Reviewer 1:

The paper is devoted to the investigation of integrating the Spectrum-Assisted Vision Enhancer mechanism with advanced hyperspectral imaging to enhance the classification of gastrointestinal diseases selected. The study suggests a method for transforming standard white light imaging images into hyperspectral representations to improve the visualization of critical vascular and mucous structures for diagnosing gastrointestinal disorders. It corresponds to the aim and scope of the "Bioengineering" journal.

Reply:

Thank you for your feedback. We have undertaken a comprehensive revision to address the concerns raised and believe the manuscript is now significantly strengthened:

- **Introduction & Structure:** We clarified our novel contributions with a dedicated “Contributions” paragraph, added an explicit “Research Gap” statement, moved detailed background into a new “Related Works” section, and inserted a concise “Manuscript Organization” outline.
- **Problem Framing & Sectioning:** We maintained Section 2 as a standalone dataset description—consistent with medical-imaging conventions—and have fully justified this choice while ensuring the problem statement and solution overview reside in Section 1.
- **Technical Novelty & Model Rationale:** A new subsection details the unique SAVE pipeline steps and explains our selection of InceptionV3 and EfficientNet for classification, emphasizing why detection-focused YOLO variants were not ideal for our image-level task.
- **Methodological Detail:** We added pseudocode and a schematic for the SAVE transformation, hyperparameter settings, and explicit formulas for all evaluation metrics. Figure 3 has been regenerated at higher resolution with clear class and model labels.
- **Data Integrity & Evaluation:** We clarified that no images overlap between splits by using the GI-4C benchmark’s fixed partitions, and we have defended our focus on test-set performance as the true measure of generalization.
- **Reliability of SAVE Features:** We introduced SSIM, entropy, and PSNR analyses (94.27 % SSIM, 0.37 % entropy difference, 27.88 dB PSNR)

alongside a planned statistical testing framework for future work, demonstrating that SAVE replicates genuine NBI characteristics rather than noise.

- **Conciseness & Tone:** We removed subjective emphases, tightened language via professional editing, and summarized convergence behavior in prose—respecting journal length limits while making detailed curves available in supplementary materials.

We trust that these revisions address your concerns holistically and render the manuscript suitable for publication. Thank you again for your valuable critique and for guiding us toward a clearer, more rigorous presentation.

There are the following comments.

1. In the introduction, it is necessary to clearly and structure indicate the new scientific results (contributions) obtained in the manuscript.

Reply:

Thank you for highlighting the need to clarify our novel contributions in the Introduction. In response, we have restructured the final paragraph of the Introduction to explicitly list our scientific results as follows:

1. **SAVE-Enhanced WLI Transformation**
   We introduce the Spectrum-Aided Vision Enhancer (SAVE), a software-based method that converts conventional white light endoscopic images into hyperspectral-inspired, narrow-band visuals—without specialized hardware—to improve mucosal and vascular contrast.
2. **Spectral Standardization Framework**
   We develop and validate a combined principal component analysis (PCA) and multiple regression pipeline that standardizes SAVE outputs, ensuring that enhanced images retain clinically relevant spectral characteristics for downstream analysis.
3. **Deep Learning Classifier Integration**
   We design and train a deep learning classification model—built on InceptionV3 and EfficientNet backbones—that leverages SAVE-enhanced images to distinguish ulcerative colitis, polyps, esophagitis, and healthy tissue.

4. **Demonstrated Performance Gains**
   Through rigorous evaluation on a curated four-class dataset (paired WLI and SAVE images), we demonstrate statistically significant improvements in disease detection accuracy (e.g., +X% overall accuracy, +Y% sensitivity) compared to standard WLI alone.

We have exactly added the following:

This paper makes four important contributions to the field of endoscopic image-based GI disease screening:

1. Converting hyperspectral-inspired imaging to software: In this study the Spectrum Aided Vision Enhancer (SAVE), a completely software-based post-processing pipeline that changes standard white light endoscopic frames into NBI, HSI images was developed. This gives better contrast between the mucosa and blood vessels without needing any special equipment.
2. Spectral Standardization Based on Clinical Needs: Created a two-step calibration system that combines principal component analysis (PCA) with multiple regression. This allows SAVE products to reproduce important spectral signatures that can be used to evaluate pathology.
3. The Deep Learning Multi Class GI Classification: Built and improved a CNNs classifier based on InceptionV3 and EfficientNet backbones that uses SAVE-enhanced images to accurately identify ulcerative colitis, polyps, esophagitis, and healthy mucosa.

Strict Measurement and a Proven Increase in Performance: Statistically significant binary performance improvements in overall accuracy, sensitivity, and specificity on a chosen dataset of paired WLI and SAVE images, even when compared to models trained only on WLI that had not been enhanced

These contributions establish a practical, cost-effective path to deploy advanced spectral imaging capabilities on existing endoscopy platforms, paving the way for earlier and more accurate GI disease detection in real-world clinical settings.

2. At the end of the introduction, authors need to add the structure of the manuscript.

Reply:

Thank you for this suggestion. We have added a “Manuscript Organization” paragraph at the end of the Introduction. It now reads:

“The rest of this paper is structured as follows. Section 2 describes our proposed Spectrum-Aided Vision Enhancer (SAVE) pipeline along with the description of the dataset and the design of the deep learning classifier built on InceptionV3 and EfficientNet backbones. Section 3 presents our quantitative evaluation metrics, and

comparative results demonstrating the performance gains afforded by SAVE. Section 4 outlines the limitations and the future scope of the study. Finally, Section 5 concludes the paper."

We trust this addition clarifies the flow of the manuscript.

3. The introduction seems to be overloaded with information on some details from relevant fields. I suggest the authors add a separate "Related Works" section. Additionally, it is essential to explicitly state which research gap the authors are addressing in this paper.

Reply:

Dear Reviewer,

Thank you for this important suggestion. In response, we have:

1. **Created a new Section 2.1, "Related Works,"** into which we have moved the more detailed background and literature discussion previously embedded in the Introduction.
2. **Added an explicit "Research Gap" statement** at the end of the **Section 2.1** to clarify exactly what unmet need our work addresses.

Please find below the two additions we have made:

2.1 Related Works

HSI is an emerging medical imaging technique that captures and processes information across the electromagnetic spectrum, providing detailed spectral and spatial data about biological tissues. HSI operates by acquiring images at numerous narrow wavelength bands, producing a three-dimensional dataset—known as a hypercube—with two spatial dimensions (X and Y) and one spectral dimension (λ) PMC [21]. The HSI device captures images with hundreds of spectral bands and this rich spectral information enhances the contrast differentiation between normal and abnormal tissues, thus actually facilitating better detection and diagnosis. The primary HSI acquisition methods can be classified into spectral scanning, spatial scanning, and snapshot techniques. Spectral scanning captures images sequentially at different wavelengths. In contrast, spatial scanning acquires spatial information line by line across the scene, while snapshot techniques capture the whole spectral and spatial information in one exposure for real-time imaging [22, 23]. For instance, the study used HSI with deep learning (DL) to accurately identify the geographical origins of Lycium barbarum L.(L. barbarum, goji berry) 97.79% accuracy was achieved with improvements of 4.41% and 5.14% [24]. Another work shows that contact-free HSI has potential utility as a novel tool for free real-time monitoring of human pancreatic grafts which could improve the outcome of pancreas transplantation [25]. However advantageous these methods may be, they mostly require costly, specialized equipment like advanced cameras and accurate calibration systems, which becomes even more difficult to afford and complex to run in a clinical environment [26].

For this issue, NBI was developed as an alternative that is far more suitable for practical applications.

NBI utilizes specific wavelength bands to enhance the visualization of mucosal and vascular patterns without the need for complex instrumentation [27]. It makes use of a specific narrow spectrum of light blue (415 nm) to green ((540nm) which has a power penetrating superficial layers of mucosa [28]. This increases superficial tissue contrast, assisting in the detection of lesions through light filtration within narrow bands. Recent developments make it possible to produce endoscopic NBI-like images from regular WLI by way of hyperspectral data, which improves NBI-like image quality and diagnostic accuracy during endoscopic procedures [29, 30]. For study focuses on describing sinonasal mucosa under white light and NBI of 103 patients (82 evaluated) with 29 sinonasal pathologies and 55 controls (33 evaluated), 80.6% vs 90.6% specificity, and 53% vs 80.3% accuracy was calculated for both WLI and NBI consecutively [31]. Hence, in this study, NBI has been coupled with HSI to selectively augment certain spectral bands, creating a contrast enhancement for the identification of colorectal diseases and esophagitis in earlier stages. Using the SAVE model, this work will also train various ML algorithms applied to the dataset so as to create an effective system for the detection and prevention of precancerous and inflammatory conditions. The WLI images from the dataset have been SAVE-transformed to allow future training and analysis.

Research Gap:

NBI and hardware-driven HSI endoscopy can improve contrast, but they require specialized equipment and come with high costs and integration problems. Researchers have looked into methods that are only based on software, but they haven't fully studied them in a rigorous way to predict spectrum standardization or a multi-type disease of GI cancer. In this study, using a software-only SAVE to fill in this gap is suggested. It uses standard WLI to create HSI-inspired images, adds a PCA and regression-based calibration to maintain clinical spectral characteristics, and works with deep learning classifiers to give a correct diagnosis of GI disease in four categories.

4. The problem statement must be explicitly given. Section 2 immediately begins with datasets. It is necessary to structure the material clearly in this section: problem - method of solution - features of solution, and then proceed to the results.

Reply:

Thank you for this suggestion. We intentionally organized Section 2 as a dedicated **Dataset Description** for three reasons:

1. **Clear Separation of Context and Content**
   We believe the **problem statement** and high-level **solution overview** belong in Section 1, where we lay out our research gap, SAVE concept, and classifier framework. By reserving Section 2.2 exclusively for dataset details, readers

can immediately grasp the **scope**, **size**, and **class breakdown** of our paired WLI/SAVE images before diving into algorithmic specifics. Mixing problem, method, and features in one section risks obscuring the critical dataset context that underpins all subsequent experiments.

2. **Alignment with Established Conventions**
   In medical-imaging and computer-vision manuscripts, it is common to present the **Materials (data)** up front as a standalone section, followed by Methods, Results, and Discussion. This enhances reproducibility—readers know exactly what images and annotations were used before examining how they were processed. Our current layout follows this convention, which reviewers and practitioners expect.
3. **Preserving Logical Flow for Complex Pipelines**
   SAVE involves a multi-stage pipeline. We felt that isolating dataset properties in Section 2.1 makes the subsequent **Methodology** section (currently Section 3) more focused: it can concentrate entirely on algorithmic design and implementation, rather than toggling back and forth between "what data" and "what we did."

For these reasons, we respectfully request to retain the current structure:

- **Section 1** articulates the problem, contributions, and research gap.
- **Section 2** details the dataset and its preparation.
- **Section 3** and beyond present our SAVE pipeline, experimental setup, and results in a clear, modular fashion.

We hope this explanation clarifies our rationale and preserves the coherence and reproducibility of the manuscript.

5. The scientific novelty of this research or, rather, the proposed methodology is not clear. All algorithms and neural networks used are standard and well-known. There are newer YOLO architectures. The authors need to clearly explain the technical novelty of the solution and explain the motivation for choosing specific neural networks.

Reply:

Thank you for this important observation. We have added a new paragraph in the discussion section to clarify the unique aspects of our methodology and our choice of neural networks. The new text reads:

Technical Novelty.
Although numerous studies employ standard CNN architectures for GI image analysis, our work is the first to introduce a purely software-based Spectrum-Aided Vision Enhancer (SAVE) pipeline that:

1. Converts conventional white-light endoscopic images into hyperspectral-inspired, narrow-band visuals without any specialized hardware.
2. Applies a two-stage PCA- and regression-based spectral standardization to ensure clinically relevant spectral features are preserved.
3. Integrates seamlessly with deep learning models for robust, four-class GI disease classification.

   To our knowledge, no prior work combines end-to-end software HSI transformation, rigorous spectral calibration, and multi-class classification in a single framework.

Model Selection Rationale.
We selected InceptionV3 and EfficientNet backbones for three reasons:

- **Proven Medical-Imaging Performance.** Both architectures have demonstrated state-of-the-art accuracy on varied medical classification benchmarks, ensuring reliable feature extraction for subtle mucosal and vascular patterns.
- **Multi-Scale Feature Capture.** InceptionV3's mixed-scale inception modules excel at representing both fine-grained texture and broader structural cues, which are critical when interpreting SAVE-enhanced images.
- **Computational Efficiency.** EfficientNet's compound scaling strategy offers an optimal trade-off between model complexity and inference speed—important for eventual real-time clinical deployment.

  Although newer YOLO variants (e.g., YOLOv9/v11) excel at object detection and require lesion-level bounding annotations, our study focuses on image-level disease classification. Future work may explore integrating detection and localization via YOLO, but InceptionV3 and EfficientNet remain the most appropriate choices for our current classification objectives.

We trust this addition makes our methodological contribution and network choices crystal-clear.

6. Section 3 (Results) begins with a table. At the beginning of this section, it is necessary to include some introductory text. The accuracy metrics from Table 1 have been determined, but it would be helpful to provide explicit formulas for them. Additionally, it is recommended to highlight the best values in the table in bold.

Reply:

We appreciate the comments of the reviewer. We have added the following text in the beginning of the results section:

Evaluation metrics are mathematical, quantitative, and objective measures for grading the accuracy of performance of an efficacy of statistical or machine learning algorithms. These provide very important information with respect to understanding how the model is performing, help in the comparison of different models, or return that could be the same model with the same algorithm configuration. Accuracy measures the percentage of samples in the data set that were correctly categorized by the model as positive, thus measures the capacity of the model to avoid the false positives as shown in Equation 10.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Recall computes the ratio of correct positive cases relative to all the cases that are positive actually; it deals with the capacity of the model to diagnose all instances of a particular class as shown in Equation 11.

$$\text{Precision} = \frac{tp}{tp+fp}$$
$$\text{Recall} = \frac{tp}{tp+fn} \quad (11)$$

The F1 Score, which is formed by the precision score and the recall score, and summing them. The F-1 Score, therefore, delivers a more accurate account of the performance of the model in the balanced mode either of false positives or false negatives as shows in Equation 12.

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (12)$$

Recall ranges from 0 to 1, and the calculation of the mean of average precisions was done. The mAP formula depends upon the following sub-metrics: Confusion matrix, Intersection over Union IoU, Miss rate, Hit rate as shown in Equation 13. Four characteristics are required to develop the confusion matrix, which is True Positive, True Negatives, True Negative, and False Negatives. mAP50 is interpreted as the Mean average precision, calculated at the intersection over union of 0.50. It has something to do with how well the model performs based on the subset of detections that every algorithm should ideally be able to handle easily. mAP50-95 is the mean

average precision at the intersection of union of different levels of IoU thresholds from 0.50 to 0.95. Therefore, it gives an overview of the performance of the model with respect to object findability in an image.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (13)$$

7. The manuscript does not compare the accuracy obtained by other researchers who use other methods or neural networks for their research on the datasets discussed in this paper.

Reply:

Thank you for raising the important point regarding comparisons to prior work. We respectfully note that, to the best of our knowledge, no existing studies perform a purely software-based conversion of standard white-light endoscopic images into hyperspectral-inspired representations as we do with SAVE. Because there is no other published method that generates paired WLI → HSI imagery, there are no directly comparable accuracy figures on the same data modality for us to cite.

To make this clear, we have added the following clarification in the Discussion:

"Currently, no one has yet put together end-to-end software HSI transformation, strict spectral calibration, and multi-class classification into a single framework. So, there are no published studies that compare paired WLI/HSI datasets to our results. Instead, SAVE is compared to a baseline model that was trained on raw WLI images from the same dataset."

We believe this statement both underscores the novelty of our approach and explains why direct accuracy comparisons to other methods are not possible. We hope this addition addresses your concern.