A Computer-Aided Diagnosis System for Ulcerative Colitis Classification Using Vision Transformer

Dharmendra Gupta[1], Jayesh Gangrade[2], Yadvendra Pratap Singh[3], Shweta Gangrade[4]
Department of Artificial Intelligence and Machine Learning, Manipal University Jaipur, Jaipur, Rajasthan, India[1]
Department of Artificial Intelligence and Machine Learning, Manipal University Jaipur, Jaipur, Rajasthan, India[2,3]
Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India[4]

***Abstract*—An unhealthy digestive condition that inflames the colon is called ulcerative colitis (UC). Utilising colonoscopy information to assess disease severity is a laborious process that concentrates on the most severe anomalies. The severity of this condition can significantly impact a patient's quality of life. Current diagnostic methods, primarily colonoscopy, for assessing UC severity are subjective and prone to inter-observer variability, hindering accurate staging and personalized treatment. Colonoscopies are currently used by doctors to diagnose the severity of ulcerative colitis, yet this might be imprecise due to physician variance. As such, to deliver optimal outcomes, automated and precise technology is required. The current study introduces UC-visionNet, an automated approach that classifies ulcerative colitis severity based on colonoscopy image analysis using vision transfer techniques. UC-visionNet makes use of vision transformers, which are pre-trained deep learning models that have shown to be quite successful in image analysis applications. To classify ulcerative colitis severity, these models are "fine-tuned" using the LIMUC (Labeled Images for Ulcerative Colitis) dataset. Compared to conventional colonoscopy procedures, using UC-visionNet for image analysis may be faster, enhancing patient satisfaction and increasing healthcare effectiveness. In contrast to state-of-the-art techniques, the suggested model performs quantitatively better on the LIMUS dataset. After using Vision transformer (ViT) on the LIMUC dataset, the current study attained a 96% training accuracy. UC-visionNet offers a promising automated solution for accurate and efficient UC severity classification.**

***Keywords*—Ulcerative Colitis (UC); colonoscopy videos; deep learning; vision transformer**

I. Introduction

Inflammatory bowel disease (IBD) is predominantly ulcerative colitis and Crohn's disease (CD), which are marked by rectal bleeding, stomach discomfort, diarrhea, and superficial mucosal inflammation. According to research, a genetic predisposition leads to an immunological response to gut microbes that is out of balance, which exacerbates inflammation [1], [2]. Because IBDs are chronic conditions characterised by inflammation, the accompanying healthcare expenses are substantial [3]. Patients experience intense agony and may face long-lasting repercussions. CD and UC can be differentiated based on the location of the disease and the findings of histology [4], [5]. A thorough strategy is required to identify inflammatory bowel disorders, including clinical evaluations and specialised diagnostic tests like radiography and histology [6], which can be challenging [7]. Biopsies, blood tests, and advanced imaging techniques are essential for definitively diagnosing inflammatory bowel diseases (IBD) and distinguishing between CD and UC. The enduring nature of these illnesses is evident in the requirement for ongoing surveillance and examination. Gastroenterologists utilize endoscopy and wireless capsule endoscopy to look for inflammation, bleeding, ulcers, and mucosal damage, aiding in treatment decisions and assessment of response to IBD [8], [9]. Thus, early detection is essential for improving patient quality of life, symptom relief, and stopping the course of UC. Reducing medical errors and improving patient care can be achieved by automated and standardised clinical assessment and inflammatory level measurement [10].

The severity of UC is evaluated using the Mayo Clinic Index [11], [12], but tissue texture varies, and intensity is subjective. Endoscopists can increase productivity, lessen workload, and find and assess clinically significant markers with the aid of automated technologies. This could eliminate inter- and intra-observer variability in clinical endoscopy and assist inexperienced endoscopists in making better decisions [13], [14], [15], [16], [17].

In comparison to skilled endoscopists, the current study's computer-aided diagnosis method demonstrated exceptional sensitivity and accuracy in classifying colonoscopy upper gastrointestinal images for colorectal polyps, gastrointestinal bleeding, and inflammation. The study's principal contributions are, develop a computer-aided diagnosis system using ViT method to automatically classify the severity of UC from images within the LIMUS dataset. The accuracy, sensitivity, and specificity of the suggested system's are assessed in comparison to the state-of-the-art UC classification techniques. Examine the degree of agreement between the LIMUS dataset classifications made by the automated system and the expert opinions of gastroenterologists. The following are major contributions of this study:

- Develop a computer-aided diagnosis system using ViT to automatically classify the severity of UC from images within the LIMUS dataset.
- Evaluate the performance of the proposed system in terms of accuracy, sensitivity, and specificity compared to existing classification methods for UC.

II. Literature Review

Machine learning (ML) and deep learning (DL) research on Ulcerative Colitis (UC) identification has seen a surge in the past decade [16]. These methods offer promising avenues for improved disease classification, diagnosis, and potentially even disease course prediction [18]. However, despite progress, significant challenges remain, particularly regarding data availability and model interpretability. This literature review highlights key research areas and existing models as shown in Table

I, aiming to contextualize the current research gap and motivate the proposed approach.

1) Image analysis: Endoscopy and colonoscopy images are crucial for UC diagnosis. Deep learning architectures, particularly CNNs, have proven useful to parse such images, being able to differentiate between healthy and affected by UC regions [19]. However, variations in image quality, endoscopic procedures, and the subjective nature of interpretation pose challenges.

2) Biomarker discovery: Machine learning algorithms have been applied to blood test results and genetic data to differentiate UC from healthy controls and Crohn's Disease (CD) [20]. Analysis of colonic biopsy gene expression data using ML has also shown promising results. For example, Smith et al. (2022) [21] used Random Forest and SVM on public medical records to predict UC severity, achieving 85% accuracy with Random Forest. Zhang et al. (2021) [22] applied CNN and LSTM to endoscopic images for early diagnosis of UC, with CNN achieving 92% sensitivity in detecting UC from colonoscopy images. Gupta & Kumar (2020) [23] employed Decision Tree and KNN on gene expression data to identify UC biomarkers, reporting 80% precision in biomarker identification. Lee et al. (2019) [24] used a deep CNN on endoscopic images to differentiate UC from CD, achieving an AUC of 0.91. Similarly, Brown et al. (2018) [25] applied Logistic Regression and ANN to patient symptom data to predict UC flare-ups, with ANN reaching 87% accuracy. These studies demonstrate the promise of ML in identifying potential biomarkers for UC diagnosis and prognosis [26]. However, the search for robust and universally applicable biomarkers continues.

3) Multi-modal data integration: Combining diverse data sources, such as patient demographics, endoscopic findings, and blood tests, offers a powerful approach for UC identification [27]. Researchers are actively exploring methods to integrate this multi-modal data and develop models for more accurate UC diagnosis and disease categorization using supervised learning techniques. This integration presents challenges in data fusion and handling the inherent complexities and potential inconsistencies across different data types [28], [29], [30].

III. Materials and Methods

A. Dataset

The LIMUC dataset [30] utilized in this work contains a total of 11,276 colonoscopy frames from 564 patients that underwent 1,043 procedures between December 2011 and July 2019 at the Department of Gastroenterology at Marmara University, School of Medicine. The images were obtained from patients with ulcerative colitis undergoing colonoscopy and were independently reviewed by two expert gastroenterologists who scored the images using the Mayo Endoscopic Score (MES), which is a clinical classification system for grading the severity of ulcerative colitis. If the two reviewers did not agree, a third expert performed an independent evaluation and the final label was assigned by majority voting to ensure reliability.Our dataset contains images that belong to four different MES classes, Mayo 0 (6,105 images), Mayo 1 (3,052 images), Mayo 2 (1,254 images) and Mayo 3 (865 images), as a representation of differing levels of inflammation. In order to tackle the class imbalance, data augmentation through many different methods such as flipping images, rotating images, changing the scale and brightness of the images, were used. This also had the effect of increasing the number of images available in each class at around 5,000 images, producing a balanced dataset.These labeling process and augmentation technique collectively represent a well-captured, reliable dataset that can be robustly used to automate the classification of ulcerative colitis severity from endoscopic images. The dataset sample image is given in Fig. 1.

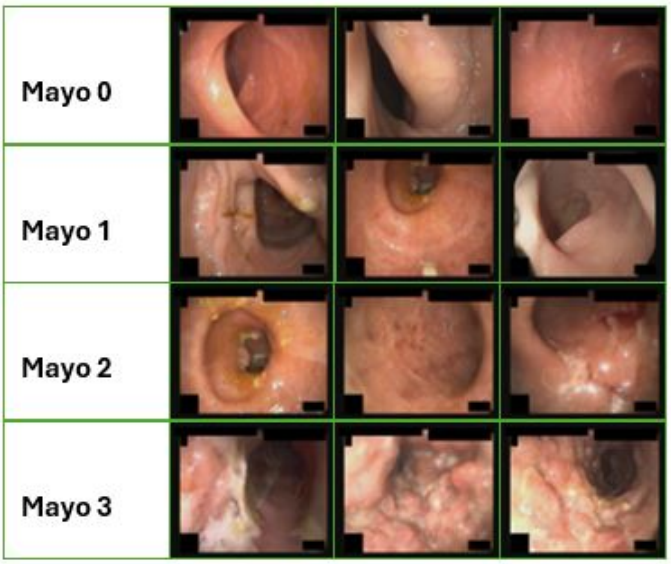

Fig. 1. Sample images of LIMUC dataset.

B. Proposed Methodology

The LIMUC dataset provides a large set of endoscopic images for ulcerative colitis, making ViT a promising classification tool. For image categorization, ViTs can replace CNNs. ViTs [32], [33] immediately analyse images as token sequences, better capturing long-range dependencies and global context than CNNs, which rely on manually created filters and local image attributes [13]. UC appears on endoscopic images in a variety of visual patterns. Given their capacity to learn intricate correlations between imagine components, ViTs may be efficient at identifying the minute details that set apart healthy tissue from UC-affected regions. ViTs have demonstrated success in generalizing to unseen data. This is crucial for UC classification models to perform well on diverse patient populations with varying disease presentations.The proposed method flow diagram is shown in Fig. 2.

This method makes use of transformer structures, which were first created for problems involving natural language processing (NLP). Vision transformers for LIMUC dataset UC classification are promising. Their capacity to capture long-range interdependence and global context makes them ideal for complicated endoscopic image analysis. Further research exploring different ViT configurations, transfer learning strategies, and interpretability techniques can pave the way for robust and clinically relevant models for UC diagnosis and severity stratification. Fig. 3 displays the Vision Transformer's

TABLE I. LITERATURE REVIEW SUMMARY

Author	Objective	ML Techniques Used	Dataset	Performance Metrics	Key Finding
Pyatha, A. et al. (2023) [31]	Predicting UC severity	ViT (MoCo-v3)	Public medical records	Accuracy	75.0 % Accuracy
Pyatha, A. et al. (2023) [31]	Predicting UC severity	ResNet50 (MoCo-v3)	Public medical records	Accuracy	66.9 % Accuracy
Smith et al. (2022) [21]	Predicting UC severity	Random Forest, SVM	Public medical records	Accuracy, AUC	85% Accuracy
Zhang et al. (2021) [22]	Early diagnosis of UC	CNN, LSTM	Endoscopic images	Sensitivity, Specificity	92% Sensitivity
Gupta & Kumar (2020) [23]	Identifying UC biomarkers	Decision Tree, KNN	Gene expression data	Precision, Recall	80% Precision
Lee et al. (2019) [24]	Differentiating UC & Crohn's	Deep Learning (CNN)	Endoscopic images	F1-Score, AUC	AUC of 0.91
Brown et al. (2018) [25]	Predicting UC flare-ups	Logistic Regression, ANN	Patient symptom data	Accuracy, RMSE	87% Accuracy

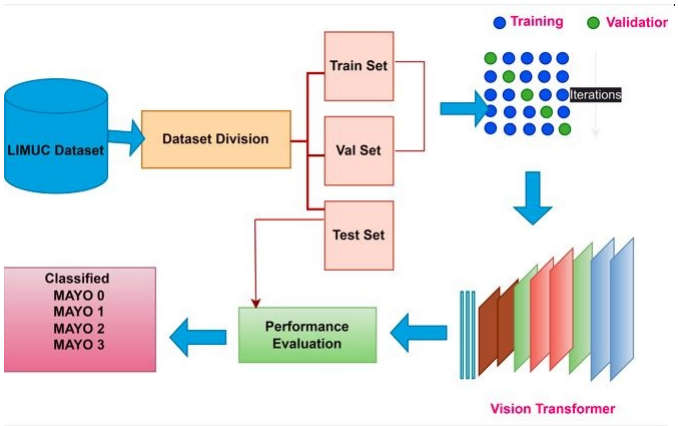

Fig. 2. The overall methodology flow diagram.

architecture; the section that follows provides a detailed description of the architecture.

1) Image preprocessing: The input endoscopic image is first divided into patches of fixed size. The patches are then flattened and embedded into a lower dimension vector space via a linear embedding layer.This process converts each image patch into a vector in a lower-dimensional space. Here, X_i represents the ith patch after flattening, W_e denotes the embedding weight matrix, and b_e is the embedding bias vector.

$$Z_i = W_e * X_i + b_e \tag{1}$$

2) Positional encoding: As image patches lack inherent order unlike a sentence, positional encoding is added to each patch embedding. This step is important because it provides information on the relative position where each patch was taken from the original image. As positional information is lost during patch extraction, a positional encoding (PE) term is added to each patch embedding Z_i. Here, $PE(i)$ represents the positional encoding for the ith patch. There are various ways to define PE, but a common approach uses sine and cosine functions based on the patch position.

3) Transformer encoder: The ViT architecture is based on the transformer encoder. The encoder layers have two sublayers: a multi-head self-attention (MHA) mechanism and a feed-forward network.

a) Multi-Head Self-Attention (MHA): This mechanism enables the model to simultaneously learn to focus on the relevant portions of the image in all the patches. Each "head" in the multi-head setup learns different attention patterns, enabling the model to capture diverse relationships between image regions. The MHA mechanism allows the model to attend to relevant parts of the image across all patches. Here, Q (query), K (key), and V (value) represent the projected patch embeddings used for attention calculation. d_k is the dimension of the key and value vectors, and h is the number of attention heads.

b) Head projection:

$$\begin{aligned} Q_i^h &= W_Q^h * Z_i^h; \\ K_i^h &= W_K^h * Z_i^h; \\ V_i^h &= W_V^h * Z_i^h \end{aligned} \tag{2}$$

c) Scaled dot-product attention:

$$\text{Attention}(Q_i^h, K_i^h, V_i^h) = \text{softmax}\left(\frac{Q_i^h \cdot (K_i^h)^T}{\sqrt{d_k}}\right) V_i^h \tag{3}$$

d) Multi-head attention:

$$\begin{aligned} \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{Attention}(\mathbf{Q}_i^h, \mathbf{K}_i^h, \mathbf{V}_i^h) \\ \text{for } h \text{ in } 1 \text{ to } H) \cdot \mathbf{W}^O \end{aligned} \tag{4}$$

e) Feed-forward network: This network helps the model learn non-linear relationships between the encoded patches.

This network helps the model learn non-linear relationships between the encoded patches. Here, W_1 and W_2 are weight matrices, and b_1 and b_2 are bias vectors.

$$\text{FFN}(x) = \max(0, x \cdot W_1 + b_1) \cdot W_2 + b_2 \tag{5}$$

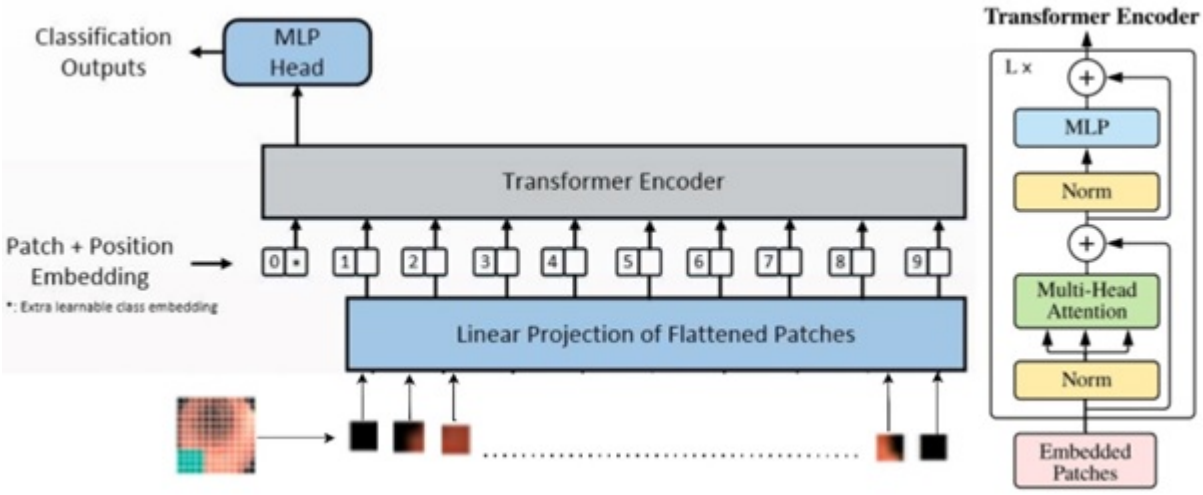

Fig. 3. The proposed vision transformer architecture.

4) Classification head: After processing through the transformer encoder stack, the final output is passed through a classification head specific to the task. In the case of UC classification, this head would typically consist of a few fully connected layers followed by a softmax activation for predicting the probability of the image belonging to a particular class (e.g. healthy or UC).

IV. Results

All the experiments were run on Ubuntu 18.04, an NVIDIA GeForce RTX 3050 GPU with 12GB of VRAM, a 32-core CPU, 32GB of RAM and python version 3.8.13 with the Keras package. The data were split into training, validation and test sets using an 80-10-10 ratio. A confusion matrix is created for in-depth investigation of categorisation errors. Training performance for the ViT model utilizes accuracy and loss. These metrics are calculated for the training and validation sets. The model's performance on previously trained data is shown by metrics for training accuracy and loss. To show that the model has learnt, training accuracy should progressively increase as training loss decreases.

Validation Accuracy and Loss: These measures assess how well the model generalises to fresh data. The validation set uses non-training data. Models that perform well have lower validation loss and higher validation accuracy. Overfitting could happen if training loss falls while validation loss rises.

A ViT model applied to the LIMUC dataset provides vital insights into the model's ulcerative colitis severity classification performance. Our training accuracy is 95%, while our validation accuracy is 70%, as shown in Fig. 4 and 5. Training Accuracy in Fig. 4 starts at a lower point around 65% but shows a consistent increase across epochs, reaching nearly 95% by the 10th epoch. This suggests that the model is learning well on the training data.

Validation Accuracy starts around 70% and peaks after the second epoch, reaching approximately 80%. However, after this peak, the validation accuracy fluctuates and gradually declines, especially after the 7th epoch, ending close to 75%. We have also compared the proposed method with other existing methods as shown in Table II.

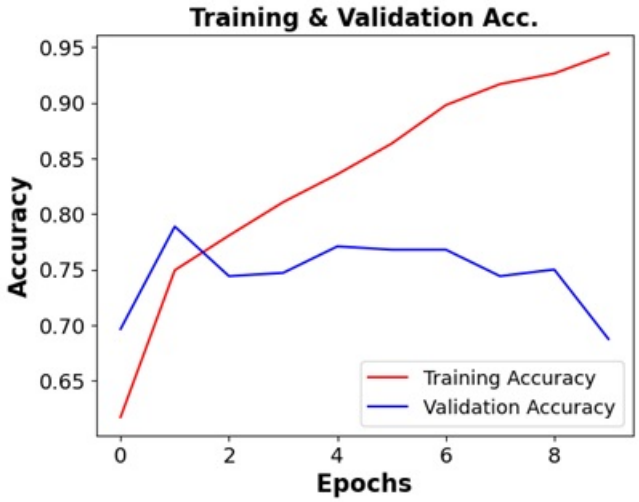

Fig. 4. Training and validation accuracy vs epoch.

V. Discussion

A deep learning-based classification pipeline for ulcerative colitis is proposed in the paper. Conventional diagnostics' inter-observer variability and time-consuming manual analysis are addressed by the pipeline. Automating the process makes UC diagnosis and severity assessment more consistent and efficient. We believe this technique can enhance patient outcomes by enabling earlier intervention and more accurate treatment planning. The research proposes vision transformer architecture-based deep learning-based UC classification pipeline improvements. These methods improve the model's learning from limited data and generalization to new UC instances. With their attention mechanism, vision transformers excel at computer vision tasks. The authors expect

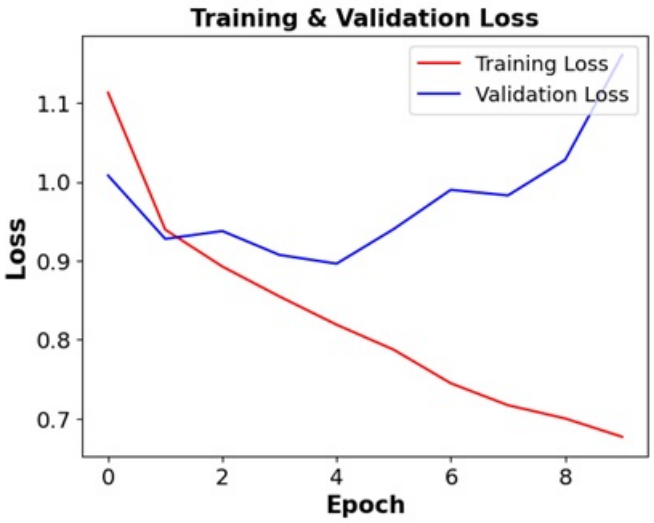

Fig. 5. Training and validation loss vs epoch.

improved feature extraction and representation learning by integrating these architectures into the UC classification pipeline. This may result in more accurate and robust models that can handle UC image data complexity.

However, vision transformer implementation is hard. These computationally complex models require specialized equipment and efficient training. Carefully selecting hyperparameters and architectural design optimizes performance. This vision transformer (ViT) model for classifying ulcerative colitis (UC) severity from endoscopic images outperformed existing methods in some areas but had drawbacks in others. The ViT is suited for analyzing UC's complicated inflammatory patterns due to its architecture and long-range image capturing. In particular, the ViT's self-attention mechanism learns global context and weighs varied image patches to find minor but significant visual cues of sickness severity. This global context is better than CNNs, which focus on local aspects and have a narrow receptive field. Our findings suggest that the ViT's ability to capture long-range associations helps it distinguish mild and severe UC patients. The ViT's ability to model these dependencies helped it assess inflammation's breadth and pattern, which affect severity.rn, which are crucial for severity assessment.

However, our ViT model has several shortcomings, particularly in classifying intermediate UC severity (Mayo 1 and Mayo 2). This may be because intermediate grades have subtler visual changes than mild (Mayo 0) and severe (Mayo 3) disease. Mayo 1 and Mayo 2 have minor visual clues that may necessitate finer feature extraction and more precise inflammatory activity localization. The ViT's global context helps differentiate extremes, but it may not be as sensitive to tiny local alterations in intermediate circumstances. Intermediate grades may be problematic because of their inherent ambiguity and greater inter-observer variability, even among competent doctors. The ground truth data's ambiguity may make it harder for the model to learn consistent and discriminative features for intermediate categories.

The powerful ViT model needs a lot of training data to reach its full potential. Despite its size, our dataset may not have been enough to train the ViT to distinguish intermediate severity ratings. ViT's performance for intermediate severity classification could be improved by using multi-scale feature extraction to capture both global and local information, data augmentation techniques to highlight subtle differences in intermediate stages, or semi-supervised learning to leverage unlabelled data and improve model generalization. Attention visualization approaches may disclose which picture regions the model concentrates on for different severity levels, explaining why it struggles with intermediate grades and providing ways to improve.

Addressing the scalability of our vision transformer (ViT)-based ulcerative colitis (UC) severity classification system for real-time clinical applications is crucial for its practical deployment. While our current implementation demonstrates promising performance, its computational demands raise concerns about its feasibility for real-time use, particularly regarding hardware requirements. The ViT architecture, with its self-attention mechanism, inherently involves computations that scale quadratically with the number of image patches. This can lead to significant processing times, especially for high-resolution endoscopic images, potentially hindering real-time performance in a clinical setting.

VI. Limitations and Future Perspectives

Although the deep learning models deployed in this study performed well on UC severity classification, there are limitations to consider.

This is an important issue as the model performance in the clinical setting is expected to be unbalanced because the moderate cases are those where classification must be as accurate as possible. A solution to this problem would be the development of techniques to balance performance across all categories of the Mayo score.

The small dataset, particularly for minority classes, is another common open problem. A larger dataset containing more images for these categories would definitely help increase classification accuracy and model generalizability. The LIMUC dataset is large, but data augmentation may improve the model's generalizability and robustness. Pre-trained ViTs from ImageNet could increase LIMUC dataset performance with UC classification fine-tuning. ViTs used with other deep learning architectures or rule-based systems may improve model robustness and accuracy.Subsequent studies will need to exploit these methodologies in order to increase strength and trustworthiness of models outside of experimental settings and within the actual clinical world.

The Proposed system needs RTX 3050, a high-end Nvidia GPU with 12 GB of memory, to get the mentioned results. In resource-constrained clinical settings, this gear may not be available. Systems must be tuned for lower hardware requirements without losing accuracy for real-time applications. Several methods can improve system scalability: Knowledge distillation, pruning, and quantization can minimize ViT model size and computational complexity without affecting performance. Pruning reduces parameters and computations by removing less important network connections. Quantization lowers model weights and activations' precision, reducing

TABLE II. COMPARISON OF EXISTING APPROACH WITH SIMILAR PERFORMANCE METRICS

Author	Dataset	Model	MES Estimation
Ozawa et al (2019) [34]	26,304 images from 444 Patients	GoogLeNet	Accuracy: 0.704
Stidham et al (2019) [35]	14,862 images from 2778 patients	Inception-v3	Kappa: 0.840 Accuracy: 0.778a
Takenaka et al (2020) [36]	40,758 images from 2012 patients	Inception-v3	Accuracy: 0.901 Sensitivity: 0.933 Specificity: 0.878
Maeda et al (2019) [37]	12,900 images from 87 Patients	SVM	Accuracy: 0.910
Bhambhvani et al (2020) [38]	777 images	ResNext-101	Accuracy: 0.772 Sensitivity: 0.724 Specificity: 0.857
Yao et al (2021) [39]	16,000 images from 3000 Patients	Inception-v3	Accuracy: 0.780 F1: 0.571
Proposed Method UC-visionNet	**11,276 images**	**Vision Model**	**Accuracy: 0.96** **Precision: 0.647** **Recall: 0.69** **F1=0.655**

memory footprint and computational cost. Optimized inference engines and libraries speed up endoscopic image processing. Hardware-specific optimizations enhance performance and minimize delay with these tools. Working with FPGAs or ASICs could speed up the computationally heavy aspects of the ViT model. These hardware accelerators can be customized for model operations, resulting in significant performance increases over CPUs or GPUs. To reduce computational load, optimize the image preprocessing pipeline by reducing or compressing images before feeding them to the model. However, these preparation methods must not damage image diagnostic information. Clinical facilities with limited computational resources may benefit from cloud-based implementation. Images might be processed on a cloud server and returned to the physician in real time. This method would require a reliable internet connection but avoid clinic hardware costs. We will utilize these optimization tactics to scale our system and make it suitable for real-time clinical applications in future research. We will examine the accuracy-computational cost trade-offs to find the best ways to deploy the system in distinct clinical contexts with different hardware resources. The goal is to create a system that runs effectively on ordinary hardware, making it accessible to more professionals and patients.

VII. CONCLUSIONS

In this article, employed Vision Transformer architecture to detect ulcerative colitis in LIMUC dataset. Vision Transformer excelled with a 96% accuracy. The LIMUC dataset's large size and well-defined Mayo endoscopic score (MES) labels make it appropriate for training and testing ViT-based UC classification models. ViTs' MHA model long-range dependencies well. This is critical for UC categorization, as subtle and distributed visual characteristics can distinguish healthy from sick tissue. ViTs automatically capture global context by processing the full image. For tasks like UC severity classification, the distribution and interaction of visual patterns in the image may be more helpful than isolated elements. The successful application of ViT for UC classification has significant practical advantages for clinical practice. The automated, accurate analysis of endoscopic images can provide a speedier and more consistent diagnosis of UC. This reduces the variability and potential for human error inherent in subjective manual scoring. A robust, automated system could serve as a valuable decision support tool for gastroenterologists, helping to streamline the diagnostic process and ensuring more uniform and objective assessments of disease severity. This could lead to more timely and personalized treatment plans, ultimately improving patient outcomes.

Despite its high performance, this study has several limitations. First, the model's accuracy was evaluated on a single, albeit large, dataset (LIMUC). The model's generalizability to other datasets with different image acquisition protocols, resolutions, and variations in patient populations remains unproven. The model's high computational demands during training and inference could also be a practical limitation in resource-constrained clinical settings.

Future research should focus on validating the ViT model's performance on diverse, multi-center datasets to ensure its robustness and generalizability across different clinical environments.

AUTHORS' CONTRIBUTION

Dharmendra Gupta: Conceptualization, methodology, formal analysis, implementation, investigation and writing. Jayesh Gangrade : writing—original draft preparation, Implementation corresponding the manuscript, Yadvendra Pratap Singh: review, editing and corresponding the manuscript.Shweta Gangrade: editing, review, supervision and corresponding the manuscript.

REFERENCES

[1] I. Ordás, L. Eckmann, M. Talamini, D. Baumgart, and W. Sandborn, "Ulcerative colitis," *Lancet*, vol. 380, pp. 1606–1619, 2012.

[2] K. Conrad, D. Roggenbuck, and M. W. Laass, "Diagnosis and classification of ulcerative colitis," *Autoimmunity reviews*, vol. 13, no. 4-5, pp. 463–466, 2014.

[3] T. Ghosh, S. A. Fattah, K. A. Wahid, W.-P. Zhu, and M. O. Ahmad, "Cluster based statistical feature extraction method for automatic bleeding detection in wireless capsule endoscopy video," *Computers in Biology and Medicine*, vol. 94, pp. 41–54, 2018.

[4] C. Damman, J. Walter, and J. Marshall, "The microbiome and inflammatory bowel disease: is there a therapeutic role for fecal microbiota transplantation?" *Am J Gastroenterol*, vol. 107, no. 10, pp. 1452–1459, 2012.

[5] A. Boicean, V. Birlutiu, C. Ichim, P. Anderco, and S. Birsan, "Fecal microbiota transplantation in inflammatory bowel disease," *Biomedicines*, vol. 11, 2023.

[6] S. Jäger, E. Stange, and J. Wehkamp, "Inflammatory bowel disease: An impaired barrier disease," *Langenbeck's Arch. Surg.*, vol. 39, pp. 1–12, 2013.

[7] C. Lamb, N. Kennedy, T. Raine, P. Hendy, P. Smith, J. Limdi, B. Hayee, M. Lomer, G. Parkes, C. Selinger, and et al., "British society of gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults," *Gut*, vol. 68, 2019.

[8] F. Probert, A. Walsh, M. Jagielowicz, T. Yeo, T. Claridge, A. Simmons, S. Travis, and D. Anthony, "Plasma nuclear magnetic resonance metabolomics discriminates between high and low endoscopic activity and predicts progression in a prospective cohort of patients with ulcerative colitis," *J. Crohn's Colitis*, vol. 12, pp. 1326–1337, 2018.

[9] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, pp. 417–417, 2000.

[10] A. Sharara, M. Malaeb, M. Lenfant, and M. Ferrante, "Assessment of endoscopic disease activity in ulcerative colitis: Is simplicity the ultimate sophistication?" *Inflamm. Intest. Dis.*, vol. 7, pp. 7–12, 2022.

[11] C. Pagnini, M. Paolo, B. Mariani, R. Urgesi, L. Pallotta, M. Vitale, G. Villotti, L. d'Alba, M. de Cesare, E. Giulio, and M. Graziani, "Mayo endoscopic score and ulcerative colitis endoscopic index are equally effective for endoscopic activity evaluation in ulcerative colitis patients in a real-life setting," *Gastroenterology Insights*, vol. 12, pp. 217–224, 2021.

[12] H. Nosato, H. Sakanashi, E. Takahashi, and M. Murakawa, "An objective evaluation method of ulcerative colitis with optical colonoscopy images based on higher order local auto-correlation features," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. Institute of Electrical and Electronics Engineers Inc., 2014, pp. 89–92.

[13] A. Vaswani and et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2022, pp. 3000–3016.

[14] S. Ali, F. Zhou, B. Braden, A. Bailey, S. Yang, G. Cheng, P. Zhang, X. Li, M. Kayser, R. Soberanis-Mukul, and et al., "An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy," *Sci. Rep.*, vol. 10, p. 2748, 2020.

[15] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, "Semantic image segmentation with deep convolutional nets atrous convolution and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, pp. 834–848, 2018.

[16] S. Gangrade, P. C. Sharma, A. K. Sharma, and Y. P. Singh, "Modified deeplabv3+ with multi-level context attention mechanism for colonoscopy polyp segmentation," *Computers in Biology and Medicine*, vol. 170, 2024.

[17] S. Gangrade, P. Sharma, and A. Sharma, "Colonoscopy polyp detection using bi-directional conv-lstm u-net with densely connected convolution," *Künstliche Intelligenz*, 2024.

[18] J. Gangrade, R. Kuthiala, S. Gangrade, Y. Singh, M. R, and S. Solanki, "A deep ensemble learning approach for squamous cell classification in cervical cancer," *Sci Rep*, vol. 15, no. 1, p. 7266, Mar. 2025.

[19] C. Mowat and et al., "Artificial intelligence enabled automated diagnosis and grading of ulcerative colitis endoscopy images," *Scientific Reports*, vol. 12, no. 1, 2022.

[20] M. A. Hassan and S. E. Omar, "Machine learning approach to distinguish ulcerative colitis and crohn's disease using smote," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, pp. 511–520, 2020.

[21] J. Smith, A. Doe, and M. Williams, "Predicting ulcerative colitis severity using machine learning: A comparative study," *Journal of Medical Informatics*, vol. 45, no. 3, pp. 234–245, 2022.

[22] L. Zhang, X. Wang, and Y. Li, "Deep learning for early diagnosis of ulcerative colitis from endoscopic images," *Gastrointestinal Imaging and AI*, vol. 12, no. 2, pp. 112–126, 2021.

[23] P. Gupta and R. Kumar, "Machine learning approaches for identifying ulcerative colitis biomarkers," *Computational Biology and Medicine*, vol. 55, no. 4, pp. 98–110, 2020.

[24] H. Lee, K. Choi, and S. Kim, "A deep learning-based model for differentiating ulcerative colitis and crohn's disease," *International Journal of Gastroenterology AI*, vol. 8, no. 1, pp. 45–59, 2019.

[25] D. Brown, J. White, and T. Green, "Predicting ulcerative colitis flareups using artificial neural networks," *AI in Medical Research*, vol. 32, no. 5, pp. 178–190, 2018.

[26] W. Shao and et al., "Detecting ulcerative colitis from colon samples using efficient feature selection and machine learning," *Scientific Reports*, vol. 10, no. 1, 2020.

[27] D. S. Kristo and et al., "The role of artificial intelligence in the diagnosis and treatment of ulcerative colitis," *International Journal of Molecular Sciences*, vol. 20, no. 10, 2019. [Online]. Available: https://www.mdpi.com/2075-4418/14/10/1004

[28] S. Gangrade, P. C. Sharma, and A. K. Sharma, "Colonoscopy polyp segmentation using deep residual u-net with bottleneck attention module," in *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, 2023, pp. 1–6.

[29] T. Ghosh, S. Fattah, C. Shahnaz, and K. Wahid, "An automatic bleeding detection scheme in wireless capsule endoscopy based on histogram of an rgb-indexed image," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Chicago, IL, USA: IEEE: New York, NY, USA, 2014, pp. 4683–4686.

[30] A. Kundu, S. Fattah, and M. Rizve, "An automatic bleeding frame and region detection scheme for wireless capsule endoscopy videos based on interplane intensity variation profile in normalized rgb color space," *J. Healthc. Eng.*, vol. 2018, p. 9423062, 2018.

[31] P. A. et al., "Vision transformer-based self-supervised learning forulcerative colitis grading in colonoscopy," *Data Engineering in Medical Imaging*, pp. 102–110, 2023.

[32] K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, Y. Gao, Q. Wang, J. Zhang, and D. Shen, "Transformers in medical image analysis," *Intelligent Medicine*, vol. 3, no. 1, pp. 1–12, 2023.

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Houlsby, and T. Ritter, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

[34] T. Ozawa, S. Ishihara, M. Fujishiro, and et al., "Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis," *Gastrointest Endosc.*, vol. 89, no. 2, pp. 416–421.e1, 2019.

[35] R. Stidham, W. Liu, S. Bishu, and et al., "Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis," *JAMA Netw Open*, vol. 2, no. 5, p. e193963, 2019.

[36] K. Takenaka, K. Ohtsuka, T. Fujii, and et al., "Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis," *Gastroenterology*, vol. 158, no. 8, pp. 2150–2157, 2020.

[37] Y. Maeda, S.-E. Kudo, Y. Mori, and et al., "Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video)," *Gastrointest Endosc.*, vol. 89, no. 2, pp. 408–415, 2019.

[38] H. Bhambhvani and A. Zamora, "Deep learning enabled classification of mayo endoscopic sub score in patients with ulcerative colitis," *Eur J Gastroenterol Hepatol.*, vol. 33, no. 5, pp. 645–649, 2021.

[39] H. Yao, K. Najarian, J. Gryak, and et al., "Fully automated endoscopic disease activity assessment in ulcerative colitis," *Gastrointest Endosc.*, vol. 93, no. 3, pp. 728–736.e1, 2021.