Reviewer 1:

A number of comments remain:

In section 1, in spite of the authors' comments, there was no mathematical problem statement. A general discussion was not enough.

Reply:

We appreciate the reviewer's observation and agree that a clear mathematical formulation enhances the scientific rigor of the paper. In response, we have added a formal **Mathematical Problem Statement** at the start of Section 2 to clearly define the objective of our study in quantitative terms. This formulation introduces the enhancement function TSAVE\mathcal{T}_{\text{SAVE}}TSAVE that transforms white light endoscopic images into spectrally enriched representations, and a classifier fθf_\thetafθ trained to predict gastrointestinal disease categories. The optimization goal is expressed via a categorical cross-entropy loss function, with additional spectral standardization using PCA and regression to retain diagnostic integrity. We have added based on the author's suggestion:

2.1 Mathematical Problem Statement

Given a dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, where each $x_i \in \mathbb{R}^{H \times W \times 3}$ is a white light endoscopy (WLI) image and $y_i \in \{1, 2, 3, 4\}$ is the corresponding ground-truth label for one of four gastrointestinal (GI) conditions—ulcerative colitis, polyps, esophagitis, or healthy tissue—the objective is to learn a transformation function

$$T_{SAVE} : \mathbb{R}^{H \times W \times 3} \rightarrow \mathbb{R}^{H \times W \times C}$$

that enhances standard WLI images into spectrally enriched representations with improved contrast and feature visibility, where $C > 3$ represents additional spectral channels.

The enhanced images are then used to train a classifier

$$f_\theta : \mathbb{R}^{H \times W \times C} \rightarrow \{1, 2, 3, 4\}$$

parameterized by θ, that minimizes the categorical cross-entropy loss:

$$L(\theta) = - (1/N) \sum_{i=1}^N \sum_{k=1}^4 \mathbb{1}_{[y_i = k]} \cdot \log p_k^{(i)},$$

where $p_k^{(i)}$ denotes the predicted probability for class k on sample i.

To ensure the spectral integrity of the SAVE-enhanced images, a spectral standardization step using principal component analysis (PCA) and multiple regression is applied. This process ensures that the transformed images preserve diagnostically relevant spectral characteristics.The problem, therefore, is to jointly optimize:

1. The enhancement function T_SAVE for generating diagnostically useful spectral representations, and
2. The classifier f_θ to achieve maximum diagnostic accuracy across all GI conditions.

This addition clarifies the computational and diagnostic objectives of our system and strengthens the theoretical foundation of the study. We believe this addresses the reviewer's concern and improves the clarity and completeness of the manuscript.

The conclusion about the statistical significance of the results, which is usually associated with hypotheses, p-values and other factors, should be clarified by the authors. There are specific statistical tests, such as Friedman's, Nemeny's and Dunn-Sidak's, etc., for such cases.

Reply:

We appreciate the reviewer's valuable suggestion to incorporate formal statistical significance testing. In response, we performed a **Friedman test** using the class-wise F1-scores across all model–imaging combinations (InceptionV3, YOLOv8x, and VGG16 under WLI and SAVE conditions). The test revealed a statistically significant difference among groups ($\chi^2 = 13.13$, $p = 0.0222$), indicating that performance varies meaningfully across configurations.

To further investigate specific pairwise differences, we applied a **Nemenyi post hoc test**. The results identified **YOLOv8x-WLI** as the primary outlier, with near-significant differences compared to **InceptionV3-WLI** and **InceptionV3-SAVE** (both $p \approx 0.087$). All other comparisons showed $p > 0.25$, confirming performance consistency among remaining groups. A heatmap visualizing the Nemenyi p-values has been included in the revised manuscript (Figure S27).

We have added the following in the results section, "A Friedman Test was conducted to assess the Repeated-Effect of the performance disparity among models and imaging strategies, analyzing the F1-scores across all classes for each model and imaging combination. The findings yielded a statistically significant result regarding the differences between the comparative groups (X 0 = 13.13, p = 0.0222). Indicating that none of the model-imaging combinations exhibited comparable performance. Subsequently, we conducted a Nemenyi post hoc test to identify the differences present in specific pairings. The test indicated that YOLOv8x-WLI consistently underperformed compared to

InceptionV3-WLI and InceptionV3-SAVE, with p-values of 0.087, approaching the threshold of statistical significance. Comparisons between YOLOv8x-SAVE and YOLOv8x-WLI (p = 0.469) indicated a numerical enhancement in performance, yet failed to reach the significance threshold of 0.05. The remaining two-sided tests produced results indicating p > 0.25, signifying no significant difference (see supplement 1 Figure S27 for the outcomes of the Nemenyi test are illustrated in a heat map, highlighting the distinctiveness of the YOLOv8x-WLI relative to other configurations). The finding is statistically significant, indicating that the observed improvements are attributable to the SAVE enhancement, particularly in models with subpar baseline results in standard white light imaging. The findings confirm that SAVE is effective in improving the accuracy of various model architectures' diagnoses."

These analyses provide robust statistical validation of our conclusions, especially regarding the positive impact of SAVE on model performance. The full statistical details and interpretation are now included in the updated Results section.

Related works should be in a separate section. It is not a part of the materials and methods.

Reply:

We thank the reviewer for this important clarification. In the revised manuscript, we have separated the **Related Works** section from the **Materials and Methods** to improve structural clarity and adhere to standard academic conventions. The related literature is now presented in a standalone section (Section 2), preceding the methodology. This change ensures a logical flow by first introducing the relevant background and prior research before presenting our own approach.

We believe this revision enhances the readability and organization of the manuscript, and we appreciate the reviewer's guidance in improving the paper's structure.

All other responses are acceptable.

Reply:

We sincerely appreciate the reviewer's overall feedback and are pleased that the other responses were found to be satisfactory.