Robust Ulcerative Colitis Detection via Integrated Convolutional Feature Encoding, Bidirectional Temporal Context, and Data Augmentation for Class Imbalance

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[1,2,3]
Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur, Rajasthan, India[4]

***Abstract*—Ulcerative Colitis (UC), a chronic inflammatory bowel disease, presents significant diagnostic challenges due to its overlapping symptoms with other gastrointestinal disorders and the complex visual patterns in endoscopic imagery. Accurate and early detection is essential to guide effective treatment and improve patient outcomes. This research introduces a robust hybrid framework that combines convolutional feature extraction with bidirectional temporal modelling for the precise identification of UC from medical imagery. The proposed approach integrates CNNs—including MobileNetV3Large, Inception v3, InceptionResNetV2, and Xception—with Bi-GRU and Bi-LSTM networks. The CNNs are responsible for capturing high-level spatial features, while the Bi-GRU and Bi-LSTM modules enhance temporal context understanding, enabling the model to effectively interpret subtle patterns and transitions characteristic of UC. Each hybrid model was designed, and thoroughly tested on an curated set of experimental data. Among the combinations, the highest accuracy was of 93.10%, obtained with the Xception + Bi-GRU + Bi-LSTM model. Inception v3 + Bi-GRU + Bi-LSTM followed closely, attaining an accuracy of 92.62%. The different data augmentation techniques is deployed to handle the class imbalance that exists in the LIMUC dataset . Notably, the bidirectional temporal modelling component significantly improved the recognition of sequential dependencies in medical image frames, enhancing the model's diagnostic robustness. The findings demonstrate that integrating CNNs with bidirectional temporal encoders offers a promising solution for UC detection, providing a valuable tool for clinicians in automated diagnostic systems. This study not only contributes to the advancement of intelligent medical imaging but also paves the way for deploying real-time UC detection models in clinical practice.**

***Keywords*—Ulcerative Colitis Detection (UCD); CNNs; Bi-GRU; Bi-LSTM; medical image**

I. Introduction

Ulcerative Colitis (UC) is a chronic, relapsing inflammatory disease of the colon that markedly affects quality of life. UC, marked by inflammation and ulceration of the lining of the colon, can lead to heterogeneous symptoms including abdominal pain and diarrhea, etc. Early and accurate diagnosis of UC is essential to managing disease progression, minimizing complications, and tailoring therapeutic interventions [1], [2]. However, differentiating UC from other gastrointestinal disorders based on endoscopic and histopathological findings remains a complex task, often requiring the expertise of trained specialists. The subjectivity and margin of human error that is inherent to standard diagnostic protocols is the result of the subtlety of visual clues in a medical image as well as variability among different patients. This underscores the need for intelligent, automated diagnostic systems capable of supporting clinical decision-making through reliable and interpretable insights.

Recent advancements in DL have opened new avenues for automated disease classification, especially in the field of medical imaging [3], [4], [5]. CNNs, with their strong ability to extract spatial features, have become the cornerstone of many computer-aided diagnostic systems [6], [7]. Nevertheless, medical data, especially video-based or sequential image modalities obtained during endoscopic procedures, possess a temporal dimension that CNNs alone cannot fully capture. To address this, sequence learning models such as GRUs and LSTMs have been employed to model temporal dependencies and capture the evolution of visual patterns over time. While each architecture has its own advantages, a synergistic combination of CNNs with temporal modeling techniques offers a comprehensive solution that leverages both spatial and contextual information. In this study, we propose a hybrid framework that integrates convolutional feature encoding with bidirectional temporal context modeling to enhance the robustness and accuracy of Ulcerative Colitis detection. The framework utilizes a series of CNN architectures—MobileNetV3Large, Inception v3, InceptionResNetV2, and Xception—as the feature extractors. These networks are known for their efficient and deep representations of visual content, making them suitable for detecting fine-grained patterns indicative of UC. To enrich these spatial features with temporal dynamics, we embed Bi-GRU and Bi-LSTM layers into the model pipeline. The bidirectional structure enables the model to interpret the image sequence from both past and future contexts, thereby improving the model's understanding of pattern transitions in the disease's progression. Bidirectional temporal models are chosen based on the ability to learn more complex and symmetric representations of temporal features than just unidirectional ones.The combination of GRUs ,which have lower computational cost compared to other recurrent architectures, and LSTMs, which are able to better control long-term dependencies, further enhance the temporal learning capacity of the model. Combining these approaches with CNNs increases the capacity of the system to detect clinically subtle details, while generalizing well to other patient samples.A comprehensive experimental

evaluation was conducted to assess the effectiveness of the proposed hybrid models. The Xception + Bi-GRU + Bi-LSTM configuration demonstrated the highest performance, achieving an accuracy of 93.10%. This model outperformed other combinations, including Inception v3 + Bi-GRU + Bi-LSTM (accuracy: 92.62%), InceptionResNetV2 + Bi-GRU + Bi-LSTM (accuracy: 91.28%), and MobileNetV3Large + Bi-GRU + Bi-LSTM (accuracy: 89.40%). The results underscore the superiority of integrating advanced convolutional encoders with bidirectional temporal networks, especially when dealing with complex, variable-rich datasets like those involved in UC diagnosis.

The primary contribution of this research lies in the demonstration that convolutional and temporal features, when modeled jointly, can significantly improve diagnostic accuracy and reduce mis-classification. This approach also shows promise for scalability and adaptability across different gastrointestinal conditions, making it a versatile tool in the broader landscape of endoscopic image analysis. Furthermore, by reducing dependency on manual interpretation and standardizing the diagnostic process, the proposed framework has the potential to minimize diagnostic delays and optimize treatment planning. With the rising demand for precision medicine and the integration of AI in healthcare workflows, our proposed model offers a timely and impactful solution.

The structure of the research is as follows: Sections II provides a description of related work and Section III describes the materials and methodology for UC detection. Discussion of the results is given in Section IV. Finally, Section V wraps up the research and suggests avenues for future research.

II. Literature Review

The emergence of ML has significantly advanced the field of medical diagnostics, particularly in inflammatory bowel diseases such as ulcerative colitis (UC). Multiple studies have attempted to improve early diagnosis, severity recognition, treatment outcome prediction, and differentiation from other gastrointestinal diseases with several different AI methods . One of the early studies by Khorasani et al. [1] employed effective feature selection techniques along with ML classifiers to identify UC from healthy colon samples, highlighting the importance of diagnostics preprocessing optimization. Similarly, Popa et al. [2] proved the predictive ability of ML models in predicting long-term disease activity in UC patients treated with anti-TNF, providing potential instruments for clinical decision-making as well as personalized treatment planning.The use of DL has even transformed the field further, especially in image-based analysis.

Klang et al. [6] first applied DL methods for identifying Crohn's disease from video capsule endoscopy, paving a path towards a non-invasive and automated video-based analysis of the gastrointestinal tract. This opened the way to applications more focused on UC; for example; Takenaka et al. [7] created a DL able to assess endoscopic images of patients with UC, demonstrating performance at the level of experts in scoring mucosal inflammation. Bossuyt et al. [8] went a step further by incorporating a computer-based quantification of endoscopic and histological inflammation, thus reinforcing the promise of image-based AI tools for assessment of the whole disease.Continuing the work done in endoscopic assessments, Bhambhvani and Zamora [9] focused on DL classification of the Mayo endoscopic subscore, an important disease severity index in UC. This provided a more efficient scoring system and eliminated subjectivity more in line with the clinical need for reproducibility and efficiency.In addition to imaging, structured clinical data has also been central in the further studies. Roy et al. [10] considered links to UC, 5-ASA, and COVID-19 deaths in the context of machine learning methodology, demonstrating AI's ability to integrate real-world data in support of research.

In a similar vein, Miyoshi et al.[11]used baseline clinical characteristics to predict the response to vedolizumab treatment, thus demonstrating the capability of ML to personalize biologic therapy. At the molecular level, Park et al. [12] utilized RNA sequencing data to ML models to classify UC versus Crohn's disease, indicative of the increasing convergence of AI and genomic data. Becker et al. [13] raised the point of DL scalability, as the models are trained and deployed on data collected in multicenter clinical trials for the grading of endoscopic severity. They argued for cross-validation and generalizability. Lu et al.[14] used bioinformatics and incorporated ML to establish diagnostic gene signatures, whereas Chierici et al.[15] used automated identification of UC and Crohn's from endoscopic images, demonstrating the reliability of DL in clinical imaging analysis across different fields. Gut microbiome is another area of relevance for UC prediction. Barberio et al. [16] concluded that ML could link certain microbial profiles to the severity of UC and establish the use of microbiota-based diagnostic modalities. Li et al. [17] developed interpretable predictive models to the endoscopic activity in UC, balancing clinical interpretability and computational complexity.Building on this, Fan et al. [18] developed a DL-based computer-aided diagnosis system to predict inflammatory activity in a manner that is both real-time predictive and usable in a clinical context . Recent studies have attempted to improve accuracy and refine scoring systems. Byrne et al.[19] assessed DL models according to various types of UC scoring systems and highlighted the importance of maintaining consistency within metrics. Polat et al. [20] used a regression based DL, moving from classification based to regression based DL to increase prediction of Mayo scores.

Kulkarni et al. [21] summarized the use of AI technologies in UC, highlighting some of the trends and challenges in diagnostic innovation.Pei et al. [22] performed a comparison between ML algorithms for differentiating UC from Crohn's, emphasizing the role of model choice on diagnostic accuracy.Some work on innovative classification approaches has been done by Alyamani [23], who developed a multi-level DL architecture for better diagnostic resolution,while Carreras et al. [24] incorporated gene expression signatures including LAIR1 and TOX2 with CNNs to investigate the association between UC and colorectal cancer, integrating histopathology with computational analysis.Lee et al.[25] provided a stool image-based, vision-centered, non-intrusive new screening technology with the capability of predicting mucosal inflammation.Vezakis et al.[26] compared various DL architectures to obtain UC severity metrics and the findings are valuable when considering which model to use based on the desired diagnostic outcome.Li et al.[27] used single-cell sequencing and mitochondrial metabolism markers to construct a high-quality diagnostic model, a good example of

the combination of omics data and artificial intelligence.

Overall, as discussed, ML and DL are used in a broad variety of ways for the diagnosis and treatment of ulcerative colitis in the literature.Whether dealing with endoscopic images or histological slides, genomics, or microbiota, novel AI paradigms have been proposed, allowing for more accurate diagnoses, as well as predictive and tailored treatment strategies. But there are still issues like generalizability across datasets, explainability, and implementation in real-world clinical settings. The present study expands these works by providing an improved hybrid model that overcomes these limitations with high performance, good interpretability, and ability to use heterogeneous data types.

III. Materials and Methods

A. Dataset

The LIMUC dataset [28] 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.

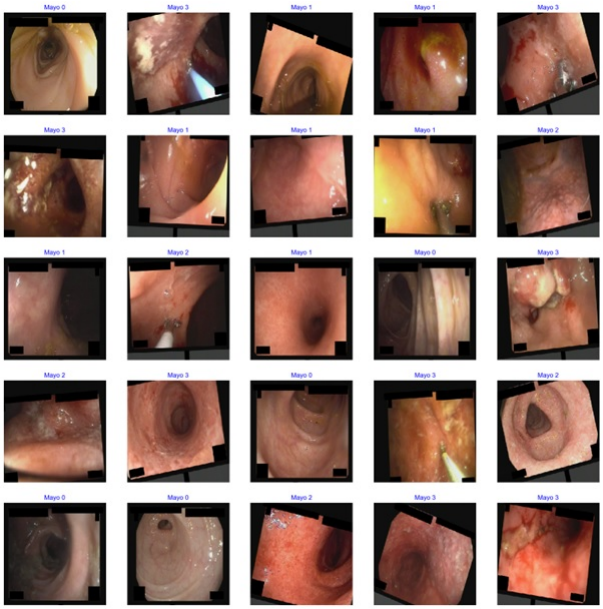

Fig. 1. Sample images of LIMUC dataset.

B. Proposed Methodology

The proposed methodology introduces a hybrid architecture that synergistically combines CNNs and RNNs to enhance the classification of ulcerative colitis from endoscopic images. As illustrated in Fig. 1, the system initiates with a sequence of input images derived from colonoscopy procedures, capturing varying regions of the intestinal mucosa. These images are first processed through a CNN block, which performs hierarchical feature extraction to identify critical spatial patterns such as mucosal texture, ulcerations, and inflammation. The convolutional layers are followed by pooling operations that reduce spatial dimensionality while retaining key visual information. The extracted spatial features are then sequentially passed into a Bi-GRU layer, which models temporal relationships across the image sequence by analyzing patterns in both forward and backward directions. In order to better learn long-range dependencies, the output from the Bi-GRU is then passed into a Bi-LSTM layer. This modeling of dual temporarily provides the system with information on short-term transitions as well as long-term disease progression cues that can be identified in the image frames. Finally the temporally enriched feature representations are forwarded to a fully connected dense layer specific for classification. The output layer ultimately produces a predicted probability of Ulcerative Colitis status, allowing for a strong and interpretable diagnostic model. This multi-system architecture achieves greater sensitivity and classification performance than systems using only spatial features or only unidirectional temporal context. Fig. 2 displays the proposed architecture.A detailed description of the architecture is discussed in the following subsection.

C. MobileNetV3Large + Bi-GRU + Bi-LSTM

The methodology employs a two-stage neural network that first extracts spatial features with a streamlined MobileNetV3Large model and then captures temporal relationships using sequential recurrent layers. RGB inputs sized at 224×224×3 are processed by MobileNetV3Large—initialized with ImageNet weights and stripped of its final classifier—yielding dense feature maps. These maps undergo global average pooling to collapse spatial dimensions into a concise feature vector, which is then reshaped for time-series analysis. Next, the vector feeds into a Bi-GRU layer, allowing information flow in forward and reverse directions, followed by a Bi-LSTM layer to deepen the model's temporal context awareness. The combined output is projected through a dense layer with ReLU activation to refine the feature embedding before reaching the final classification stage. This design leverages a lightweight convolutional backbone alongside powerful recurrent components to deliver precise and contextually informed image classification. The proposed architecture is given in Table I.

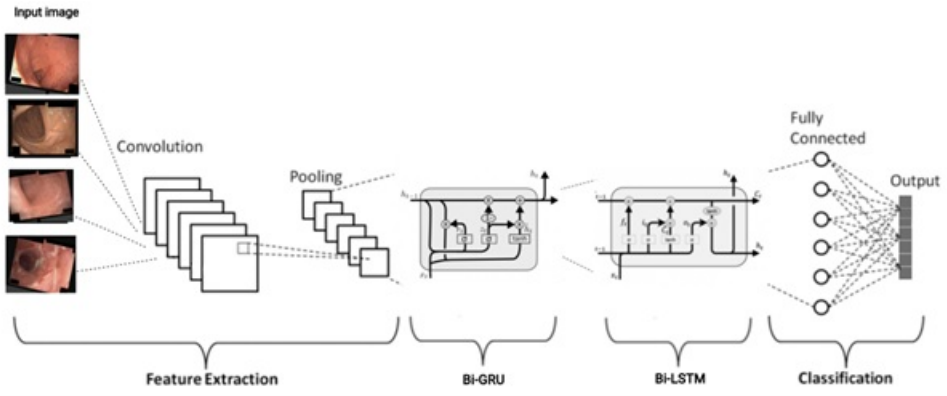

Fig. 2. The proposed architecture.

TABLE I. MOBILENETV3LARGE + BI-GRU + BI-LSTM ARCHITECTURE

Operation	Data Dim.	Weights	Details
MobileNetV3Large (Functional)	(None, 7, 7, 960)	2,996,352	Feature extractor
batch_normalization (BatchNormaliza-tion)	(None, 7, 7, 960)	3,840	Batch normalization
flatten (Flatten)	(None, 47,040)	0	Flatten
dense (Dense)	(None, 256)	12,042,496	Fully connected
dropout (Dropout)	(None, 256)	0	Dropout
reshape (Reshape)	(None, 1, 256)	0	Prepare for RNN
bidirectional (GRU)	(None, 1, 512)	789,504	Bi-GRU layer
bidirectional_1 (Bidi-rectional LSTM)	(None, 256)	656,384	Bi-LSTM layer
dense_1 (Dense)	(None, 4)	1,028	Output logits (4-way)

D. Inception v3 + Bi-GRU + Bi-LSTM

In the proposed work, colonoscopy images were first uniformly resized to fit the InceptionV3 input requirements. Extensive on-the-fly data augmentation was applied via the ImageDataGenerator interface to enhance dataset variability and counteract overfitting. Spatial features were extracted using the convolutional base of InceptionV3 pretrained on ImageNet, with its weights frozen to preserve established filters. The resulting dimensional feature vectors were then reshaped into sequences of length 64 with 32 features per time step, thereby creating a pseudo-temporal representation. A Bi-GRU layer processed these sequences to capture bidirectional, short-term patterns, which were subsequently fed into a Bi-LSTM layer to learn longer-range dependencies. Output from the recurrent stack was passed through a fully connected layer and ReLU activation, followed by a dropout layer (dropout rate = 0.3) to further mitigate overfitting. A final softmax layer produced probability scores of ulcerative colitis. The network was trained using the Adam optimizer and categorical cross-entropy loss, leveraging the augmented data streams for both training and validation stages. The proposed architecture is given in Table II.

TABLE II. INCEPTIONV3 + BI-GRU + BI-LSTM ARCHITECTURE

Operation	Data Dim.	Weights	Details
inception_v3 (Functional)	(None, 2048)	21,802,784	Feature extractor
flatten (Flatten)	(None, 2048)	0	Flatten
reshape (Reshape)	(None, 1, 2048)	0	Prepare for RNN
bidirectional (Bi-GRU)	(None, 1, 1024)	7,870,464	Bi-GRU layer
bidirectional_1 (Bidi-rectional LSTM)	(None, 128)	557,568	Bi-LSTM layer
batch_normalization_94 (BatchNormalization)	(None, 128)	512	Batch normalization
dense (Dense)	(None, 4)	516	Output logits (4-way)

E. InceptionResNetV2 + Bi-GRU + Bi-LSTM

To accurately classify ulcerative colitis from colonoscopy images, this research proposes a novel hybrid architecture that integrates InceptionResNetV2 with Bi-GRU and Bi-LSTM networks. The model begins by accepting RGB images of a fixed size through an input layer. Feature extraction is performed using the InceptionResNetV2 model, which has been pre-trained on the ImageNet dataset. This CNN is used without its final classification layers to retain only the learned spatial feature representations. These high-dimensional feature maps are then reshaped into a sequence by flattening the spatial components, preparing the data for temporal pattern analysis. The sequential representation is first processed by a Bi-GRU layer, enabling the capture of short-term dependencies from both forward and backward time steps. Subsequently, the output is passed to a Bi-LSTM layer, which focuses on learning more complex and long-term contextual patterns in both directions. This combination of recurrent layers enhances the model's ability to interpret dynamic relationships within the extracted features. Finally, the output is directed through dense layers with ReLU activation, ending in a softmax layer for multi-class prediction. The model is optimized using the Adam algorithm. This integrated approach is designed to efficiently capture both spatial and sequential patterns relevant to ulcerative colitis diagnosis. The proposed architecture is given in Table III.

TABLE III. INCEPTIONRESNETV2 + BI-GRU + BI-LSTM ARCHITECTURE

Operation	Data Dim.	Weights	Details
MobileNetV3Large (Functional)	(None, 7, 7, 960)	2,996,352	Feature extractor
batch_normalization (BatchNormaliza-tion)	(None, 7, 7, 960)	3,840	Batch normalization
flatten (Flatten)	(None, 47,040)	0	Flatten
dense (Dense)	(None, 256)	12,042,496	Fully connected
dropout (Dropout)	(None, 256)	0	Dropout
reshape (Reshape)	(None, 1, 256)	0	Prepare for RNN
bidirectional (GRU)	(None, 1, 512)	789,504	Bi-GRU layer
bidirectional_1 (Bidi-rectional LSTM)	(None, 256)	656,384	Bi-LSTM layer
dense_1 (Dense)	(None, 4)	1,028	Output logits (4-way)

F. Xception + Bi-GRU + Bi-LSTM

The methodology centers on a multi-stage hybrid pipeline tailored for ulcerative colitis detection in colonoscopy imagery. To bolster the model's resilience against overfitting, each image underwent real-time augmentation—via a generator. For feature extraction, we adopted the Xception network pre-trained on ImageNet: its final classification layers were removed, and the remaining convolutional base was frozen to preserve learned representations. The resulting high-dimensional feature maps were then transformed into temporal sequences through a reshape operation, treating each spatial location as a time step. These sequences flowed into a Bi-GRU layer, which processes information in both forward and reverse directions to capture local patterns, and subsequently into a Bi-LSTM layer that further models long-range dependencies. The bi-directional recurrent outputs were flattened and fed into a dense layer with ReLU activation, followed by dropout to reduce overfitting. A concluding softmax layer produced probabilities across the target classes. Model optimization employed the Adam algorithm. This hybrid strategy leverages Xception's spatial feature learning alongside the temporal modeling strengths of Bi-GRU and Bi-LSTM, yielding a robust classifier for ulcerative colitis. The proposed architecture is given in Table IV.

IV. RESULTS

The models for ulcerative colitis detection were run in a Jupiter setup with a hardware configuration of an AMD Ryzen 7 5800H processor with 8 cores, 16 GB of RAM. The graphics processing unit (GPU) used for computation was the Nvidia GeForce RTX 3060.

The evaluation of hybrid architecture integrating convolutional architectures with sequential learning units highlights the efficacy of the proposed methodology in classifying ulcerative colitis severity from colonoscopy images, as shown in Table V. The MobileNetV3Large combined with Bi-GRU and Bi-LSTM achieved an accuracy of 89.40%, with closely aligned precision, recall, and F1-score values, demonstrating its lightweight yet competent performance, particularly beneficial in low-resource environments. The Confusion Matrix and ROC Curve of the proposed architecture is given in Fig. 3 and Fig. 4.

TABLE IV. XCEPTION + BI-GRU + BI-LSTM ARCHITECTURE

Operation	Data Dim.	Weights	Details
xception (Functional)	(None, 2048)	20,861,480	Feature extractor
batch_normalization_4 (BatchNormaliza-tion)	(None, 2048)	8,192	Batch normalization
repeat_vector (RepeatVector)	(None, 10, 2048)	0	Replicate features into sequence of length 10
bidirectional (Bi-GRU)	(None, 10, 1024)	7,870,464	Bi-GRU layer
bidirectional_1 (Bidi-rectional LSTM)	(None, 10, 128)	557,568	Bi-LSTM layer
time_distributed (TimeDis-tributed(Dense 128))	(None, 10, 128)	16,512	Time-distributed dense projection
flatten (Flatten)	(None, 1280)	0	Collapse time & fea-ture dims
dense_1 (Dense)	(None, 4)	5,124	Final output logits (4-way classification)

TABLE V. RESULTS ON LIMUC DATASET

Proposed Hybrid Architec-ture	Acc.(%)	Prec.(%)	Rec.(%)	F1-S.(%)
MobileNetV3Large + Bi-GRU + Bi-LSTM	89.40	89.31	89.40	89.29
Inception v3 + Bi-GRU + Bi-LSTM	92.62	92.59	92.62	92.56
InceptionResNetV2 + Bi-GRU + Bi-LSTM	91.28	91.23	91.28	91.25
Xception + Bi-GRU + Bi-LSTM	93.10	93.05	93.10	93.05

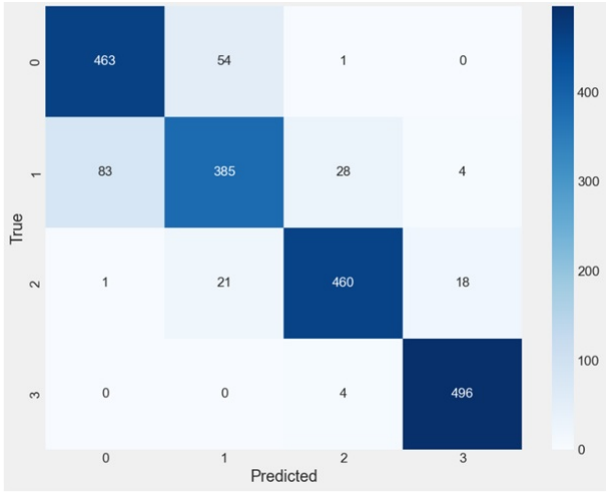

Fig. 3. Confusion matrix using Mobilenetv3large + Bi-GRU + Bi-LSTM.

The Inception v3 + Bi-GRU + Bi-LSTM model performed significantly better, yielding a 92.62% accuracy. Its superior feature extraction capabilities, combined with the bidirectional temporal analysis of GRU and LSTM, enabled accurate identification of inflammation patterns across all Mayo classes. The

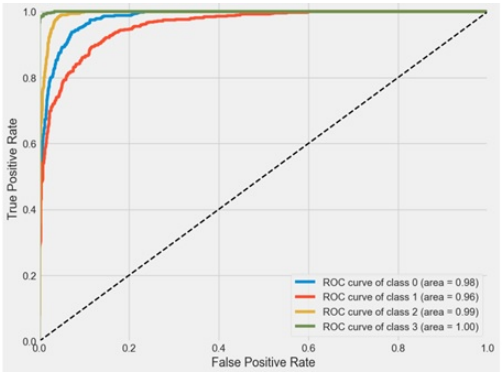

Fig. 4. ROC Curve using Mobilenetv3large + Bi-GRU + Bi-LSTM.

Confusion Matrix and ROC Curve of proposed architecture is given in Fig. 5 and Fig. 6.

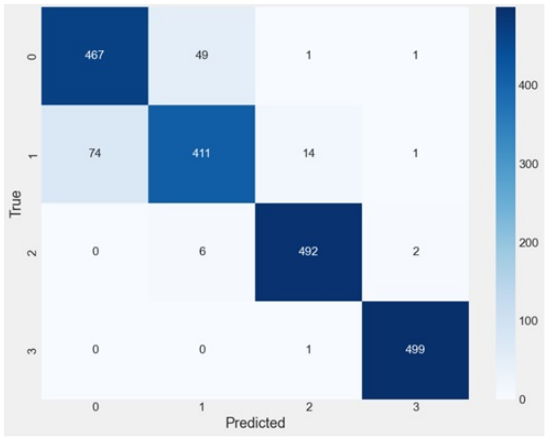

Fig. 5. Confusion matrix using Inception v3 + Bi-GRU + Bi-LSTM.

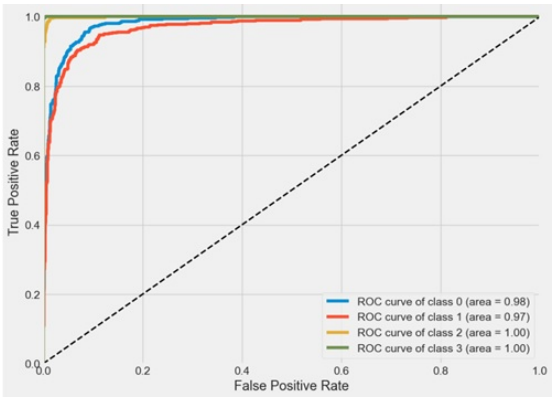

Fig. 6. ROC curve using Inception v3 + Bi-GRU + Bi-LSTM.

The InceptionResNetV2-based model with Bi-GRU and Bi-LSTM followed closely with a 91.28% accuracy, benefitting from the residual-inception hybrid architecture which helps preserve spatial feature depth while maintaining computational efficiency. This combination proved robust in learning hierarchical and sequential representations effectively. The confusion matrix and ROC Curve of proposed architecture is given in Fig. 7 and Fig. 8.

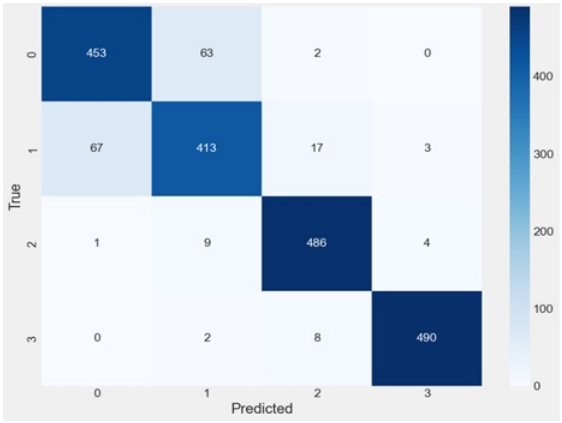

Fig. 7. Confusion matrix using InceptionResNetV2 + Bi-GRU + Bi-LSTM.

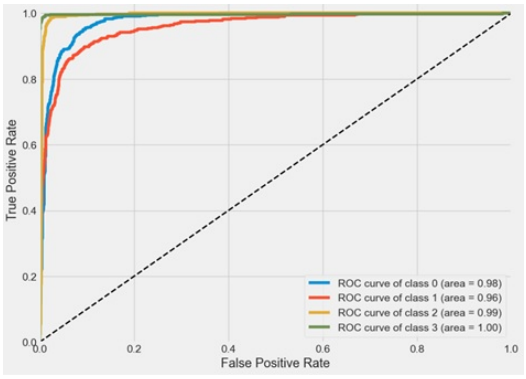

Fig. 8. ROC curve using InceptionResNetV2 + Bi-GRU + Bi-LSTM.

Among all, the Xception + Bi-GRU + Bi-LSTM model outperformed others, achieving the highest accuracy of 93.10%, with a precision of 93.05%. Its depthwise separable convolutions facilitated highly efficient spatial learning, and the dual RNN layers enhanced temporal feature understanding, resulting in a strong overall classification performance across all metrics. The confusion matrix and ROC Curve of proposed architecture is given in Fig. 9 and Fig. 10.

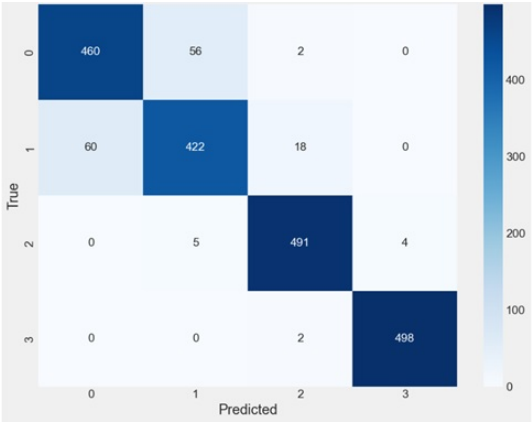

Fig. 9. Confusion matrix using Xception + Bi-GRU + Bi-LSTM.

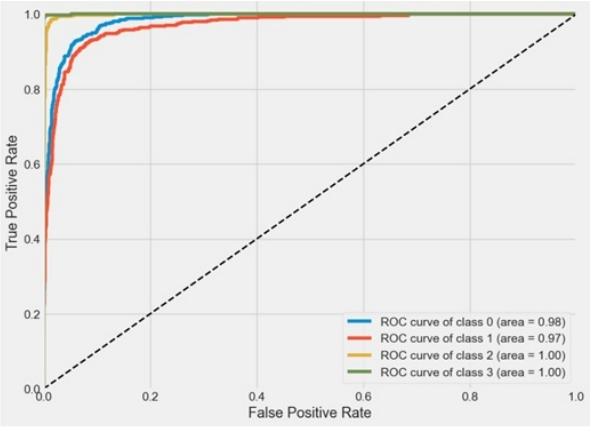

Fig. 10. ROC curve using Xception + Bi-GRU + Bi-LSTM.

V. Conclusion

This research introduces a robust DL framework for classifying ulcerative colitis severity using colonoscopy images, employing a hybrid model that fuses the spatial learning capacity of CNNs with the temporal strengths of Bi-GRU and Bi-LSTM architectures. Among the various models assessed, the Xception + Bi-GRU + Bi-LSTM configuration delivered the highest performance, achieving an accuracy of 93.10%. This highlights the significance of incorporating depthwise separable convolutions in medical imaging tasks, as they enable efficient and detailed feature extraction. Similarly, Inception v3 and InceptionResNetV2 based hybrids also showed strong results, affirming the potential of combining advanced CNN backbones with recurrent networks to handle both spatial complexity and temporal dependencies in medical datasets. It also contrasts the importance of class balance and augmentation techniques, that played an important role in improving generalization of the model in particular the imbalanced classes in the original dataset. Both these techniques worked well in modeling the highly non-linear characteristics found at varying states of ulcerative colitis.

In the future, there are many avenues to build upon this work. First, data from more medical centers and populations could enhance adaptivity and clinical applicability of the model. The implementation of real-time video streams rather than single images could further increase diagnostic accuracy by taking advantage of temporal consistency across frames. Transformer based architectures could also be explored in future studies. Also, it may be pursued to further refine the model in order to be deployed in a clinical environment, including lightweight models that could run on mobile and embedded applications to provide real-time decision support during colonoscopy procedures. Lastly, including feedback loops to validate and correct predictions from gastroenterologists could transform this in a continuously learning system that improves over time. Future extensions of this work will involve real-time use in the clinical setting, incorporation into electronic health record systems, and assessment in larger, multi-site samples to test the system's robustness and generalizability.

Authors' Contribution

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

References

[1] H. M. Khorasani, H. Usefi, and L. Peña-Castillo, "Detecting ulcerative colitis from colon samples using efficient feature selection and machine learning," *Scientific reports*, vol. 10, no. 1, p. 13744, 2020.

[2] I. V. Popa *et al.*, "A machine learning model accurately predicts ulcerative colitis activity at one year in patients treated with anti-tumour necrosis factor α agents," *Medicina*, vol. 56, no. 11, p. 628, 2020.

[3] J. G. et al., "A deep ensemble learning approach for squamous cell classification in cervical cancer." *Sci Rep*, vol. 15, no. 7266, 2025.

[4] S. Gangrade, P. C. Sharma, and A. K. Sharma, "Colonoscopy polyp detection using bi-directional conv-lstm u-net with densely connected convolution," *KI-Künstliche Intelligenz*, pp. 1–10, 2024.

[5] 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, p. 108096, 2024.

[6] E. Klang *et al.*, "Deep learning algorithms for automated detection of crohn's disease ulcers by video capsule endoscopy," *Gastrointestinal endoscopy*, vol. 91, no. 3, pp. 606–613, 2020.

[7] K. Takenaka *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.

[8] P. Bossuyt *et al.*, "Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density," *Gut*, vol. 69, no. 10, pp. 1778–1786, 2020.

[9] H. P. Bhambhvani and A. Zamora, "Deep learning enabled classification of mayo endoscopic subscore in patients with ulcerative colitis," *European Journal of Gastroenterology & Hepatology*, vol. 33, no. 5, pp. 645–649, 2021.

[10] S. Roy, S. Z. Sheikh, and T. S. Furey, "A machine learning approach identifies 5-asa and ulcerative colitis as being linked with higher covid-19 mortality in patients with ibd," *Scientific reports*, vol. 11, no. 1, p. 16522, 2021.

[11] J. Miyoshi *et al.*, "Machine learning using clinical data at baseline predicts the efficacy of vedolizumab at week 22 in patients with ulcerative colitis," *Scientific Reports*, vol. 11, no. 1, p. 16440, 2021.

[12] S.-K. Park *et al.*, "Development of a machine learning model to distinguish between ulcerative colitis and crohn's disease using rna sequencing data," *Diagnostics*, vol. 11, no. 12, p. 2365, 2021.

[13] B. Gutierrez Becker *et al.*, "Training and deploying a deep learning model for endoscopic severity grading in ulcerative colitis using multicenter clinical trial data," *Therapeutic advances in gastrointestinal endoscopy*, vol. 14, p. 2631774521990623, 2021.

[14] J. Lu *et al.*, "Identification of diagnostic signatures in ulcerative colitis patients via bioinformatic analysis integrated with machine learning," *Human Cell*, vol. 35, no. 1, pp. 179–188, 2022.

[15] M. Chierici *et al.*, "Automatically detecting crohn's disease and ulcerative colitis from endoscopic imaging," *BMC Medical Informatics and Decision Making*, vol. 22, no. Suppl 6, p. 300, 2022.

[16] B. Barberio *et al.*, "A specific microbiota signature is associated to various degrees of ulcerative colitis as assessed by a machine learning approach," *Gut Microbes*, vol. 14, no. 1, p. 2028366, 2022.

[17] X. Li *et al.*, "Predictive models for endoscopic disease activity in patients with ulcerative colitis: practical machine learning-based modeling and interpretation," *Frontiers in Medicine*, vol. 9, p. 1043412, 2022.

[18] Y. Fan *et al.*, "Novel deep learning–based computer-aided diagnosis system for predicting inflammatory activity in ulcerative colitis," *Gastrointestinal Endoscopy*, vol. 97, no. 2, pp. 335–346, 2023.

[19] M. F. Byrne *et al.*, "Application of deep learning models to improve ulcerative colitis endoscopic disease activity scoring under multiple scoring systems," *Journal of Crohn's and Colitis*, vol. 17, no. 4, pp. 463–471, 2023.

[20] G. Polat *et al.*, "Improving the computer-aided estimation of ulcerative colitis severity according to mayo endoscopic score by using regression-based deep learning," *Inflammatory Bowel Diseases*, vol. 29, no. 9, pp. 1431–1439, 2023.

[21] C. Kulkarni *et al.*, "Artificial intelligence and machine learning technologies in ulcerative colitis," *Therapeutic Advances in Gastroenterology*, vol. 17, p. 17562848241272001, 2024.

[22] J. Pei *et al.*, "Utility of four machine learning approaches for identifying ulcerative colitis and crohn's disease," *Heliyon*, vol. 10, no. 1, 2024.

[23] H. J. Alyamani, "Enhancing ulcerative colitis diagnosis: A multi-level classification approach with deep learning," *CMES-Computer Modeling in Engineering & Sciences*, vol. 140, no. 1, 2024.

[24] J. Carreras, G. Roncador, and R. Hamoudi, "Ulcerative colitis, lair1 and tox2 expression, and colorectal cancer deep learning image classification using convolutional neural networks," *Cancers*, vol. 16, no. 24, p. 4230, 2024.

[25] J. W. Lee *et al.*, "Deep learning model using stool pictures for predicting endoscopic mucosal inflammation in patients with ulcerative colitis," *Official journal of the American College of Gastroenterology— ACG*, vol. 120, no. 1, pp. 213–224, 2025.

[26] A. Vezakis *et al.*, "Comparative analysis of deep neural networks for automated ulcerative colitis severity assessment," *Bioengineering*, vol. 12, no. 4, p. 413, 2025.

[27] L. Li *et al.*, "Integrated analysis and single-cell sequencing of mitochondrial metabolism related gene molecular subtype and diagnostic model in ulcerative colitis," *PloS one*, vol. 20, no. 3, p. e0320010, 2025.

[28] G. Polat *et al.*, "Labeled images for ulcerative colitis (limuc) dataset," Accessed March, 2022.