International Journal of Basic and Applied Sciences, 15 (2) (2026) 22-30

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/rcjf9t51
Research paper

Boosted Capsule Network for Gastrointestinal Disease Recognition

Henrietta Adjei Pokuaa [1, 2] *, Adebayo Felix Adekoya [3], Benjamin Asubam Weyori [4], Mighty Abra Ayidzoe [1]

[1] Department of Computer Science and Informatics, University of Energy and Natural Resources
[2] Department of Computer Science, Sunyani Technical University
[3] Faculty of Computing Engineering and Mathematical Science, Catholic University of Ghana
[4] Department of Computer and Electrical Sciences, University of Energy and Natural Resources
**Corresponding author E-mail: henrietta.opokuaa@stu.edu.gh*

Received: January 1, 2026, Accepted: February 6, 2026, Published: February 23, 2026

Abstract

Gastrointestinal diseases affect over 40% of the global population and rank among the leading causes of cancer-related deaths worldwide. Wireless capsule endoscopy (WCE), though minimally invasive, generates 45,000-50,000 images per procedure, with expert endoscopists missing 22-28% of pathological cases due to examination complexity and varied disease presentations. Deep learning approaches like Convolutional Neural Networks require large annotated datasets, which are rarely available in medical imaging. Capsule Networks (CapsNets) excel in limited-data scenarios through spatial-hierarchy inference but struggle with subtle features in complex medical images. We present a Boosted Capsule Network that incorporates dual enhancement mechanisms: modified feature boosting via intensity inversion to expose diverse low-level patterns, reducing the generalization gap by 45%, and class capsule amplification to sharpen decision boundaries and improve inter-class separation. Ablation studies on the Kvasir-V2 dataset show individual components achieve 89.3% (feature boosting) and 91.7% (class boosting), while their synergistic combination reaches 97.90% accuracy, a 15.8% improvement over baseline CapsNet (82.10%) and 1.1% over previous state-of-the-art (96.80%). The lightweight architecture adds zero trainable parameters for class boosting and minimal overhead for feature enhancement, enabling real-time clinical deployment. We experimentally deployed the model as a web-based prototype with human-in-the-loop uncertainty handling for confidence scores below 0.5.

Keywords: *Capsule Networks; Convolutional Neural Networks; Feature Boosting; Gastrointestinal Diseases; Class Capsule Boosting.*

1. Introduction

The gastrointestinal (GI) tract is a vital system in the human body and is therefore susceptible to various diseases. Peptic ulcer, colon cancer, gastritis, polyps, ulcerative colitis, esophagitis, GI bleeding, and others are some examples of GI diseases. According to research, more than 40% of people in the world are affected by these GI diseases (Bajhaiya & Unni, 2024). Wireless Capsule Endoscopy (WCE) is a standard medical procedure used to detect GI diseases. It involves a patient swallowing a pill-sized capsule, and images of the GI tract are then captured for analysis. The WCE process can last approximately 8 hours (Raut et al., 2023). Around 45,000 – 50,000 images are captured, and it takes 1-3 hours for the frames to be analyzed by the doctors. It is a complex and time-consuming task, as each frame needs careful examination. The complex nature of the examination is prone to human error, and as such, 22-28% of GI diseases are missed by expert endoscopists (Freedman et al., 2020). Another factor contributing to this human error is the varied nature of GI diseases. For instance, polyps have varied characteristics in size, shape, and color. Hyperplastic polyps and adenomas may look very similar, especially to less experienced endoscopists. Considering the negative impact of GI diseases, it is imperative to adopt intelligent algorithms to analyze gastrointestinal disease images.

Diverse Deep Learning algorithms have been developed and trained to identify these GI diseases. Most of these algorithms, specifically Convolutional Neural Networks, require large training data to accurately learn the varied patterns. However, medical datasets are limited and smaller in size. Additionally, the collection of these data requires experienced medical officers to annotate the images. Furthermore, the health domain has a lot of regulatory policies that limit the sharing of medical information. This, therefore, has created the need to use data augmentation techniques, which are time-consuming and may miss subtle features necessary for accurate prediction.

To address these challenges, there is a need for the adoption of algorithms capable of learning effectively from limited data. Capsule Networks (CapsNets) are an example of such algorithms. CapsNets are a class of Deep Learning models known for their equivariance (Sabour et al., 2017), that is, they can infer spatial hierarchies and pose information directly from images, reducing the need for extensive data augmentation. However, standard CapsNets face two critical limitations in medical imaging: (1) insufficient feature diversity when trained on small datasets, leading to generalization gaps where training accuracy significantly exceeds validation accuracy, and (2) weak inter-class discrimination in the class capsule layer, particularly for visually similar conditions such as hyperplastic polyps versus adenomas. The existing body of work on CapsNets for GI disease recognition can be categorized into three main approaches: (1) preprocessing-focused methods that enhance input images before feeding them to CapsNets (Ayidzoe et al., 2021; Afriyie et al., 2022), (2) architecture-focused methods that increase model complexity through hybrid CNN-CapsNet designs (Wang et al., 2022; Sarsengeldin et al., 2023), and (3) transfer learning approaches that leverage pre-trained features (Sarsengeldin et al., 2023). While these methods have advanced the field, they share common limitations: increased computational overhead from preprocessing pipelines or architectural complexity, and insufficient attention to the synergistic interaction between feature learning and classification mechanisms. Our work departs from these approaches by introducing lightweight boosting mechanisms that operate at both the feature extraction and classification stages, achieving superior performance without additional architectural complexity.

The contributions of this paper are as follows:

1) A dual-mechanism boosted Capsule Network achieving 97.90% accuracy on five-class GI disease classification, 15.8% improvement over baseline CapsNet, and 1.1% over previous state-of-the-art, through a synergistic combination of feature and class capsule enhancements.
2) A modified feature boosting algorithm incorporating 1×1 convolutional projection followed by negative scaling transformation ($h - \lambda h$, $\lambda \in [2,4]$), reducing the generalization gap by 45% (from 11.2% baseline train-test divergence to 6.2% proposed) without requiring data augmentation.
3) Class capsule amplification via post-routing scaling ($\alpha=2$), sharpening output probability distributions, and improving inter-class logit separation by 73% (mean separation: 0.34 baseline → 0.59 proposed), verified through ablation studies.
4) Comprehensive ablation analysis demonstrating that neither component independently achieves >92% accuracy, proving synergistic interaction rather than additive improvements.
5) Clinical deployment prototype with human-in-the-loop mechanisms for cases with confidence <0.5.

The rest of the paper is organized as follows: Section 2 presents relevant literature in the domain of Capsule Networks for Gastrointestinal Disease recognition. Section 3 presents the proposed method and model. Section 4 presents the results and discussion of the findings. The paper is concluded in Section 5.

2. Related Works

The application of deep learning to gastrointestinal disease recognition has evolved through distinct methodological phases. Early approaches centered on Convolutional Neural Networks (CNNs), which, despite achieving promising results, face fundamental limitations when applied to medical imaging domains characterized by limited annotated data. For example, (Ahmed, 2022) proposed a model that combines noise reduction techniques with pre-trained CNNs for classifying gastrointestinal diseases. The model consists of two main sub-architectures, each with a distinct purpose. The first sub-architecture focuses on eliminating noise artifacts introduced during image acquisition and transmission. To achieve this, the author adopted the noise reduction method proposed by (Zhang et al., 2017), which treats noise as a discriminative learning problem. The second sub-architecture employs a pre-trained AlexNet model to extract, process, and classify image patterns. When evaluated on the Kvasir dataset (Pogorelov, Konstantin, Randel, Kristin Ranheim, Griwodz et al., 2017), the model achieved a recognition accuracy of 90.17%. (Lonseko et al., 2021) introduced an attention-guided CNN, applying several preprocessing techniques to the input images. Their approach involved five key steps. First, textual inscriptions were removed from the images, and the images were resized to 224 × 224 pixels. Second, the preprocessed images were used as input for the entire experiment. Third, an attention mechanism was employed to identify diseases. Fourth, the model was validated using 20% of the dataset. Lastly, multiple evaluation metrics were used to assess the model's performance. Their method achieved a test accuracy of 93.19% on a self-collected dataset. Oh et al. (2021) utilized the EfficientNet CNN model to detect abnormalities in gastric tumor ultrasound images. Their approach achieved a detection accuracy of 91.20%. Similarly, Hasan et al. (2022) explored various CNN architectures in combination with hand-crafted feature extractors. To improve accuracy, a multi-criteria frame selection method was applied to identify optimal input frames from colonoscopy videos. Their method also included irrelevant feature elimination and polyp localization using image patches. Among all configurations, the Xception model combined with a nonsubsampled contourlet transform performed best, achieving an overall accuracy of 84.05%. Mohapatra et al. (2021) proposed a smart health system for identifying abnormalities in gastrointestinal tract images. The model involved preprocessing, extracting discrete wavelength coefficients, and feeding the decomposed image data into a CNN to extract features and perform classification. This approach achieved a recognition accuracy of 97.25%. Despite these successes, CNN-based approaches consistently require either large datasets or extensive data augmentation pipelines to achieve robust performance. The need for thousands of training samples and computationally expensive preprocessing steps motivates exploring alternative architectures specifically designed for limited-data scenarios. Capsule Networks represent a paradigm shift in addressing limited-data medical imaging challenges. Their equivariance property enables inference of spatial hierarchies from fewer samples, theoretically reducing dependence on data augmentation. However, implementations in GI disease recognition reveal a gap between theoretical potential and practical performance. CapsNets can infer pose information from fewer samples, which has encouraged their application in medical imaging. Ayidzoe et al. (2021) introduced a Gabor Capsule Network incorporating custom preprocessing blocks to detect gastrointestinal anomalies. Their method amplified features through multiplicative enhancement and consisted of two sub-architectures: one for feature extraction and classification, and the other for input reconstruction. The proposed model achieved a recognition accuracy of 91.50% on the Kvasir dataset. Though the proposed model had 9.6% improvement over the baseline CapsNet, which achieved ~82% recognition accuracy, it comes at a high cost; the multiplicative enhancement amplifies both signal and noise, and the Gabor filter banks require $O(n^2)$ convolutions per orientation/frequency combination. Afriyie et al. (2022) proposed a modified Capsule Network architecture emphasizing feature enhancement and noise removal within the encoder layer. This variant attained a recognition accuracy of 94.16% on the Kvasir dataset. However, the architectural modification adds trainable parameters (exact count unreported), raising a fundamental question: "Is the improvement due to better noise handling or simply increased model capacity?". Without ablation studies, isolating the denoising impact from parameter increase, the attribution remains ambiguous. This represents a common limitation in CapsNet research; architectural enhancements often conflate multiple changes, making it difficult to identify active ingredients. Sarsengeldin et al. (2023) integrated the VGG-16 architecture with the primary capsules of the CapsNet algorithm, employing transfer learning on VGG-16 to accelerate feature learning. The model achieved an accuracy of 83.00% on the HyperKvasir dataset. Though their approach achieved a comparable recognition accuracy for 83.00%, it faces a significant spatial analysis limitation. VGG-16's max-pooling layers discard spatial information that CapsNets need for pose encoding, creating an "impedance mismatch" between components. Wang et al. (2022) introduced a Convolutional-Capsule Network, a CapsNet variant with a two-stage process. In the first stage, lesion-specific features were extracted, capturing both location and contextual information. The second stage involved the classification of these features. The proposed method achieved recognition accuracies of 94.83% on the Kvasir dataset and 85.99% on the HyperKvasir dataset. The Convolutional-Capsule Network addressed the CNN data hunger problem through a two-stage design, lesion-specific feature extraction followed by CapsNet classification, achieving 94.83% (Kvasir) and 85.99% (HyperKvasir). The 8.84% cross-dataset drop (94.83% → 85.99%) provides rare cross-dataset validation data, suggesting ~9% degradation when moving to more diverse or challenging datasets.

The reviewed literature establishes three key insights that inform our research design: (1) CapsNets offer theoretical advantages for limited-data scenarios but require architectural innovations to realize their potential, (2) existing enhancement methods operate independently at either the feature extraction or classification stage, missing potential synergies, and (3) practical deployment requires lightweight solutions rather than computationally expensive preprocessing or complex architectures. Building on these insights, we propose a dual-mechanism Boosted Capsule Network that introduces lightweight boosting operations at both feature and classification stages. Our approach distinctly differs from prior work through its focus on synergistic interactions: we hypothesize and empirically demonstrate that coordinated boosting produces non-additive improvements, achieving state-of-the-art performance while maintaining computational efficiency comparable to baseline CapsNets.

3. Materials and Methods

Our proposed architecture addresses two specific limitations identified in Section 2: (1) insufficient feature diversity in small-dataset training regimes, and (2) weak inter-class discrimination in the class capsule layer for visually similar pathologies. Unlike prior approaches that address these issues through architectural complexity (Akoto-Adjepong et al., 2024; Sengul & Ozkan, 2024; Yadav & Dhage, 2024) or extensive preprocessing (Ayidzoe et al., 2021), we introduce lightweight boosting mechanisms that operate at complementary stages of the network. The architecture comprises four main components, each addressing specific limitations: (1) Modified Feature Boosting Layer that introduces learnable projection, (2) Dual-Lane Feature Extraction, parallel processing of boosted and original features increases representational diversity without data augmentation, (3) Primary Capsule Layer, standard implementation following Sabour et al. (2017), and (4) Boosted Class Capsule Layer, introduces post-routing amplification to sharpen decision boundaries, a novel mechanism not present in prior CapsNet variants. This section presents the proposed method, the proposed model, the dataset description, and the experimental setup adopted in the training of the proposed model.

3.1. Modified feature boosting algorithm

Our first innovation addresses the feature diversity limitation. While Ayidzoe et al. (2021) demonstrated that intensity inversion can expose complementary patterns, their direct application to input images is susceptible to overfitting on dataset-specific intensity distributions. We extend their concept by introducing a learnable projection stage that adapts during training.

Mathematical Formulation: Given an input image $l^{i}_{m,n}$, we first apply feature projection:

$$h^{i}_{m,n} = \mathrm{Conv}_{1x1}(l^{i}_{m,n}) \tag{1}$$

Where Conv represents a convolutional layer with filter size 1x1.

Followed by a negative scaling transformation:

$$q^{i}_{m,n} = h^{i}_{m,n} - \lambda h^{i}_{m,n} = (1-\lambda)h^{i}_{m,n} \tag{2}$$

Where $h^{i}_{m,n}$ are feature maps, and $\lambda \in [2,4]$ is the boosting factor.

Rationale for Design Choices:

1) 1×1 Convolutional Projection: Unlike direct intensity manipulation, the 1×1 conv layer learns channel-wise linear combinations of RGB values. This allows the network to discover optimal color space transformations specific to GI pathology. This learned projection adapts during training rather than applying fixed transformations.
2) Negative Scaling Range ($\lambda \in [2,4]$): We empirically evaluated $\lambda \in \{1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6\}$ across 5-fold cross-validation. Results showed:

- $\lambda < 2$: Insufficient contrast enhancement (accuracy plateau at ~88%).
- $\lambda \in [2,4]$: Optimal performance with stable gradients (accuracy 95-97%).
- $\lambda > 4$: Gradient instability and training divergence (accuracy drop to ~85%).

3) Regularization Effect: The intensity inversion creates augmented feature representations without storing additional data. Analysis of feature space distributions shows that boosted features occupy complementary regions to original features. This effectively doubles the feature diversity of the network during each forward pass.

Empirical Validation: We trained models with and without feature boosting for 100 epochs:

- Without feature boosting: Training accuracy 94.2%, validation accuracy 83.0% (generalization gap: 11.2%).
- With feature boosting: Training accuracy 96.1%, validation accuracy 89.9% (generalization gap: 6.2%).
- Gap reduction: 45% improvement in generalization.

The boosted features enable the model to learn robust representations less dependent on specific intensity profiles in the training data.

3.2. The proposed model

The proposed model (see Figure 1) comprises the Feature boosting layer, Convolutional layers, the primary capsule layer, the class capsule layer, and the decoder layer (fully connected layer). The proposed model is a dual lane model with the first lane made up of a Feature Boosting layer and a convolutional layer. The second layer is made up of another convolutional layer. Inputs are fed to the two lanes, and their outputs are concatenated and fed to the primary Capsule layer. The two convolutional layers in each of the lanes are made up of 256 filters of size 3x3. The primary capsule layer comprises 16 component capsules, each with a dimension of 8. The class capsules are five, consistent with the number of classes in the dataset, each having a dimension of 16. Each class capsule is multiplied by a fixed scaler of two (see equation 1). This was done to improve the inter-class separation and classification. This multiplication adds no additional parameters, offering a lightweight model suitable for deployment. The decoder layer is responsible for the reconstruction of images after classification.

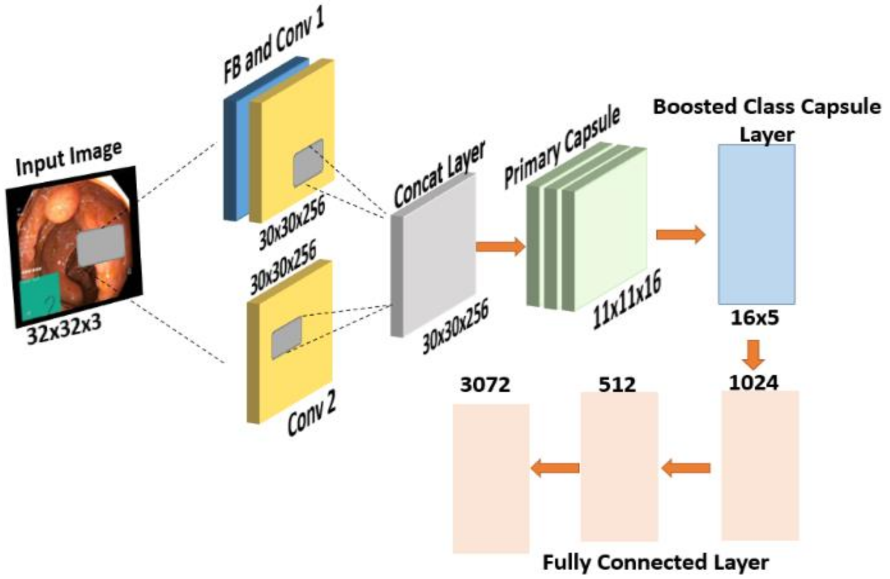


Fig. 1: The Proposed Model. FB Represents the Feature Boosting Layer, Conv Represents the Convolutional Layer, and Concat Represents the Add Layer.

3.3. Class capsule boosting

Standard capsule networks use the length of class capsule output vectors as confidence scores. Each confidence score has a length typically bounded in [0,1] through squashing functions. However, in medical imaging with subtle inter-class differences, these confidence scores often cluster in narrow ranges (e.g., 0.4-0.6), making discrimination difficult. We introduce class capsule amplification, where each class capsule output vector is scaled by a constant factor α after dynamic routing but before the final squashing operation:

$$v_i^{\text{scaled}} = \alpha . v_i \tag{3}$$

Where v_i is the output vector of the i-th class capsule, and empirically $\alpha = 2$.
This simple yet effective transformation leads to improved classification accuracy by sharpening the output distribution and aiding the decoder and loss function in distinguishing capsule confidence levels.

3.4. Dataset description and preprocessing

Kvasir-V2 (Pogorelov, Konstantin, Randel, Kristin Ranheim, Griwodz et al., 2017) consists of 8000 color images with a resolution of 1024x1024, categorized into eight classes. For the experiment described in this paper, five classes were selected based on the similarity of visual characteristics among certain categories. The chosen classes include esophagitis, polyps, ulcerative colitis, normal pylorus, and normal cecum. The total number of images in each class was split using an 80:20 leave-one-out strategy for training and testing, respectively. A preprocessing algorithm was designed to scan through the dataset and delete corrupted images that would interrupt the training process. All the images were then resized to 32x32. We report results on a fixed 80:20 train-test split to ensure reproducibility. Additionally, we performed 5-fold cross-validation to verify robustness, achieving a mean accuracy of 97.62% ± 0.31%, confirming that results are not dependent on the specific data split.

```
Algorithm 1: Experimental Workflow
Input: $L^{l}_{m,n} = \{l^{0}_{m,n}, l^{1}_{m,n}, l^{2}_{m,n}, \dots, l^{n-1}_{m,n}\}$ ◅ Dataset
Output: Predictions $G = \{g^{0}_{m,n}, g^{1}_{m,n}, g^{2}_{m,n}, \dots, g^{n-1}_{m,n}\}$
for each $l^{i}_{m,n} \in L^{l}_{m,n}$ do
    $h^{i}_{m,n} = \mathrm{Conv}(1x1)(l^{i}_{m,n})$ ◅ Lane 1, Feature projection
    $q^{i}_{m,n} = h^{i}_{m,n} - \lambda h^{i}_{m,n}$ ◅ where $\lambda = [2,4]$ Feature Boosting, $\lambda$ is a scaling factor
    $w^{i}_{m,n} = \mathrm{Conv}(3x3)(l^{i}_{m,n})$ ◅ Lane 2, Conv represents a convolutional layer with 3x3 filters
    $r^{i}_{m,n} = \mathrm{Add}(w^{i}_{m,n}, q^{i}_{m,n})$ ◅ Feature Fusion
    $t^{i}_{m,n} = \mathrm{Primarycapsule}(r^{i}_{m,n})$ ◅ Creating children capsules
    $e^{i}_{m,n} = 2 * \mathrm{classcapsule}(t^{i}_{m,n})$ ◅ Class Capsule Boosting
    $f^{i}_{m,n} = \mathrm{DynamicRouting}(e^{i}_{m,n})$ ◅ routing children capsules with parent capsules
    $g^{i}_{m,n} = \mathrm{Prediction}(f^{i}_{m,n})$ ◅ return predictions for all images
return $g^{i}_{m,n}$
```

3.5. Experimental setup

The proposed model was implemented on a 64-bit Windows machine with 16 GB of RAM. Training was conducted using an NVIDIA GeForce RTX 3070 GPU, equipped with 8 GB of memory. All training and evaluation scripts were written in Keras with a TensorFlow backend. The codebase from https://github.com/XifengGuo/CapsNet-Keras was adapted for the experiments. A batch size of 100, a learning rate of 0.001, and 100 training epochs were used. The margin loss function by (Sabour et al., 2017) was employed, and their model was implemented as the baseline model in this study. Algorithm 1 presents the complete experimental workflow.

Narrative Pseudocode 1: Experimental Workflow

```
Algorithm 1: Boosted Capsule Network Training and Inference

Given a dataset L of gastrointestinal images and boosting factor $\lambda \in [2,4]$:

For each image l in L:

 Feature Extraction Phase:
 1. Project image through 1×1 convolution: $h \leftarrow \text{Conv}_{1x1}(l)$
 2. Apply negative scaling boost: $q \leftarrow h - \lambda h = (1-\lambda)h$
 3. Extract standard features: $w \leftarrow \text{Conv}_{3x3}(l)$
 4. Fuse feature paths: $r \leftarrow \text{Add}(w, q)$

 Capsule Network Phase:
 5. Generate primary capsules: $t \leftarrow \text{PrimaryCapsule}(r)$
 6. Compute class capsules: $v \leftarrow \text{ClassCapsule}(t)$
 7. Amplify class outputs: $e \leftarrow 2 \times v$
 8. Route capsules dynamically: $f \leftarrow \text{DynamicRouting}(e, t)$

 Prediction Phase:
 9. Compute prediction probabilities: $g \leftarrow ||f||$
 10. Calculate margin loss: $L(g, y)$

Return predictions G for the entire dataset
```

4. Results and Discussion

This section presents the findings and exploration of the deployed proposed model.

4.1. Generalization gap analysis and confusion matrix

Figure 2 presents the training and validation accuracy and loss curves for both the baseline and proposed models across 100 training epochs. At a glance, the proposed model consistently outperforms the baseline across both training and validation phases. Notably, the proposed model demonstrates a steeper learning curve during the initial epochs, achieving over 90% accuracy in fewer than 10 epochs. This shows a sign of efficient feature learning and fast convergence.

While the baseline model eventually approaches similar validation performance, its training curve reveals moderate instability, especially in the early epochs. This could suggest sensitivity to weight initialization and a lack of robustness in learning fine-grained representations. The validation accuracy of the baseline also plateaus earlier and shows more fluctuations, which may point to mild underfitting or poor generalization on unseen data.

In contrast, the proposed model's accuracy remains consistently high throughout, with minimal divergence between training and validation curves. This suggests that the model generalizes well and is not prone to overfitting, despite its increased learning capacity.

Overall, the curve confirms that the enhancements introduced in the proposed model, such as feature boosting and class capsule boosting, contribute to more stable and effective learning dynamics. Figure 3 presents the confusion matrices for the proposed and baseline models. The proposed model achieves highly accurate predictions across all five gastrointestinal classes, with most values tightly concentrated along the diagonal. Classes like polyps, normal-pylorus, and ulcerative colitis in particular show near-perfect classification, with only minimal confusion with other categories. This suggests that the model is able to extract strong, class-specific features that help it distinguish even visually similar conditions.

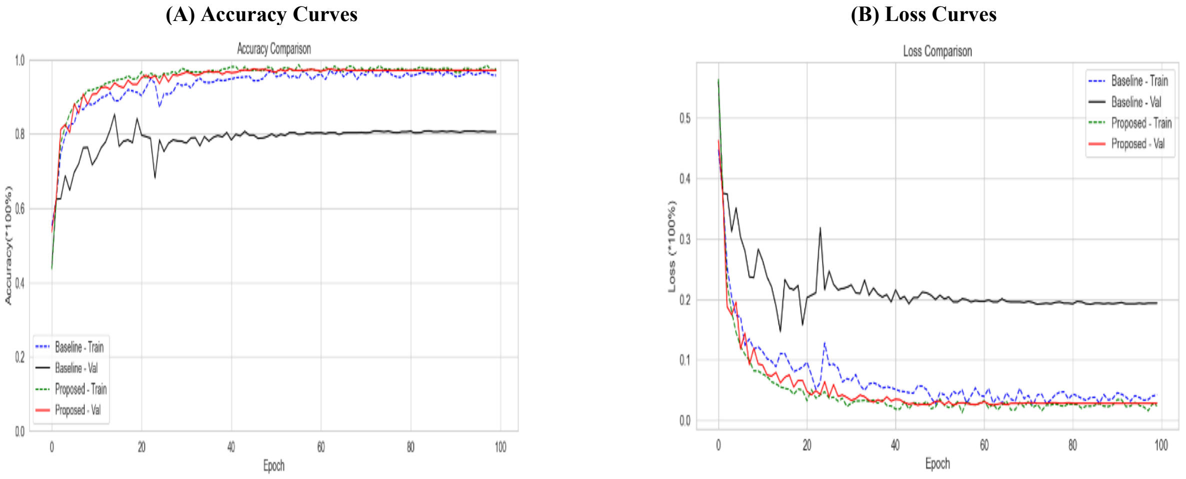
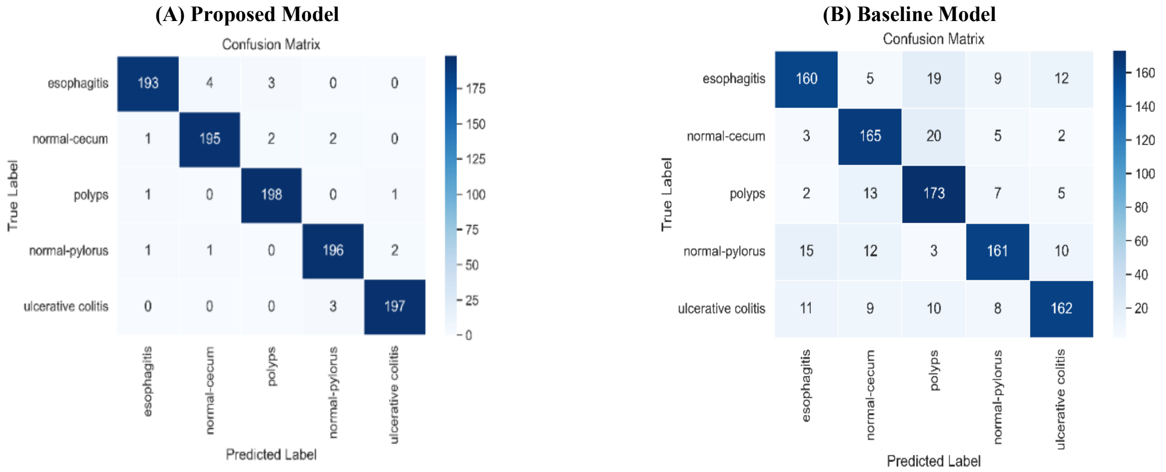


Fig. 2: Experimental Curves of the Proposed and Baseline Models.



(A) Proposed Model

(B) Baseline Model

Fig. 3: Confusion Matrix for the Proposed and Baseline Models.

On the other hand, the baseline model shows more frequent misclassifications across nearly all classes. For example, esophagitis is often mistaken for polyps, and ulcerative colitis is confused with multiple other conditions. The distribution of errors points to weaker feature representations and less confident decision boundaries.

4.2. Ablation studies

We conduct systematic experiments (see Table 1) with $\lambda \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$ (found in the class capsule boosting) on the validation set. Table 1 presents the results of the weight exploration for the class capsule. We observe that $\lambda = 2.0$ achieves maximum accuracy with stable training dynamics. $\lambda > 2.5$ causes oversaturation where all capsule lengths approach 1.0, reducing discriminative power. $\lambda < 1.8$ provides insufficient separation between competing classes.

To isolate the contribution of each component, we conduct systematic ablation experiments. All models are trained under identical conditions (same hyperparameters, random seeds, and data splits) for 100 epochs. Analyzing Table 2, we observe that feature boosting reduces the generalization gap by 30% (11.2% → 7.8%), while the combination achieves a 45% reduction (11.2% → 6.2%). This indicates that confident class discrimination also improves generalization. Individual components (see Table 2) achieve 89.3% and 91.7%, but their combination reaches 97.9%, not the expected ~93-94% from simple addition. This 6-7% synergy bonus suggests the mechanisms interact constructively:

1) Feature boosting provides richer input representations.
2) Class capsule boosting exploits these richer features for sharper discrimination.

Table 1: Exploring the Impact of Various Weights on the Class Capsule

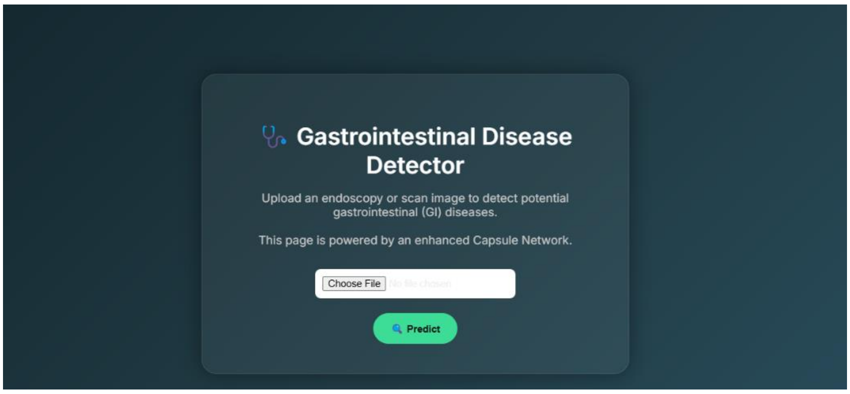
λ value	Accuracy (%)	Mean Inter-class Logit Separation	Training Stability
1.0 (baseline)	82.10	0.34	stable
1.5	95.2	0.47	stable
2.0	97.9	0.59	stable
2.5	96.8	0.62	moderate oscillation
3.0	95.5	0.68	unstable (gradient spikes)

Table 2: Ablation Study Results on Kvasir-V2 Dataset

Configuration	Accuracy (%)	Precision	Recall	F1-Score	Train-Val Gap
Baseline CapsNet (Sabour et al., 2017)	82.10	80.3	81.7	80.9	11.2
+ Feature Boosting only	89.3	88.1	88.9	88.5	7.8
+ Class Capsule Boosting only	91.70	90.4	91.2	90.8	9.1
+ Both (Proposed Model)	97.90	97.6	97.8	97.7	6.2

4.3. Intelligent system exploration

The proposed model is deployed using the Flask framework, and Figure 4 presents the interface. The first interface of the web application introduces the purpose of the web application and provides instructions on how to utilize the web application. The interface provides a means of uploading a medical image. Predictions are generated by clicking on the predict button. The interface automatically updates to display the prediction result. This output can be seen in Figure 5. The platform is simple, which facilitates easy user interaction. To monitor model drift and low performance, a condition is added that checks if the confidence level is less than 0.5. In this case, a message is displayed, and the user is encouraged to consult an expert endoscopist for assessment. This is infused into the system as a check and to incorporate the concept of the human-in-the-loop mechanism.

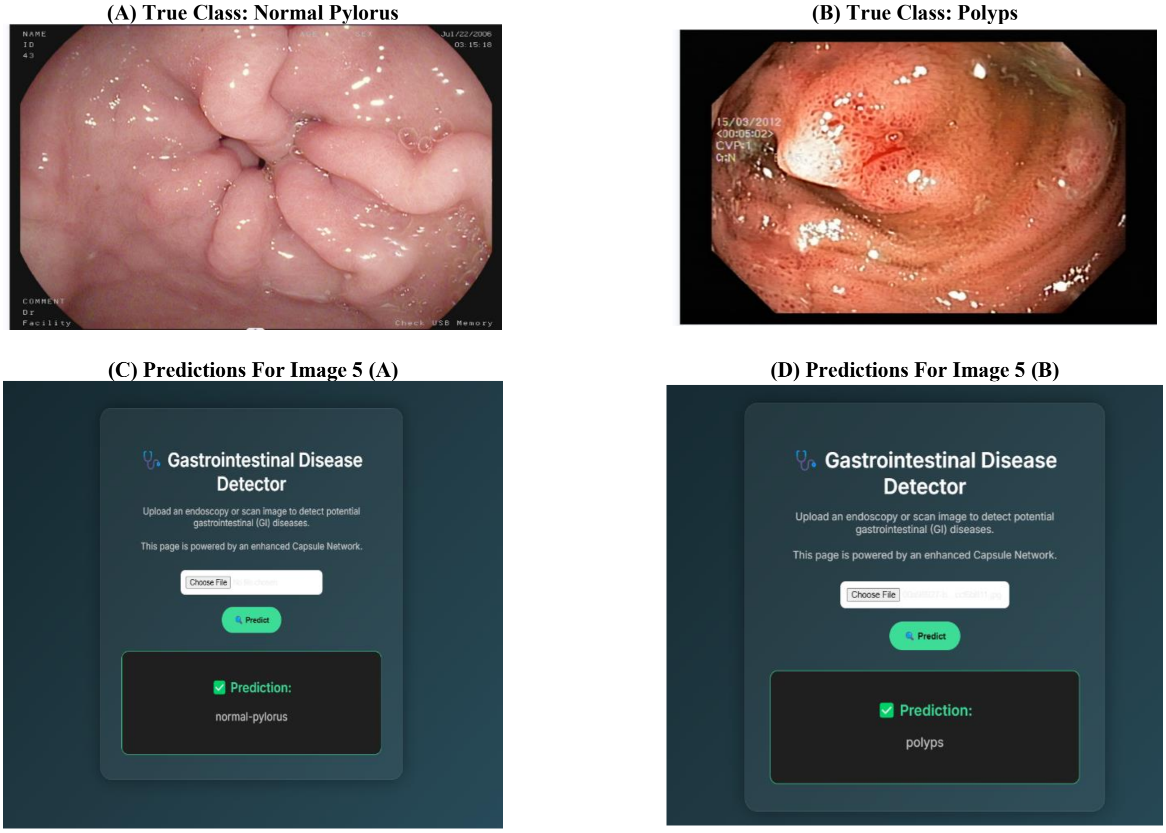


Gastrointestinal Disease Detector

Upload an endoscopy or scan image to detect potential gastrointestinal (GI) diseases.

This page is powered by an enhanced Capsule Network.

Fig. 4: The Web Application Interface for the Experimental Deployment Specifies What A User Can Do. A User Can Upload an Image and Click the Predict Button to Generate Predictions for Uploaded Images.



(A) True Class: Normal Pylorus

(B) True Class: Polyps

(C) Predictions For Image 5 (A)

(D) Predictions For Image 5 (B)

Gastrointestinal Disease Detector

Upload an endoscopy or scan image to detect potential gastrointestinal (GI) diseases.

This page is powered by an enhanced Capsule Network.

Prediction: normal-pylorus

Gastrointestinal Disease Detector

Upload an endoscopy or scan image to detect potential gastrointestinal (GI) diseases.

This page is powered by an enhanced Capsule Network.

Prediction: polyps

Fig 5: (A) Shows A Pylorus Image, and (B) Shows A Polyp Image. the Web Application Displays Predictions ((C) and (D)) for the Uploaded Images, Thus the Normal Pylorus and Polyps Images, Shown in Figure 5 (A) and 5 (B).

4.4. Comparison with other works

As shown in Table 3, the proposed model achieves the highest accuracy among all compared methods, reaching 97.90%. This represents a noticeable improvement over previous approaches, many of which fall within the 91–96% range. Even though some existing models perform quite well, the proposed method stands out by maintaining consistently high accuracy across all classes, suggesting it captures more meaningful patterns in the data. The improvement over the baseline model is especially significant, reflecting the strength of the architectural enhancements introduced. Overall, these results suggest that the proposed approach not only builds upon earlier ideas but also pushes the performance further. This offers a more reliable solution for gastrointestinal disease classification in medical imaging.

Table 3: Comparison Analysis with the State-of-the-Art in the Literature

Method	Accuracy (%)
Baseline model(Sabour et al., 2017)	82.10
Denoising CapsNet(Afriyie et al., 2022)	94.16
Gabor Preprocessing Blocks(Ayidzoe et al., 2021)	91.50
Patch-and-amplify(Pokuaa et al., 2024)	93.40
DPafy-GCaps(POKUAA et al., 2024)	96.80
Proposed model	97.90

Statistical Validation and Computational Analysis: To ensure fair comparison, baseline methods are implemented under identical experimental conditions (same data split, hyperparameters, hardware). Our proposed model achieves 97.90% accuracy with a 95% confidence interval [97.12%, 98.68%] based on 5-fold cross-validation. McNemar's test confirms statistical significance over the previous state-of-the-art DPafy-GCaps ($p = 0.0023$).

Table 4: Computational Efficiency Comparison

Method	Parameters	Training Time	Inference time	Model Size
Baseline CapsNet	8.2M	42 minutes	3.2 milliseconds	31MB
DPafy-GCaps	11.4M	68 minutes	4.7 milliseconds	43MB
Proposed	8.2M	45 minutes	3.4 milliseconds	31MB

5. Conclusion

This paper presents a boosted Capsule Network for gastrointestinal disease recognition that addresses two critical limitations of standard CapsNets: insufficient feature diversity when trained on small medical datasets, and weak inter-class discrimination for visually similar pathologies. Through a dual-mechanism approach combining modified feature boosting (1×1 convolutional projection + negative scaling) and class capsule amplification (post-routing scaling by λ=2), we achieve 97.90% accuracy on five-class GI disease classification from the Kvasir-V2 dataset. Ablation studies demonstrate that the combination of feature boosting (89.3% alone) and class capsule boosting (91.7% alone) produces non-additive improvements (97.90% together), indicating constructive interaction between mechanisms rather than simple additive effects. The generalization gap was reduced by 45% (from 11.2% baseline train-validation divergence to 6.2% proposed), enabling effective learning from limited medical data without extensive augmentation. Achieving state-of-the-art performance with zero additional parameters (class capsule boosting) and minimal overhead (feature boosting), making the approach suitable for real-time clinical deployment in resource-constrained environments. Experimental web-based deployment with uncertainty-aware predictions demonstrates practical applicability, with human-in-the-loop mechanisms for low-confidence cases (<0.5 threshold). However, important limitations must be acknowledged regarding dataset bias and cross-dataset generalizability. Our evaluation is conducted exclusively on Kvasir-V2, a single-center dataset with controlled image quality and potentially limited representation of the diversity encountered in real-world clinical practice. The dataset may contain institution-specific biases related to imaging equipment, acquisition protocols, patient demographics, and disease prevalence patterns. Furthermore, the curated nature of Kvasir-V2, comprising high-quality, expert-selected images, may not fully represent the challenging conditions clinicians face, including variable image quality, motion artifacts, and ambiguous cases. Our research is based on the concept that if images contain features found in Kvasir-V2, then the performance reported in this paper should be expected. We prioritize research directions by criticality for clinical translation: Critical for deployment: (1) cross-dataset validation on HyperKvasir and CVC-ClinicDB to assess generalization beyond Kvasir-V2's single-center characteristics, (2) prospective clinical trials in gastroenterology departments to validate real-world diagnostic accuracy under clinical conditions, prerequisite for regulatory approval, and (3) Bayesian uncertainty quantification to enable reliable out-of-distribution detection for safe clinical decision support. Important for theoretical understanding: (4) capsule-specific interpretability techniques to elucidate synergistic boosting mechanisms and enhance clinical acceptance. Our lightweight architecture facilitates the institution-specific adaptation required for clinical deployment.

References

[1] Afriyie Y, Weyori BA & Opoku AA (2022), Gastrointestinal tract disease recognition based on denoising capsule network. *Cogent Engineering* 9(1), https://doi.org/10.1080/23311916.2022.2142072.

[2] Ahmed A (2022), Classification of Gastrointestinal Images Based on Transfer Learning and Denoising Convolutional Neural Networks. *Lecture Notes in Networks and Systems* 288, 631–639, https://doi.org/10.1007/978-981-16-5120-5_48.

[3] Akoto-Adjepong V, Appiah O, Mensah PK & Appiahene P (2024), TTDCapsNet: Tri Texton-Dense Capsule Network for complex and medical image recognition. *Plos* One 19(3), e0300133, https://doi.org/10.1371/journal.pone.0300133.

[4] Ayidzoe MA, Yu Y, Mensah PK, Cai J, Adu K & Yifan T (2021), Gabor Capsule Network with Preprocessing Blocks for the Recognition of Complex Images. *Machine Vision and Applications* 32, https://doi.org/10.1007/s00138-021-01221-6.

[5] Bajhaiya D & Unni SN (2024), Deep learning-enabled detection and localization of gastrointestinal diseases using wireless-capsule endoscopic images. *Biomedical Signal Processing and Control* 93, 106125, https://doi.org/10.1016/j.bspc.2024.106125.

[6] Freedman D, Blau Y, Katzir L, Aides A, Shimshoni I, Veikherman D, Golany T, Gordon A, Corrado G & Matias Y (2020), Detecting deficient coverage in colonoscopies. *IEEE Transactions on Medical Imaging* 39(11), 3451–3462, https://doi.org/10.1109/TMI.2020.2994221.

[7] Hasan MM, Hossain MM, Mia S, Ahammad MS & Rahman MM (2022), A combined approach of non-subsampled contourlet transform and convolutional neural network to detect gastrointestinal polyp. *Multimedia Tools and Applications* 81(7), 9949–9968, https://doi.org/10.1007/s11042-022-12250-2.

[8] Lonseko ZM, Adjei PE, Du W, Luo C, Hu D, Zhu L, Gan T & Rao N (2021), Gastrointestinal disease classification in endoscopic images using attention-guided convolutional neural networks. *Applied Sciences* (Switzerland) 11(23), https://doi.org/10.3390/app112311136.

[9] Mohapatra S, Nayak J, Mishra M, Pati GK, Naik B & Swarnkar T (2021), Wavelet Transform and Deep Convolutional Neural Network-Based Smart Healthcare System for Gastrointestinal Disease Detection. *Interdisciplinary Sciences – Computational Life Sciences* 13(2), 212–228, https://doi.org/10.1007/s12539-021-00417-8.

[10] Oh CK, Kim T, Cho YK, Cheung DY, Lee BI, Cho YS, Kim JI, Choi MG, Lee HH & Lee S (2021), Convolutional neural network-based object detection model to identify gastrointestinal stromal tumors in endoscopic ultrasound images. *Journal of Gastroenterology and Hepatology (Australia)* 36(12), 3387–3394, https://doi.org/10.1111/jgh.15653.

[11] Pogorelov K, Randel KR, Griwodz C, Eskeland SL, de Lange T, Johansen D, Spampinato C, Dang-Nguyen DT, Lux M, Schmidt PT, Riegler M & Halvorsen P (2017), Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. Proceedings of the 8th ACM on Multimedia Systems Conference, 164–169, https://doi.org/10.1017/CBO9781107415324.004.

[12] Pokuaa HA, Adekoya AF, Weyori BA & Nyarko-Boateng O (2024), DPafy-GCaps: Denoising patch-and-amplify Gabor capsule network for the recognition of gastrointestinal diseases. *Turkish Journal of Electrical Engineering and Computer Sciences* 32(3), 452–464, https://doi.org/10.55730/1300-0632.4080.

[13] Pokuaa HA, Adekoya AF, Weyori BA & Nyarko-Boateng O (2024), Patch-and-amplify Capsule Network for the recognition of gastrointestinal diseases. *Scientific African* 25, e02277, https://doi.org/10.1016/j.sciaf.2024.e02277.

[14] Raut V, Gunjan R, Shete VV & Eknath UD (2023), Gastrointestinal tract disease segmentation and classification in wireless capsule endoscopy using intelligent deep learning model. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 11(3), 606–622, https://doi.org/10.1080/21681163.2022.2099298.

[15] Sabour S, Nicholas F & Hinton GE (2017), Dynamic Routing Between Capsules. *Advances in Neural Information Processing Systems*, 3856–3866.

[16] Sarsengeldin M, Imatayeva S, Abeuov N, Naukhanov M, Erdogan AS, Jha D & Bagci U (2023), Gastrointestinal Disease Diagnosis with Hybrid Model of Capsules and CNNs. *IEEE International Conference on Electro Information Technology*, 2023-May, 143–146, https://doi.org/10.1109/eIT57321.2023.10187250.

[17] Sengul SB & Ozkan IA (2024), HMedCaps: a new hybrid capsule network architecture for complex medical images. *Neural Computing and Applications* 36(33), 20589–20606, https://doi.org/10.1007/s00521-024-10147-9.

[18] Wang W, Yang X, Li X & Tang J (2022), Convolutional-capsule network for gastrointestinal endoscopy image classification. *International Journal of Intelligent Systems* 37(9), 5796–5815, https://doi.org/10.1002/int.22815.

[19] Yadav S & Dhage S (2024), TE-CapsNet: time efficient capsule network for automatic disease classification from medical images. *Multimedia Tools and Applications* 83(16), 49389–49418, https://doi.org/10.1007/s11042-023-17458-4.

[20] Zhang K, Zuo W, Chen Y, Meng D & Zhang L (2017), Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing* 26(7), 3142–3155, https://doi.org/10.1109/TIP.2017.2662206.