INFLAMMATION DETECTION USING ENSEMBLE ENDOSCOPIC MULTIMODAL ASSESSMENT IN INFLAMMATORY BOWEL DISEASE

Bisi Bode Kolawole[1] *Ujwala Chaudhari*[1] *Giovanni Santacroce*[2] *Irene Zammarchi*[2]
Rocio Del Amor[4] *Pablo Meseguer*[4] *Andrea Buda* [3] *Raf Bisschops* [5] *Valery Naranjo*[4]
Subrata Ghosh[2] *Marietta Iacucci*[2] *Enrico Grisan*[1] *PICaSSO Group*[⋆]

[1] School of Engineering, London South Bank University, London, UK

[2]APC Microbiome Ireland, College of Medicine and Health, University College of Cork, Cork, Ireland
[3]Division of Gastrointestinal Oncological Surgery, Santa Maria del Prato, Feltre (Italy)
[4] Instituto de Investigación e Innovación en Bioingeniería, HUMAN-tech,
Universitat Politècnica de València, València, Spain

[5] Division of Gastroenterology, University Hospitals Leuven, Leuven, Belgium

ABSTRACT

Inflammatory bowel diseases (IBD), comprising Crohn's disease (CD) and ulcerative colitis (UC), present chronic inflammatory gastrointestinal disorders with substantial implications for patients' quality of life. Traditional endoscopic evaluation remain pivotal for monitoring and managing IBD. Recent advancements in Virtual Chromoendoscopy (VCE) technologies, such as Flexible Spectral Imaging Color Enhancement (FICE) and iScan with digital enhancement, offer noninvasive alternatives for evaluating gastrointestinal diseases. While overcoming some limitations of White Light Endoscopy (WLE), these technologies introduce challenges related to scoring systems and deep learning algorithm training due to the qualitative nature of existing endoscopic scores. To address these challenges, we propose a combination of a generative (cycleGAN) and an ensemble model that integrates assessments from white light endoscopy (WLE), and generated Virtual Chromoendoscopy (VCE) to enhance inflammation detection and prediction. The ensemble model aims to combine the strengths of diverse modalities, providing a holistic understanding of a patient's inflammation status. Experiments demonstrated in this paper show that by integrating endoscopic findings with other modalities using an ensemble learning method can greatly improve the accuracy of prediction of IBD.

Index Terms— Multiple instance learning, Virtual Chromoendoscopy (VCE), White Light Endoscopy (WLE), Endoscopy enhancement, Esemble learning.

1. INTRODUCTION

Inflammatory bowel diseases (IBD), such as Crohn's disease (CD) and ulcerative colitis (UC), are chronic inflammatory gastrointestinal disorders that significantly impact the quality of life and general well-being of affected individuals [1]. Monitoring IBD severity and remission traditionally involves endoscopy, which is essential for patient management and personalized therapeutic plans. However, adherence to the treat-to-target strategy in clinical management necessitates reliable endoscopic evaluations. Despite dedicated training, endoscopic assessments suffer from high interobserver variability and low reproducibility, even among experts.

The advent of Virtual Chromoendoscopy (VCE) technologies such as iScan with digital enhancement by PENTAX, has revolutionized the evaluation of IBD. These technologies address the limitations of White Light Endoscopy (WLE) and are currently employed for various diagnostic tasks related to gastrointestinal diseases [2]. Recently, more and more endoscopy image enhancement tools have been proposed, especially with the use of deep learning by exploiting contrast fusion [3], whereas generative approaches StillGAN was proposed for improving the general appearance of the images [4]. Other approach targets the visibility of specific tissues and abnormalities to generate the enhanced endoscopy data using GAN [5] for improving polyps detection.

To enhance endoscopy evaluation further, other deep learning models were proposed for endoscopy evaluation [6]. These systems, which utilize either VCE or WLE, aim to standardize and simplify evaluations, thereby reducing discrepancies among endoscopists and pathologists.

However, despite the advance in research, existing scoring systems for inflamation grading face limitations due to interobserver variation and the qualitative nature of grading

⋆ Pradeep Bhandari, Gert de Hertogh, Jose G. Ferraz, Martin Goetz, Xianyong Gui, Bu'Hussain Hayee, Ralf Kiesslich, Chiara Metelli, Mark Lazarev, Remo Panaccione, Adolfo Parra-Blanco, Luca Pastorelli, Timo Rath, Elin Synnøve Røyset, Michael Vieth, Vincenzo Villanacci, Davide Zardo.

scores, making it challenging to train these deep learning systems.

To address this challenge, [7] developed a new electronic virtual chromoendoscopy (EVC) score called PICaSSO, capable of reflecting mucosal and vascular changes, including mucosal healing in UC. This innovative approach separates the assessment of mucosal and vascular patterns, translating into a computer-based system for detecting endoscopic remission and inflammation.

In recent studies combining different models to achieve the same objective have been seen to greatly improve the performace of deep learning models on various task. Based on our current knwoledge combining WLE with generated VLE with ensemble learning has not being fully exploited. In this research we aim to generate VCE(iScan2, and iScan3) with a generative model (cycleGAN) and use the ensemble learning method to combine and evaluate this based on the degree of inflamation. By leveraging on the strengths of these modalities the model will be able to provide a more nuanced understanding of the patient's inflammation status by integrating information from these assessments.

2. METHODOLOGY

Given a data of single modality (WLE), we exploit the use of cycleGAN for the conversion of this data into different modalities, VCE (iScan2 and iScan3). This is to create a robust model on the prediction of inflamation in the medical data. We exploit the use of a CNN framework with a multiple-instance learning approach. The first framework follows a unimodal model approach, while the second adopts a multimodal multi-model (ensemble learning) approach. This means we explore features from various modalities of endoscopy videos by applying different machine learning methods. Using an ensemble approach as shown in Fig. 1, we ensemble different ResNet models by inputting each modality into individual models. The output features of each model are then combined to form a unified output, which is subsequently utilized for making predictions.

2.1. Generating the modalities with cycleGAN

The CycleGAN [6] is employed for the conversion of WLE to VCE. A notable feature of CycleGAN is its capacity to perform this conversion without the necessity of paired data. Acquiring paired data is particularly challenging in the context of endoscopy or any clinical data. Previous studies [6] have observed that CycleGAN demonstrates effectiveness in image conversions involving color and texture changes, aligning well with the characteristics of endoscopy data.

Given a set of video $\{W_1, W_2, ..., W_n\}$, where $W_i = \{w_{i1}, w_{i2}, ..., w_{im}\}$, and the video frame $w_{ij} \in W$. The objective is to convert the video frames from input domain (W) to desired domain (V) with the help of the mapping function M such that, $M : W \rightarrow V$ and inverse mapping function $I : V \rightarrow W$. The generator model does the mapping from one domain to another domain, and the discriminator model checks whether it differs from the desired domain or not.

We incorporate three distinct loss functions, as outlined in [12]: Adversarial Loss, Cycle Consistency Loss, and Identity Loss. Giving a total generative Loss L_G of:

$$L_G(M, V) = L_{GAN} + \lambda * L_{cyc} + 0.5 + \lambda * L_{id} \quad (1)$$

where λ is a weight assigned for the cycle loss and identity loss.

2.2. CNN with Multiple Instance Learning

We trained a CNN network to detect subtle vascular and mucosal changes reflecting chronic and acute inflammation in the generated frames and classify them according to the degree of inflammation as either mucosal inflammation or mucosal healing. Due to the nature of the data we employ a Multiple Instance Learning (MIL) approach based on CNN following it's use in [8]. Given the generated data which consists of bags $\{X_1, X_2, ..., X_n\}$ and bag labels $\{Y_1, Y_2, ..., Y_n\}$, where $X_i = \{x_{i1}, x_{i2}, ..., x_{im}\}$, $x_{ij} \in X$ and $Y_i \in \{0, 1\}$. All bags are of varying sizes m. Using the pertained CNN (ResNet) the features from each instance were extracted, and the extracted features from these instances were merged to yield an aggregated feature vector X.

The goal is to train a bag classifier $H(X) : X^m \rightarrow Y$. The loss function used to optimize the end-to-end MIL approach is the cross-entropy cost function:

$$Loss = \sum_i \left(I(Y_i = 1) \log \hat{Y}_i + I(Y_i = 0) \log \left(1 - \hat{Y}_i\right)\right) \quad (2)$$

with $I(.)$ being an indicator function.

2.3. Ensemble Learning

The ensemble learning method is employed for multi-modality learning. Given different inputs of different modalities, $X1$, $X2$, and $X3$. Each modality is passed into an individual ResNet model, such that $F_{R1} = ResNet(X1)$, $F_{R2} = ResNet(X2)$ and $F_{R3} = ResNet(X3)$. The features extracted from the three ResNet models, denoted as F_{R1},F_{R2},F_{R3}, are concatenated before the final classification layer. This can be expressed as follows:

$$C = concat(F_{R1}, F_{R2}, F_{R3}) \quad (3)$$

with C being the concatinated output that is passed into the final classification layer.

3. RESULTS AND DISCUSSION

3.1. Implementation Details

Our model is implemented based on the Pytorch framework. During the cycleGAN training, the training parameters are configured as follows: 50 epochs, a batch size of 4, and a learning rate of 0.0002. The weight λ for Cycle consistency

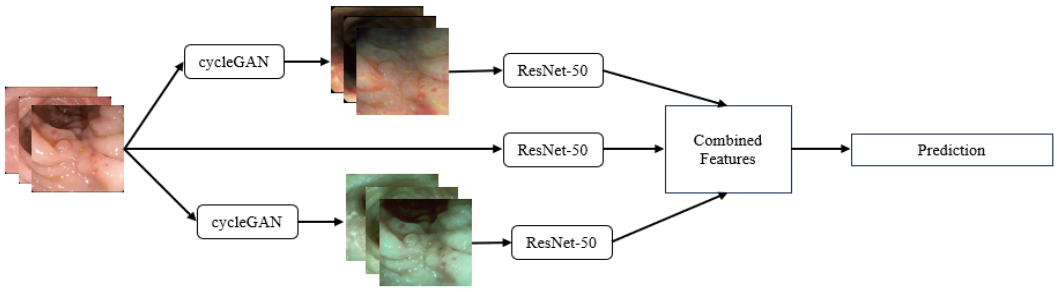

Fig. 1. The proposed model architecture, outlining the conversion of the WLE using cycleGAN to iScan2 and iScan3. Subsequently, the inputs are fed into various ResNet-50 models to create an ensemble. The output features are then aggregated for prediction.

loss is set to 10, while the weight for identity loss is set to $\lambda * 0.5$.

In constructing our ensemble model, we utilize ResNet-50 pretrained on ImageNet. We adapt the output layers to align with our specific task. The ensemble network is trained using a stochastic gradient descent (SGD) optimizer with softmax activation, employing a batch size of 15 and training for 50 epochs. The learning rate, decay, and momentum are set to default values. For faster training, all training is executed on the Intel(R) Xeon(R) Gold 6330 CPU with A40 Nvidia GPU.

3.2. Dataset

This study analyzed 240 WLE data from the PICaSSO group [7], sourced from ulcerative colitis patients across 11 centers in Europe and North America. The data is obtained from the sigmoid and rectum of each patient. The most inflamed areas or those exhibiting representative features of endoscopic remission (ER) determined by PICaSSO are made as the target. Each endoscopy video within the dataset underwent a downsampling procedure. This procedure standardized the videos into a uniform sequence of precisely 15 frames totaling 4,605 frames/images in all the videos combined. We made sure equivalent modalities of all frames are generated when using the cycleGAN. After generation, we eliminated some frames in each video that are of poor quality giving videos of varying sizes. For unimodal and multimodal training we split the data into 70%(168 samples) for training and 30% (72 samples) percent for validation.

3.3. Modality conversion (cycleGAN) result

Frames of dimensions 512 x 512 from videos serve as input for the conversion to all other modality classes (i.e WLE to iScan2, WLE to iScan3) which is done with the cycleGAN. Prior to the conversion process, the frames are processed and normalized for effective training. The losses observed are as follows: Generator Loss: 1.107125, Adversarial Loss: 0.642476, Cycle Consistency Loss: 0.034698, Identity Loss: 0.023534. From Fig. 2, it can be seen that the model performed decent work on the conversion of WLE to iScan2 and iScan3.

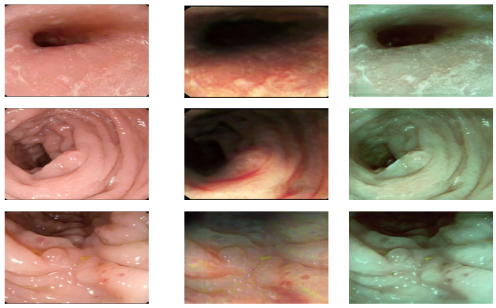

Fig. 2. The image shows the WLE and the corresponding generated iScan2 and iScan3.

3.4. Generated modalities results

The generated dataset was partitioned into three subsets: WLE, iScan2, and iScan3. The frames in each video of each modality were processed and resized into 224×224 and augmentation was performed. We conducted various experiments, initially employing each modality as an input individually. This configuration constitutes the unimodal model. Subsequently, we refer to the second model, which utilizes ensemble learning, as the multimodal mode. The unimodal model processes a single modality. It undergoes training on each of the three modalities separately to assess their performance. This step is crucial for validating our ensemble learning approach. In contrast, the multimodal model leverages the ensemble learning method to incorporate the three different inputs. All models, both unimodal and multimodal, are trained using identical parameters.

The validation results, presented in Table 1, and the ROC as shown in Fig. 3, indicate that training an ensemble model based on the generated modalities appears to enhance the prediction accuracy.

4. CONCLUSION

In this work, we propose an inflammation detection using ensemble endoscopic multimodal assessment in inflammatory bowel disease. The model uses different modalities generated

Table 1. The table illustrates the evaluation results obtained from the unimodal and multimodal strategies.

	Accuracy	Sensitivity	Specificity	PPV	NPV
Unimodal$_{WLE}$	0.82	0.81	0.82	0.57	0.94
Unimodal$_{iScan2}$	0.82	0.71	0.88	0.74	0.86
Unimodal$_{iScan3}$	0.82	0.73	0.86	0.70	0.88
Multimodal$_{WLE+iScan2+iScan3}$	0.89	0.80	0.94	0.87	0.90

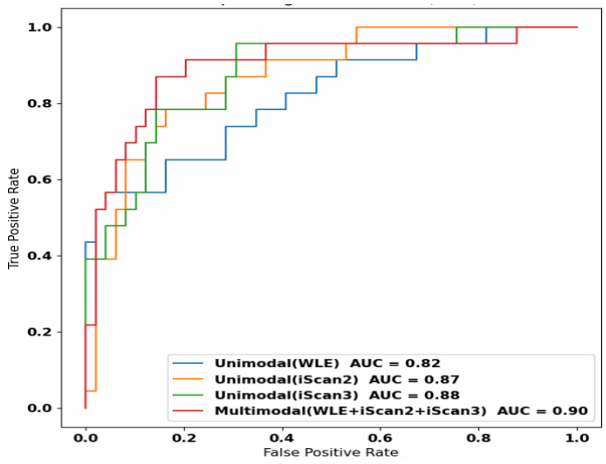

Fig. 3. The ROC curves for both unimodal and multimodal strategies

from a single modality using the cycleGAN model. We were able to achieve a better performance using the combination of WLE and its generated modalities as compared to that of using a single modality.

Future efforts will focus on incorporating additional modalities to aid in assessing inflammation in patients with Inflammatory Bowel Disease (IBD). This expansion is anticipated to contribute to improved accuracy in inflammation prediction. Furthermore, the research can be extended to include other medical datasets for predicting or detecting specific disease patterns.

5. COMPLIANCE WITH ETHICAL STANDARDS

Approval for ethical considerations was granted by the West Midlands Research Ethics Committee (17/WM/0223) and the institutional ethics committees of all participating centers. Each patient provided informed consent before participating in the study.

6. REFERENCES

[1] Daniel C Baumgart and Simon R Carding, "Inflammatory bowel disease: cause and immunobiology," *The Lancet*, vol. 369, no. 9573, pp. 1627–1640, 2007.

[2] Nataliia A Obukhova, Alexandr A Motyko, Alexandr A Pozdeev, and Armine R Hachaturyan, "Image processing algorithm for virtual chromoendoscopy (tone enhancement) in clinical decision support system," in *2018 23rd Conference of Open Innovations Association (FRUCT)*. IEEE, 2018, pp. 293–299.

[3] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[4] Yuhui Ma, Jiang Liu, Yonghuai Liu, Huazhu Fu, Yan Hu, Jun Cheng, Hong Qi, Yufei Wu, Jiong Zhang, and Yitian Zhao, "Structure and illumination constrained gan for medical image enhancement," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3955–3967, 2021.

[5] Fan He, Sizhe Chen, Shuaiyi Li, Lu Zhou, Haiqin Zhang, Haixia Peng, and Xiaolin Huang, "Colonoscopic image synthesis for polyp detector enhancement via gan and adversarial training," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 1887–1891.

[6] Yuzhe Zhu, Chao Xu, Bo Feng, and Shanshan Fang, "Research on endoscopic image enhancement algorithm based on contrast fusion," in *2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2021, pp. 210–215.

[7] Marietta Iacucci, Marco Daperno, Mark Lazarev, Razvan Arsenascu, Gian Eugenio Tontini, Oluseyi Akinola, Xianyong Sean Gui, Vincenzo Villanacci, Martin Goetz, Mark Lowerison, et al., "Development and reliability of the new endoscopic virtual chromoendoscopy score: the picasso (paddington international virtual chromoendoscopy score) in ulcerative colitis," *Gastrointestinal endoscopy*, vol. 86, no. 6, pp. 1118–1127, 2017.

[8] Rocío Del Amor, Pablo Meseguer, Tommaso Lorenzo Parigi, Vincenzo Villanacci, Adrián Colomer, Laëtitia Launet, Alina Bazarova, Gian Eugenio Tontini, Raf Bisschops, Gert de Hertogh, et al., "Constrained multiple instance learning for ulcerative colitis prediction using histological images," *Computer methods and programs in biomedicine*, vol. 224, pp. 107012, 2022.