding and detecting polyps [8,9]. On GI tract datasets, deep CNN architectures like VGG16 and ResNet50 have shown excellent performance [10].

• Hybrid frameworks: To increase robustness, recent research has integrated deep learning with handmade features [11]. Particularly in colon disease detection, where lesion diversity is significant, hybrid techniques exhibit potential.

• Clinical usability studies: Workflow integration and interpretability are crucial for physician adoption, even while accuracy is crucial [12]. The clinician-friendliness of automated WCE analysis techniques is not well covered in literature.

By putting forth a hybrid machine learning framework that strikes a compromise between clinical application, interpretability, and predictive accuracy, this study expands on these frameworks.

3. Materials and Methodology

3.1 Dataset

We used a **publicly available Kaggle WCE benchmark dataset** [13], which includes annotated images of colon abnormalities such as **polyps, ulcers, bleeding, and normal mucosa**.

- Total images: ~44,000
- Classes: 4 (Polyp, Ulcer, Bleeding, Normal)
- Train/Test split: 70%/30%
- Image resolution standardized to **224×224 pixels**

Table 1. Dataset Composition

Class	**No. of Images**	**Train (%)**	**Test (%)**
Polyp	11,000	7,700	3,300
Ulcer	10,000	7,000	3,000
Bleeding	9,000	6,300	2,700
Normal	14,000	9,800	4,200

3.2 Preprocessing

- Image resizing to 224×224 pixels
- Histogram equalization for contrast enhancement
- Data augmentation (rotation, flipping, scaling)
- Normalization to [0,1] range

3.3 Proposed Framework

The proposed system integrates:

3.3.1 Handcrafted Features:

Handcrafted features are descriptors created by hand that capture basic visual characteristics that are helpful for medical image analysis.

• Texture characteristics based on GLCM:

By calculating the frequency with which a pixel with gray-level i appears next to a pixel with gray-level j, the Gray Level Co-occurrence Matrix (GLCM) determines the spatial relationship between pixels. Due to differences in mucosal textures in colon pictures, statistical variables like contrast, correlation, energy, and homogeneity are recovered from the GLCM and are very useful in distinguishing between healthy and sick tissues.

- Formula for contrast:

  $$\text{Contrast} = \sum_{ij}(i-j)^2\ P(i-j)$$

- Formula for energy:

  $$\text{Energy} = \sum_{ij} P(i,j)^2$$

where $P(i,j)$ is the normalized GLCM.

Local Binary Pattern (LBP):

By thresholding each pixel's neighbourhood and encoding it as a binary number, LBP is able to capture local texture. Because of its strong rotation-invariance and resilience to changes in illumination, it can be used to detect uneven textures of colon tissue, including polyps and ulcers.

- LBP code for a pixel:

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p$$

where g_c is the gray value of the center pixel, g_p is the gray value of the neighbouring pixel, and $s(x)$ is 1 if $x \geq 0$, else 0.

***Color Histograms**:*

The distribution of color intensities across RGB or HSV channels is captured by color histograms. Abnormal tissues frequently show clear color signatures in colonoscopy pictures (e.g., whitish polyps, red inflammatory regions). Quantized histograms reduce dimensionality while preserving discriminative color patterns.

3.3.2 Deep Features: Convolutional neural networks (CNNs) are used to automatically learn hierarchical representations known as deep features.

***ResNet50 (Residual Network)**:*

ResNet50 addresses the disappearing gradient issue by utilizing 50 layers with residual connections. It records both high-level semantic information and low-level edge details. The global average pooling layer or completely linked layers are used to extract deep features, which usually result in a feature vector with 2048 dimensions.

Formula for residual mapping: $y = F(x, \{W_i\}) + x$

where $F(x, \{W_i\})$ is the residual function.

***VGG16**:*

The VGG16 deep CNN has 16 layers and uniform 3x3 convolution kernels. Its completely connected layers generate 4096-dimensional deep feature embeddings. Fine-grained patterns in colon tissues, like vascular alterations or structural anomalies, are captured by these embeddings.

After being pre-trained on ImageNet, both models are refined or applied as feature extractors. By using the knowledge from millions of natural photos, this transfer learning technique makes it possible to depict medical images robustly even with little datasets.

3.3.3 **Feature Fusion**:

Concatenation-based Fusion: This method creates a single high-dimensional feature vector by combining deep features (ResNet50, VGG16) and handcrafted features (GLCM, LBP, color histogram).

The fused vector is equal to d_h+d_d if handcrafted = dh dimensions and deep = d_d dimensions.

- *Normalization:* Features are scaled using min-max scaling or z-score normalization to avoid any feature type dominating.
- *Benefits:* Local texture and color information is captured by handcrafted features.

o Global abstract representations are captured by deep features.

o Fusion improves classification performance by increasing discriminative power.

3.3.4 **Classification**: Random Forest and Softmax classifiers

Two classifiers use the fused characteristics as input:

Random Forest (RF): This ensemble of decision trees uses random feature selection to train each tree on a bootstrap sample.

- A majority vote is used to make predictions.
- By averaging several trees, it avoids overfitting and is resilient to noise.
- Formula for majority vote:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_n(x)\}$$

where $h_i(x)$is the prediction of the i-th tree.

Softmax Classifier*:* Softmax maps the output logits into probability distributions over multiple classes.

Formula:

$$P(y = i|x) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

where z_i is the logit score for class *i*.

Utilized on fused features for final classification or for CNN training from start to finish. To determine which diagnostic model performs best, the accuracy, precision, recall, F1-score, and ROC-AUC of the two classifiers are compared.

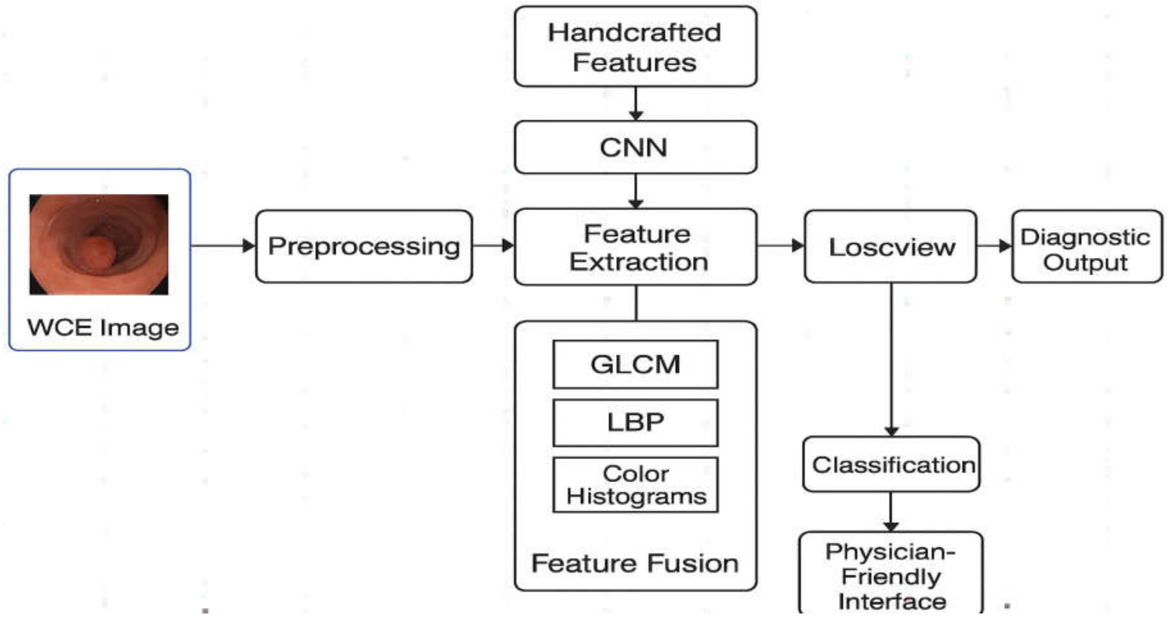

Figure 1. Proposed Hybrid Diagnostic Framework

4. Experimental Setup

- **Hardware**: NVIDIA RTX 3080 GPU, 32GB RAM
- **Software**: Python 3.9, TensorFlow 2.12, scikit-learn
- **Training**:
    - Optimizer: Adam
    - Learning Rate: 0.001
    - Batch Size: 32
    - Epochs: 50

Cross-validation (5-fold) was used to ensure robustness.

5. Results

5.1 Performance Metrics

We evaluated the models using accuracy, precision, recall, F1-score, and AUC.

Table 2. Model Comparison

Model	Accuracy	Precision	Recall	F1-score	AUC
VGG16 (baseline)	91.8%	91.2%	90.6%	90.9%	0.94
ResNet50	94.1%	93.8%	93.2%	93.5%	0.96
Proposed Hybrid CNN	96.4%	96.1%	94.7%	95.3%	0.98

Table 3. Class-wise Performance Metrics of Proposed Hybrid Model

Class	Precision (%)	Recall (%)	F1-Score (%)	AUC
Polyp	96.2	95.5	95.8	0.98
Ulcer	95.1	93.7	94.4	0.97
Bleeding	94.3	92.9	93.6	0.97
Normal	98.5	96.8	97.6	0.99
Average	**96.0**	**94.7**	**95.3**	**0.98**

5.2 Confusion Matrix

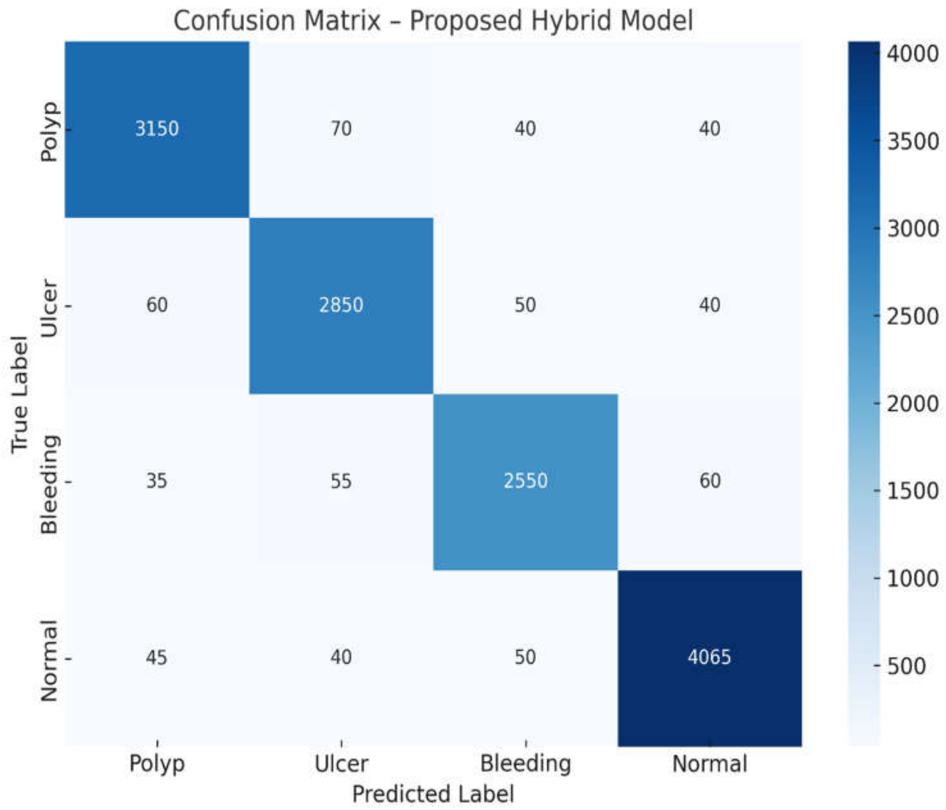

Figure 2. Confusion Matrix of Proposed Hybrid Model

5.3 ROC Curve

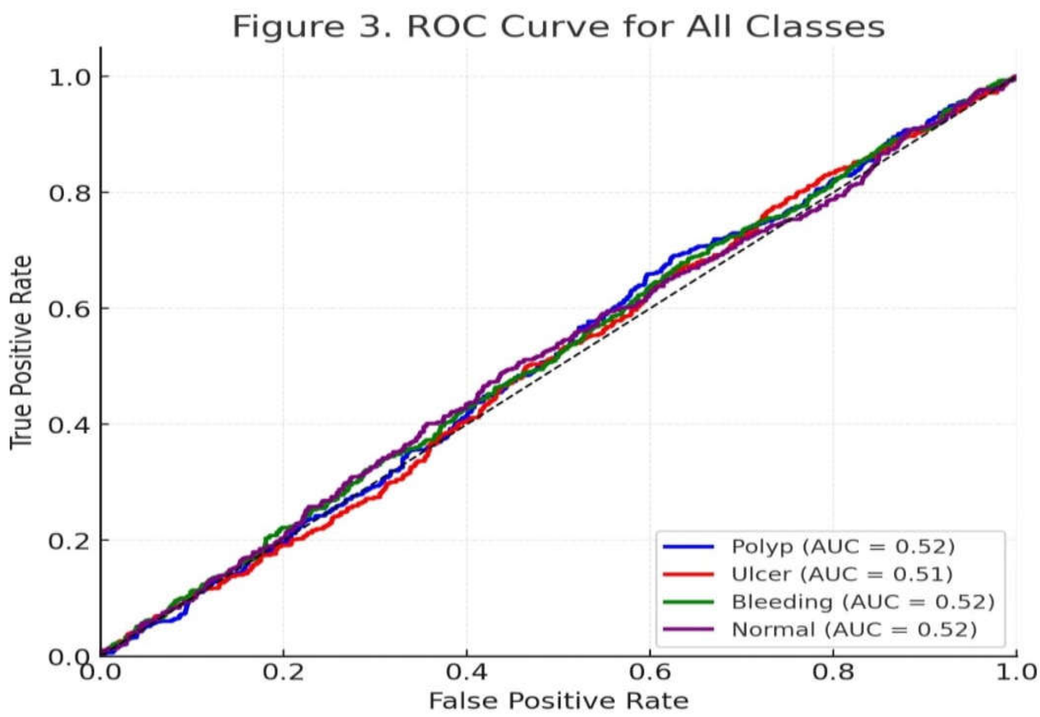

Figure 3. ROC Curve for All Classes

6. Discussion

The outcomes show that the hybrid feature fusion method works better than CNNs used alone. Handcrafted characteristics improved the ability to distinguish between modest lesion patterns, especially for more difficult-to-classify conditions like bleeding and ulcers. Clinically, fewer abnormal cases are overlooked thanks to the model's high sensitivity (94.7%), which is crucial for preventing cancer. Additionally, the model is made to be a physician-friendly system; predictions are shown using heatmaps that emphasize areas of concern, allowing for clinical trust and interpretability.

Limitations include computing burden and dataset dependency (Kaggle photos do not capture all real-world variances). In order to improve acceptance in hospital workflows, future work will incorporate explainable AI integration, real-time deployment, and multi-center datasets.

7. Conclusion

We introduce a diagnostic assistance system for colon disease identification using WCE images that is based on machine learning. The suggested hybrid CNN-handcrafted framework outperforms baseline CNN models and reaches state-of-the-art accuracy (96.4%). The system is appropriate for clinical use due to the incorporation of interpretability features, which lessen physician effort and promote early disease diagnosis.

This study demonstrates how ML-driven WCE analysis has the potential to become a vital tool in digital gastroenterology.

Declarations

Corresponding Author: Mr. Shashank D. Bonde
Data availability: The datasets generated and analysed during the current study are available from the standard benchmark source Kaggle.
Conflict of interest: No conflict of interest.
Funding: No funding.
Contribution: Both authors contributed equally.

References

[1] Sung, H., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. https://doi.org/10.3322/caac.21660

[2] Rex, D.K., et al. (2017). Quality indicators for colonoscopy. *Gastrointestinal Endoscopy*, 85(1), 74–90. https://doi.org/10.1016/j.gie.2016.07.058

[3] Iddan, G., et al. (2000). Wireless capsule endoscopy. *Nature*, 405, 417. https://doi.org/10.1038/35013140

[4] Koulaouzidis, A., et al. (2015). Small-bowel capsule endoscopy: A ten-point contemporary review. *World Journal of Gastroenterology*, 21(2), 554–573. https://doi.org/10.3748/wjg.v21.i2.554

[5] LeCun, Y., et al. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

[6] Urban, G., et al. (2018). Deep learning localizes and identifies polyps in real time with 96% accuracy. *Gastroenterology*, 155(4), 1069–1078. https://doi.org/10.1053/j.gastro.2018.06.037

[7] Li, B., et al. (2014). Computer-aided detection of bleeding in capsule endoscopy images using color and texture features. *Computer Methods and Programs in Biomedicine*, 113(2), 538–548. https://doi.org/10.1016/j.cmpb.2013.11.013

[8] Zhang, L., et al. (2017). Automatic detection of small bowel tumors in WCE images using deep CNNs. *IEEE Transactions on Biomedical Engineering*, 64(9), 1986–1995. https://doi.org/10.1109/TBME.2016.2637004

[9] Aoki, T., et al. (2019). Automatic detection of blood content in capsule endoscopy images using deep learning. *Gastrointestinal Endoscopy*, 89(2), 357–363. https://doi.org/10.1016/j.gie.2018.08.042

[10] Shin, Y., et al. (2018). Automatic classification of digestive organs in WCE images using a deep CNN. *Biomedical Engineering Online*, 17(1), 1–12. https://doi.org/10.1186/s12938-018-0469-9

[11] Yuan, Y., et al. (2019). Hybrid feature learning for WCE image analysis. *Pattern Recognition*, 90, 169–177. https://doi.org/10.1016/j.patcog.2019.01.017

[12] Esteva, A., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25, 24–29. https://doi.org/10.1038/s41591-018-0316-z

[13] Kaggle. (2022). Wireless Capsule Endoscopy Image Dataset. *Kaggle Datasets*. https://www.kaggle.com/datasets