SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I6P105

Volume 13 Issue 6, 50-64, June 2026

Original Article

Ulcerative Colitis Detection and Severity Prediction Using a Hybrid Deep Learning Model

SumedhVithalrao Dhole[1,2,4*], Chetan S. More[2], Anuradha Sagar Nigade[2], Sonali Kishore Pawar[2], A.Y. Prabhakar[2], Sangeeta Rajendra Chougule[3]

[1]*Department of Electronics and Telecommunication Engineering, Kolhapur Institute of Technology's College of Engineering (Autonomous), Kolhapur, Maharashtra, India.*
[2]*Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune. Maharashtra, India.*
[3]*Director, Ashokrao Mane Group of Institutions, Vathar, Kolhapur, Maharashtra, India.*
[4]*Shivaji University, Kolhapur, Maharashtra, India.*

[*]*Corresponding Author : svdhole@bvucoep.edu.in*

Received: 07 March 2026 Revised: 06 April 2026 Accepted: 05 May 2026 Published: 27 June 2026

Abstract - *Ulcerative Colitis Detection and Severity Prediction research creates an effective and dependable automated model to identify the severity of colon diseases using Wireless Capsule Endoscopy (WCE) images. Current methods typically use single deep learning models or conventional machine learning models, which do not readily model both fine-grained variations in mucosal texture and global contextual interactions, particularly when applied to small medical data sets. The improvements of generalization were done by data augmentation and training of the model (categorical cross-entropy loss) with optimized hyperparameters. It was applied to the Python platform with the deep learning libraries and tested on the WCE Curated Colon Disease Dataset, comprising 800 images and four severity levels. The suggested method had a precision of 97.5, which is a better performance than the current models. This system has advantages because it offers accurate diagnosis with computer-assisted assistance to gastroenterologists and aids in the early and accurate evaluation of the severity of UC. The proposed model uniquely combines- Local convolutional features via ResNet-50, Global contextual features via Vision Transformer, and Handcrafted clinical texture descriptors (GLCM + LBP). This multi-source feature fusion is reduced via PCA to preserve 95% variance and addresses the core limitations of single-architecture models that tend to either underfit local texture patterns or miss long-range spatial dependencies.*

Keywords - *Ulcerative Colitis, ResNet-50, Vision Transformer, Colonoscopy Image Classification, Deep Learning.*

1. Introduction

Ulcerative Colitis (UC) is a gastrointestinal disease that affects the innermost layer of the large intestine and causes severe pain, diarrhoea, and fatigue. It has dynamic symptoms, so it's essential to assess the severity of the disease at a suitable time for treatment. Recently, deep learning has been used for UC images, where CNNs can learn subtle visual features but need to concentrate only on local areas.

Transformer-based models can focus more on the general visual relationships of UC images, but may not perform well in identifying fine visual patterns and often require large amounts of data. Some previous studies also used handcrafted features, which can still provide valuable information for clinical use. To solve these issues in research gaps, the proposed model uses a hybrid model that combines CNNs, transformers, and selected handcrafted features for more reliable, accurate, and interpretable results with clinical support.

UC is a chronic IBD that affects the colon and the rectum, leading to persistent inflammation, ulceration, and abdominal discomfort [1]. UC is more characterized by an episodic and remission-like pattern, and therefore, effective treatment and management of the disease are dependent on appropriate diagnosis and prediction of severity. Misdiagnosed UC can progress to severe complications like colon perforation, toxic megacolon, or the risk of colorectal cancer. Conventional diagnostic strategies like colonoscopy and histopathological examinations are time-consuming, observer-dependent, and operator-dependent [2]. Given the increasing global prevalence of UC, particularly among young individuals, an urgent need for secure, automated, and intelligent platforms that facilitate early detection and severity grading. Through medical imaging and AI, especially DL methods, there is the potential to improve the scalability of UC diagnostics [3].

In recent years, numerous research studies have employed CNN in the automatic detection and classification of gastrointestinal diseases like Ulcerative Colitis. Image-based disease classification is a common use of CNNs. [4] Nonetheless, most of the current methods are constrained in their capacity to pick up on the long-distance spatial dependencies and global context that occur in colonoscopy or histopathology images [5]. Such models tend to fail with the uncertainty in image quality, non-uniform illumination, and the inconspicuousness of UC lesions. In addition, most methods rely exclusively on learned features without incorporating handcrafted features related to the domain that might improve interpretability and accuracy [6]. There are models that have been fairly accurate but are not robust and generalizable across varied datasets. Furthermore, severity grading as a vital part of disease treatment is not studied or handled correctly. The existence of these constraints highlights the need to have more integrated and hybridized designs, which can exploit the advantages of different learning processes. The limitations of the existing methods can be mitigated by a framework that integrates local and global feature extraction and clinical or handcrafted features, which can provide a superior understanding of UC patterns.

To address these challenges, a model will be created that integrates CNN and ViT with hand-crafted feature extraction algorithms to provide trustworthy outcomes in the diagnosis of UC and the prediction of its severity. The CNN section, which is the adaptation of ResNet-50, is capable of retrieving local and textural features of the mucosal textures and ulcerations. Furthermore, handcrafted features like intensity scales, texture dimensions, and colour histograms were introduced to the model to increase interpretability and clinical significance. A combination of these various features is produced through a learnable fusion layer to produce the rich feature representation to be used in classification. The WCE Curated Colon Disease Dataset was used to train and validate, and is a huge publicly accessible gastrointestinal image dataset on Kaggle. It contains over 800 endoscopic pictures of high resolution with UC cases of various severity. Our system is very accurate in the presence of UC and very accurate in predicting the severity of the condition. The level of improvement of the model performance of multi-perspective data fusion is very high, hence a strong solution to practical clinical applications in real life. This study will combine the existing gap between automated diagnosis and clinical usefulness with a combination of advanced machine learning and human-interpretable features.

1.1. The Most Important Contribution of the Research

- Produced a new hybrid deep learning model, which involved CNNs, Vision Transformers, and hand-crafted clinical features to promptly detect and forecast the severity of Ulcerative Colitis.
- Used a dual-branch pipeline that added Local features of the ResNet-50 and global features of a Vision Transformer (ViT), and fused and decoded those features to enhance the diagnostic features.
- Used publicly available WCE Curated Colon Disease Dataset. This data is comprised of various labelled endoscopic images, currently also containing images of Ulcerative Colitis and severity marking.

2. Related Works

Qiu et al. [7] proposed a new MIL model, VIGIL, a combination of endoscopy on white light and endocytic pictures and the diagnostic report text to predict the histological healing of UC patients. The model achieved 92.69 percent accuracy and 94.79 percent AUC value and outperformed the already established state-of-the-art techniques. Combining image and text data enables greater semantic understanding, minimizing annotation burdens. However, the model's performance may vary with different data sources, and its generalizability to diverse clinical settings requires further validation.

Syed et al. [8] employed ML techniques in combination with genomic profiling in order to identify meaningful biomarkers to early diagnose inflammatory bowel diseases, including UC. The studies demonstrated the ability of ML to be applied in the discovery of novel biomarkers and the enhancement of early diagnostic accuracy. Despite the fact that some of the levels of accuracy were not disclosed, the method demonstrated a higher level of detection. The constraints are the requirement of huge, changeable datasets in order to stabilize the model and the integration of genomic facts within the clinical environment.

Popa et al. [9] developed an algorithm based on machine learning to predict UC activity following a period of one year of anti-TNF-alpha use. The algorithm demonstrated good predictive capability, which may be useful in personalized planning of treatment. However, the retrospective Nature of the study and reliance on specific treatment data could diminish the generalizability of the study to different treatment scenarios.

Diaconu et al. [10] have covered the AI in IBD in terms of its utility in raising the accuracy of the diagnosis and patient care. The study also mentioned the capabilities of various models of AI, but it also highlighted such problems as data normalization and integration into clinical work. No particular measures of accuracy were provided, which shows the need to further substantiate it empirically.

Li et al. [11] built on images of endocytoscopy to determine histological remission in UC patients. The model had an accuracy of 90%. This process is an alternative to traditional methods of biopsy that are non-invasive. The model may be influenced by image quality and variability between various clinical settings.

Mokhtari et al. [12] designed a loosely-supervised deep learning model that provides the prediction of diseaseassociated features in biopsies of IBD, such as UC. The model may provide an AUC of 0.87 in the classification of Crohn's Disease vs UC and 0.8 0 in the classification of the disease severity. The method improves the interpretability using attention maps, whereas the quality of the labels and the necessity to possess a strong validation can impair it.

Chaitanya et al. [13] proposed a model of a spatio-temporal transformer, called Arges, in an attempt to determine the severity of UC with the help of endoscopy videos. The results of the model showed significant boosts in F1 scores using various severity indices, a fact that demonstrates better performance over the past approaches. However, the model has the potential to be too complicated and computationally expensive to be used in real-time in clinical practice.

Testoni et al. [14] utilized the Xception model when detecting Crohn's Disease erosions and ulcers with 92.4% accuracy and 97.1% precision; the CNNs were used. Despite its intention for UC disease, the application to UC is possible. The need for large annotated datasets and variability of endoscopic image quality constrain it.

Clinically, diagnosis is primarily endoscopic imaging followed by visual grading by experts, which adds variability, latencies, and subjectivity. Although deep learning-based automated detection has seen growing interest, existing models perform poorly when extended to real-world UC datasets. CNN delivers a strong performance in spatial hierarchy extraction but fails to learn long-range dependencies or contextual relationships in medical images. ViTs deliver enhanced contextual comprehension but tend to be data-hungry, computationally demanding, and noise-sensitive in small medical imaging datasets. Additionally, human-crafted features with explainable attributes, like patterns of texture and intensity, are not yet available with deep learning, but their clinical value is already well proven. Most existing methods fail to incorporate handcrafted domain knowledge with automatic feature acquisition, thus leading to underfitting or loss of interpretability.

Table 1. Comparative analysis of methods, results, and research gaps in mayo endoscopic score grading studies

References	Methods Used	Limitations	Research Gap	Key Findings
Lee, J., & Kim, S. 2025. [15]	Multi-task deep learning for MES classification	Requires large annotated datasets; limited explainability	Integration of XAI with multi-task learning	DenseNet121 had the best overall performance; MTL had joint-loss optimization with less underdiagnosis of severe MES stages (MES 2 and 3); short MobileNetv3-large was also performing well, although it is lightweight.
Ozdemir et al. 2025. [16]	Curriculum learning-based deep learning	Training complexity; curriculum dependency	Adaptive curriculum learning for real-world variability	CLoE outperformed all supervised and self-supervised baselines on LIMUC and HyperKvasir; improved robustness to annotation noise; generalized across CNN and Transformer architectures.
Ahmed et al. 2024. [17]	Review of AI in IBD (diagnosis, prognosis, treatment)	No experimental validation	Need for standardized benchmarking frameworks	AI can also emulate expert endoscopic scoring in IBD; NLP and LLMs are becoming more prevalent in text-based data of IBD; multimodal AI methods represent the new frontier.
Tontini et al. 2021. [18]	Systematic review of AI in GI endoscopy	Lack of standardized datasets	Real-world validation needed	In this study, good AI findings were reported; most aimed at UC mucosal activity measurement and capsule endoscopy in Crohn; the first systematic review of AI in endoscopy in IBD; publication bias was recognized.
Stidham et al. 2025. [19]	AI in clinical trial automation	Limited adoption in practice	Integration with regulatory pipelines	AI uses disease endoscopy scoring in clinical trials; LLMs screen patients and collect consents electronically, and extract data; AI image analysis is more detailed than human endoscopy scores; computer vision has the potential to perform better than MES, including new scores such as Cumulative Disease Score.
Sinonquel, P., Eelbode, T., Bossuyt, P., et al., 2021. [20]	AI-assisted endoscopy systems	Variability across systems	Standard evaluation metrics required	AI (deep learning) has systematically increased the rate of polyp/adenoma detection in RCTs; upper GI AI performs as well as experts in Barrett's and gastric neoplasia; capsule endoscopy AI saves on reading time without impeding accuracy; may be

				used to train, evaluate quality, and make therapeutic decisions.
Diaconu, C., State, M., Birligea, M., et al. 2023. [20]	AI for monitoring and prediction	Limited longitudinal validation	Real-time predictive monitoring systems	AI is capable of measuring the endoscopic activity, prognosis predicting, response to it, and neoplasia monitoring in IBD; ML can predict the prognosis after using patient databases, although multi-modal data (endoscopy, histology, biomarkers, and genomics) development is the future direction.
Jacob Broder Brodersen et al. 2024. [21]	AI-assisted analysis using deep learning on pan-enteric capsule endoscopy images/videos	Limited generalizability across devices; potential bias due to dataset size; limited temporal modeling	Need for robust real-time video-based analysis with temporal deep learning and cross-device validation.	DenseNet121 showed the highest overall performance; MTL using joint-loss minimized under the diagnosis of severe MES stages (MES 2 and 3); MobileNet-v3-large was also able to report better performance despite being lightweight.

3. Problem Statement

Ulcerative Colitis (UC) is a chronic inflammatory enteropathy that must be appropriately assessed in terms of severity to make appropriate clinical decisions and treatment choices. Even though some machine learning and deep learning techniques have been proposed to predict UC based on colonoscopic images, the existing techniques are highly flawed.

Traditional machine learning techniques are extremely sensitive to handcrafted features and manual thresholding, which may not always be able to reflect complex spatial variation and subtle mucosa changes at different stages of disease [22]. Even though CNN-based models have better performance in learning features, the current CNN architectures are primarily capable of local texture patterns, and are unable to learn long-range contextual dependencies, leading to misclassifications between neighbouring carvieties such as Mild and Moderate UC [23]. Transformer-based approaches are useful to grasp the global situation, require a huge amount of data, and fail to gather fine-grained local information currently. In addition, the majority of research done to date has an imbalance in classes, poor generalization, and bad interpretability, and no validation through real clinical data. These shortcomings restrict their practical application in the field in a clinical context.

To fill this gap, an improved and powerful framework that is capable of producing both local and global contextual information on colonoscopic pictures with a high level of detail and strong generalization is of utmost importance. This study contends this issue by postulating a hybrid deep learning algorithm that combines the robustness of state-of-the-art convolutional networks and transformer-based frameworks. Through the use of complementary feature representations and a powerful fusion strategy, the proposed solution will improve discrimination among UC severity stages. Moreover, it highlights optimized training, balanced learning, and performance assessment reliability to facilitate accurate, interpretable, and clinically applicable ulcerative colitis severity classification.

4. Methodology of Hybrid ResNet50-ViT + Handcrafted Fusion Model for UC Detection

The recent developments of medical images classification have shown hybrid deep learning architectures like Hybrid-RViT (ResNet-50 + Vision Transformer) and DEMF (EfficientNetV2 + MobileNetV2) that seek to harness the complementary feature representations. Although these models have been found to have a high performance level in the general medical imaging tasks, a number of limitations are present when used to classify the severity of Ulcerative Colitis (UC). First, the majority of current hybrid frameworks use deep feature fusion only without the use of domain-specific handcrafted texture descriptors. Alternatively, the model that has been suggested combines convolutional (ResNet-50), transformer-based (ViT), and handcrafted (GLCM and LBP) features, allowing for the capture of micro-textural irregularities as well as the global contextual dependencies. This multilevel feature representation is especially significant in UC classification, in which subtle differences in mucosal texture are diagnostic. Second, the proposed method is based on feature-level fusion and subsequent dimensionality optimization using PCA as opposed to DEMF and other such architectures, which are mainly concerned with network depth and parameter scaling.

This helps eliminate redundancy and reduce overfitting, which is of great importance in cases of limited medical datasets. Third, the proposed model is explicitly built upon severity-level UC classifications (Normal, Mild, Moderate, Severe), as opposed to binary abnormality detection, which is a more clinically relevant and finer diagnostic issue. Thus, the innovation of the present study is a three-level feature fusion approach, dimensionality optimization, and the focus on application to UC severity grading with the help of a curated WCE dataset.

DATA COLLECTION ⟹ **FEATURE EXTRACTION**

Data preprocessing
↓
Image Resizing
↓
Data Augmentation
↓
Normalization
↓
Dataset splitting

ResNet-50Local Feature Extraction
↓
ViT Global Feature Extraction
↓
Handcrafted Features(GLCM, LBP)
↓
Feature fusion and PCA
↓
Classification and Optimization
↓
UC Classification

Fig. 1 Overall workflow

4.1. Dataset Description

The WCE Curated Colon Disease Dataset is a publicly available, large-scale repository of gastrointestinal videos and images for medical image analysis. It consists of 800 images of polyps, esophagitis, and UC [24].

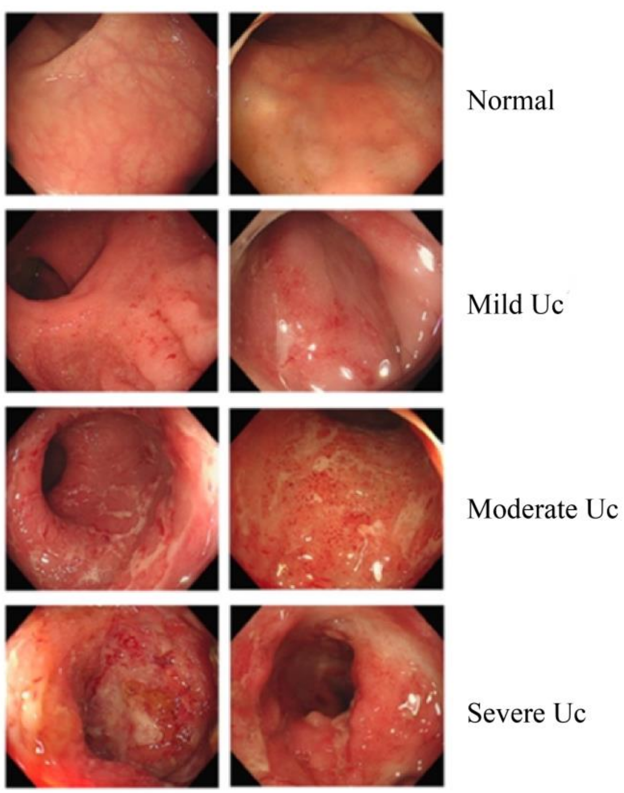

Fig. 2 Sample images from the dataset

The UC-related images are extracted and manually verified, and labelled into 4 clinically relevant classes, as shown in Figure 2.

- Normal
- Mild UC
- Moderate UC
- Severe UC

The data set offers a variety of practical cases that are necessary to have solid UC detection.

4.2. Data Pre-processing

To facilitate better generalization of models, pre-processing of data is critical in standardizing input images. The authors applied several pre-processing techniques described below:

4.2.1. Image Resizing

Images were resized to have a uniform shape of 224 × 224 pixels with 3 colour channels to conform to model input specifications. This provides uniformity to the dataset and enables batch processing in neural networks.

4.2.2. Normalization

Normalization takes the pixel values to the range between 0 and 1, which increases numerical stability during training. It reduces the effect of the various lighting environments and the Contrast of the images.

4.2.3. Data Augmentation

Data augmentation methods were used to artificially enhance the variety of training samples to solve the issue of a small dataset size and enhance the generalization of the model. The ImageDataGenerator library was used to augment the data to represent natural variations that are likely to occur in endoscopic imaging. The augmentation strategies that were used included:

- Rotation within ±15°
- Horizontal and vertical flipping.
- Zooming in the range of 0.8 to 1.2
- Luminance range in the range of 20%.

The effective training dataset size was altered through these transformations, which allowed the deep learning model to learn the features that are invariant to different imaging conditions. The method minimizes overfitting and enhances strength without the need to collect more data.

4.3. Deep Feature Extraction using Res-Net50 and ViT Model

In order to find deep and meaningful features in gastrointestinal images, a hybrid method was followed using the ResNet-50 and Vision Transformer (ViT), as these two methods are complementary to each other. One of the most popular CNN architectures that is widely popular and can be considered as the one with powerful feature extracting properties is ResNet-50, which is considered to be highly applicable to images with complex texture and pattern, including medical images. Its residual skip connection implementation makes it able to be trained to very deep architectures with no problem of hitting the vanishing gradient problem, and learn more local features such as edges, textures, and lesion geometry. The localized features play a central role in the identification of small-scale visual indicators that are indicative of disease severity in UC.

In Contrast, ViT utilizes the self-attention model to capture global contextual relationships in the whole image. In comparison to CNN, which looks at local neighbourhoods, ViT splits the image into patches and captures long-range dependency and spatial relationships among the patches. This global recognition compensates for the local feature extraction of CNN by allowing the model to have a global context understanding of the image, which is useful for complicated medical diagnosis tasks.

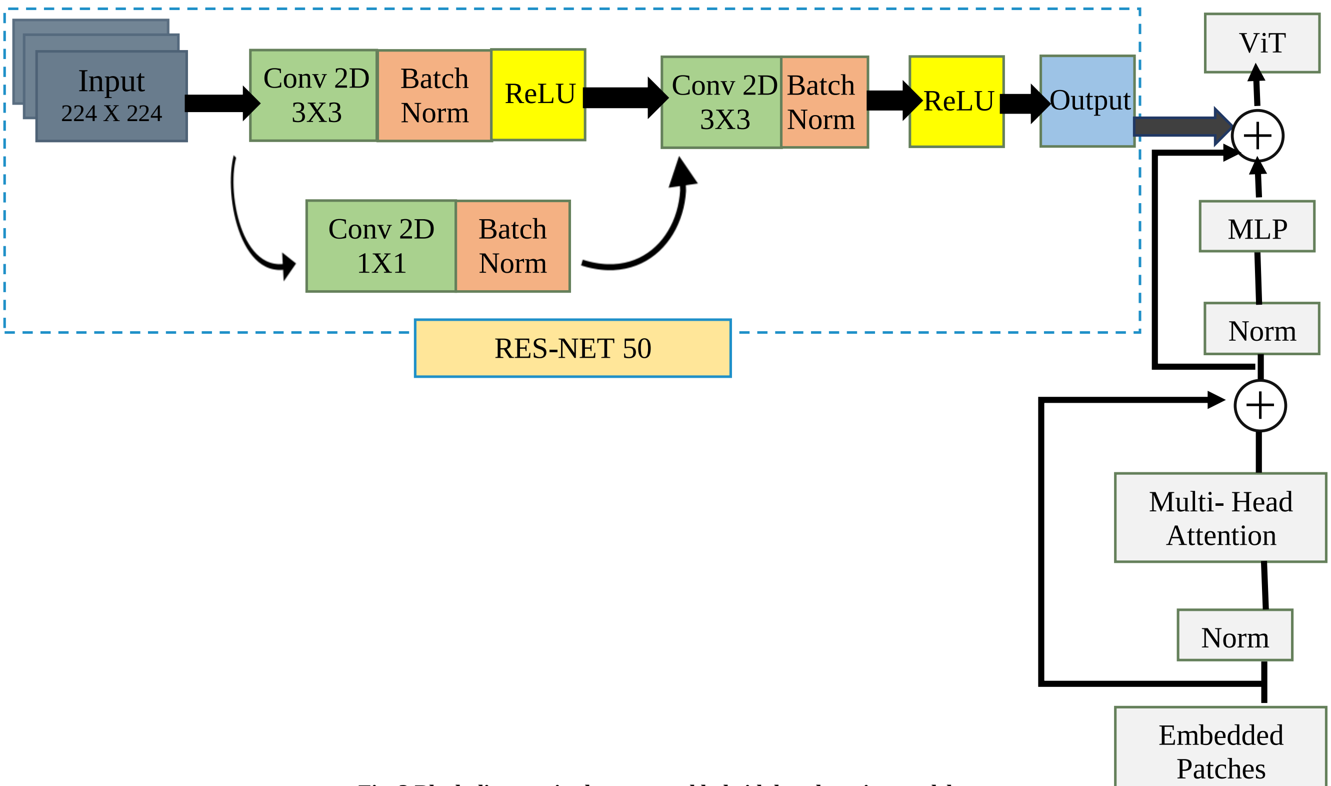

Fig. 3 Block diagram in the proposed hybrid deep learning model

By fusing ResNet-50's local texture awareness and ViT's global context perception, our model is able to robustly capture the complex visual features necessary for effective UC detection.

4.3.1. ResNet-50 Architecture

ResNet-50 is a residual network with 50 layers that adds residual skip connections to help improve the vanishing gradient problem. The skip connections enable gradients to pass directly through identity mappings, so it's possible and stable to train very deep networks.

The Feature Extraction Process is the final convolutional block that captures feature maps rich in semantics, which contain high-level semantic information. The GAP layer is used to transform these feature maps into a fixed-size feature vector as in Equation (1) [25].

$$y = F(x, \{W_i\}) + x \quad (1)$$

Here, x is the input to a residual block, F represents the residual mapping learned by stacked convolutional layers with weights W_i, and y is the output.

4.3.2. Feature Extraction Process

The final convolution block captures feature maps rich in semantics, which contain high-level semantic information. The GAP layer is used to transform these feature maps into a fixed-size feature vector.

GAP spatially averages each feature map, capturing the spatial existence of features, thus suppressing overfitting and model parameters. This vector detects local information, such as areas of inflammation, texture deformities, and lesion squares that are important in the classification of UC.

4.3.3. Output Representation

It is a vector that collimates contextual information in all patches across the globe to encode general image semantics, including the spatial location of lesions and morphology patterns, to detect UC. The architecture is provided in Figure 3.

4.4. Handcrafted Feature Extraction

Handcrafted texture descriptors were incorporated into the model to add feature diversity and rich low-level visual cues. The conventional descriptors are particularly beneficial in medical imaging, as local texture differences (i.e., surface abnormalities, inflammatory patterns, and granularity) offer important diagnostic hints. In this way, handcrafted descriptors and deep features are combined to construct a more generalizable and stronger model for ulcerative colitis classification. Two well-documented texture analysis methods were utilized: GLCM and LBP [26].

4.4.1. Gray-Level Co-occurrence Matrix (GLCM)

GLCM is a statistical approach that measures the frequency with which pairs of pixel intensities occur in a given spatial relationship in an image. This approach is very effective in measuring image texture as it contains second-order statistics.

4.4.2. *Local Binary Pattern (LBP)*

LBP is a simple but highly effective approach to describing local texture patterns. It systemically compares each pixel with its neighbours to make a binary number. It shows texture features (Contrast, Correlation, Homogeneity, Energy, ASM) for Normal and Mild UC classes. The radial chart demonstrates how the shape's texture features vary throughout the disease. These LBP measures are highly sensitive to, and depict the changes in micro-patterns (i.e., spots, edges, and flats), which are essential to detect early-stage inflammation and/or structural inconsistencies [27].

4.5. Feature Fusion and Dimensionality Reduction

Features extracted from ResNet-50, Vision Transformer, and handcrafted methods are concatenated into a single vector to leverage complementary information as in Equation (2) [28].

$$\{ Feature\ ) = \{PCA\}\backslash\{Feature - Concat\} \tag{2}$$

Where $Feature - Concat$: The concatenated feature vector, formed by combining multiple feature sets.

PCA reduces the dimensionality of the concatenated feature vector while preserving 95% of the variance. This lowers computational complexity and removes redundancy, enabling efficient downstream processing.

4.6. Classification and Optimization

Following the extraction of high-dimensional features and their dimensional reduction via PCA, an FCNN is implemented for classification. An output softmax layer to produce class probabilities in multi-class classification, as in Equation (3) [29].

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{3}$$

Here, y_i is the ground truth (1 for the correct class, 0 otherwise), and $\hat{y}_i$ is the predicted probability. Minimizing this loss improves model accuracy in multi-class classification tasks.

The Adam optimizer is employed to update network weights while training. Adam allows for effective, stable, and quick convergence even when sparse gradients or noisy data are used, so it is perfectly suitable for medical image analysis. It starts with pre-processing, which involves resizing images to 224×224×3, normalization, and data augmentation (rotation, flip, zoom, change in brightness) for model stability. Deep features are extracted with ResNet-50 to extract localized texture information and Vision Transformer (ViT) to extract global contextual relations.

Additionally, manually built features such as GLCM and LBP are also obtained to achieve an improved low-level texture coding. The features are amalgamated and reduced through the Principal Component Analysis (PCA) to preserve 95 percent of the variance, with the rest of the redundancy. In FCNN, the use of ReLU activation and softmax output is carried out by classification. The combined process offers valid and robust UC detection.

5. Result and Analysis

The target hybrid architecture, ResNet-50 + Vision Transformer (ViT), was developed and implemented using the Python platform with the help of the TensorFlow and PyTorch libraries when performing deep learning operations. The convolutional strength of ResNet-50 and the global contextual perception of ViT could be successfully used in the model to classify the severity of Ulcerative Colitis (UC). The proposed model was trained on the Dataset and was determined to have a classification accuracy of 97.5% and a weighted F1-score of 0.981, outperforming the conventional classifiers and isolated deep models. Such results suggest the strength and validity of the model and its clinical use in assisting the diagnosis of UC by examining colonoscopy images.

5.1. Experimental Setup

Table 2 shows the experiment design to classify the ulcerative colitis severity of the WCE Curated Colon Disease Dataset. The model was trained and executed with the help of a high-performance computing system with an NVIDIA RTX 3090 graphics card, an Intel i9 CPU, and 64 GB of RAM, which allowed the execution process to be performed in a limited time frame.

The model is written in TensorFlow 2.9 and uses Keras APIs to create a deep model and Scikit-learn to assess the results. The Adam optimizer was chosen with a learning rate of 0.0001 to ensure effective learning of the model. The experiment was performed with 50 epochs, which was enough to train the model, yet not to overfit. To use Categorical Cross-Entropy (CCE) as the loss function, the multi-class classification of the stages of UC severity was applied.

Table 2. Experimental setup

Parameter	Value
Dataset Name	WCE Curated Colon Disease Dataset
Hardware Components	NVIDIA RTX 3090 GPU, Intel i9 CPU, 64GB RAM
Optimizer	Adam Optimizer
Number of Epochs	50
Learning Rate	0.0001
Framework Used	Python (TensorFlow 2.9, Keras, Scikit-learn)
Loss Function	Categorical Cross-Entropy

5.2. Image Augmentation Sample Images

Different data augmentation methodologies, as in Figure 4, were applied in an effort to improve model generalization and mitigate the problem of class imbalance within the WCE Curated Colon Disease Dataset. A multi-pronged data augmentation approach was used throughout the data augmentation, using the following methods: rotation (±15°), horizontal and vertical flipping, zooming (up to 10%), adjustment of brightness or Contrast, and random cropping.

Each augmentation allows for variation randomly in the endoscope images in the real world, and maintains robustness of the model regardless of angle, scale, and lighting. Data augmentation is beneficial for classes where representation is sufficiently low, especially for classes like Severe UC.

Data augmentation provided advantages of diversity in the data, applied in a dynamic fashion while training the model on batches of data, producing new images of augmented data in distinct epochs of training, and keeping the size of the data set the same while accessing a variety of images, and helped to improve the overall classification of all categories of severity.

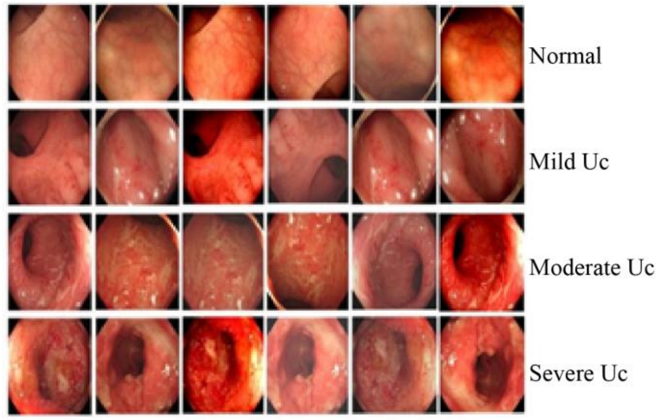

Fig. 4 Images augmented from the dataset

5.3. Analysis of Handcrafted Features (GLCM & LBP)

Figure 5 is a bar chart showing the comparison of GLCM contrast and LBP energy values between four UC classes: Normal, Mild UC, Moderate UC, and Severe UC. The chart indicates that the values of the features significantly increase with the severity of the disease, indicating the presence of more textural aberrations and surface disturbance in the inflamed regions. GLCM contrast shows the difference in intensity between neighbouring pixels, and LBP energy shows local uniformity of texture.

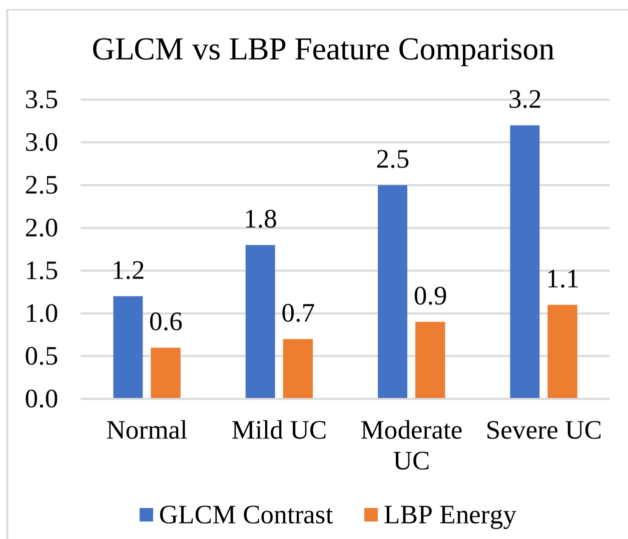

Fig. 5 GLCM and LBP features across uc classes

The radar chart in Figure 6 is a multivariate comparison of the two classes of texture-Normal and Mild UC in terms of Contrast (0.87), Correlation (0.91), Homogeneity (0.42), Energy (0.12), and ASM.

The axes represent the different texture descriptors of GLCM. The two classes have differently identifiable shapes, with which inflammation alters the spatial pixel relationships. Normal images have more Homogeneity and ASM, which are signs of smoother tissue textures.

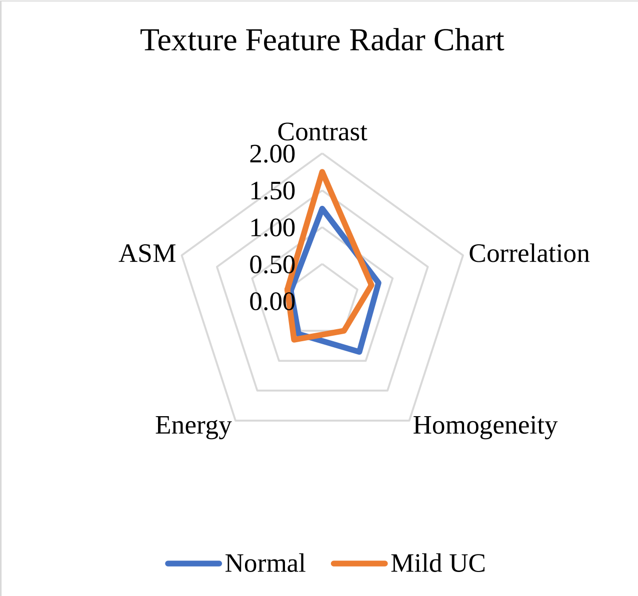

Fig. 6 Radar chart of texture features for representative classes

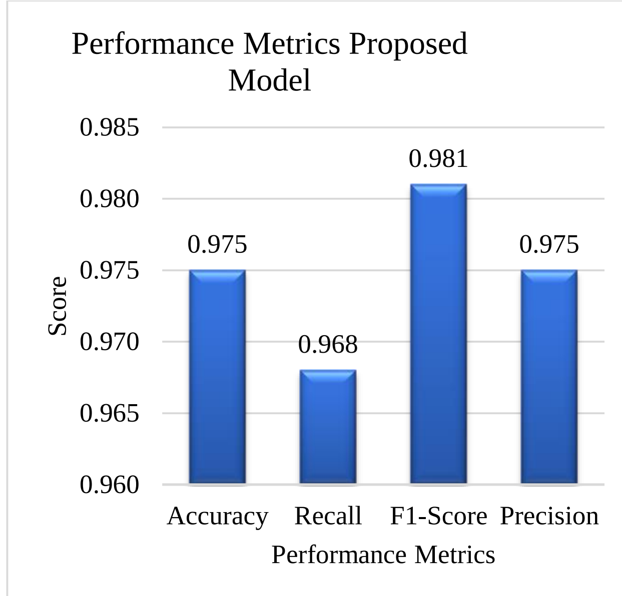

Fig. 7 Performance metrics

Mild UC has greater Contrast and less energy, which shows early mucosal irregularities. This radar visualization successfully extracts subtle variation between image texture profiles and represents an intuitive methodology with which to compare multidimensional features in medical imaging, facilitating diagnostic model interpretability.

5.4. Class-Wise Performance Metrics

Table 3 shows the sample texture features extracted, and Figure 7 provides a report of the detailed classification of the performance of the colon disease classification model on four classes, which are: Normal, Mild UC, Moderate UC, and Severe UC.

Table 3. Sample texture features extracted

Image ID	img001
Contrast	0.87
Homogeneity	0.42
Energy	0.12
Correlation	0.91
LBP Histogram Entropy	0.88

The hybrid ConvNeXt-Swin Transformer with a feature fusion model designed by hand provides good and well-established classification results in all the evaluation measures. The total accuracy of 97.5 means that the model recognizes most cases of severity of Ulcerative Colitis (UC), and this depicts the model as being reliable in multi-class prediction.

The model has a recall of 96.8%, indicating that it has been effective in predicting true positives, that is, the majority of cases that are diseased are predicted with minimal false negatives, as shown in Figure 7.

The accuracy of the prediction of the positive cases at 97.5% indicates that the positive cases predicted are very accurate, with a minimal number of false alarms. It is also important to note that the model attains an F1-score of 98.1 percent, which is the harmonic mean of the two: precision and recall, meaning that there is a very good balance between sensitivity and specificity.

The large F1-score indicates the consistency of performance of the model in all the categories of UC severity. In general, these indicators prove the strength, the ability to generalize, and clinical usefulness of the given framework for the proper ulcerative colitis severity classification, as shown in Table 3.

Table 4. Classification report

Metric	**Score (%)**
Accuracy	97.5%
Recall	96.8%
F1-Score	98.1%
Precision	97.5%

The heatmap of the CM in Figure 8 displays the classification of four types of colon conditions: Normal, Mild UC, Moderate UC, and Severe UC. The diagonal cells on the heatmap denote correct predictions, whereas off-diagonal cells denote misclassifications.

As can be seen, large values on the diagonal (e.g., 93 for Normal and Severe UC) indicate the strong capability of the model to correctly classify the majority of samples. severities. The distinct division between categories attests to the strength of training. By presenting the confusion matrix, one can understand some of the performance issues at the class level that may not be apparent in the overall accuracy.

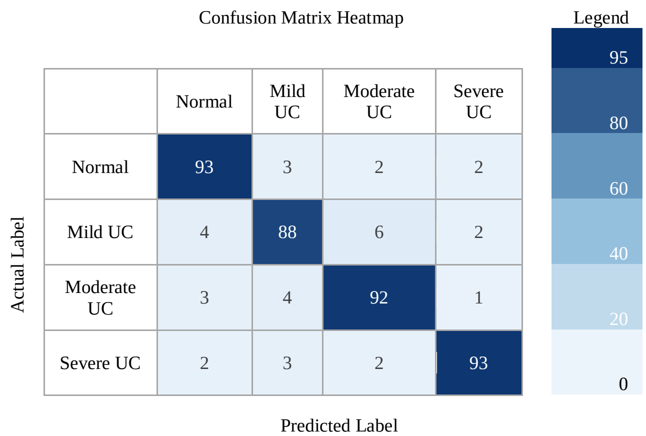

Fig. 8 Confusion matrix

All four types of colon diseases have an ROC provided in Figure 9. The graphs are the curves illustrating the TPR and the FPR with different classification thresholds. The AUC numbers are staggering: Normal (0.97), mild UC (0.94), moderate UC (0.95), and severe UC (0.96), which is evidence that the model is so good at discrimination. Large AUCs indicate that the classifier is not weak in any of the classes, and the discrimination of the Normal class is close to perfection.

This capability to identify each of the types, especially in an issue of multiple classes of health, is pivotal towards enabling real-time clinical diagnosis. ROC curves boost the predictability and reliability of the model predictions and are especially predictive and reliable in cases where the disease progression stages are context-dependent.

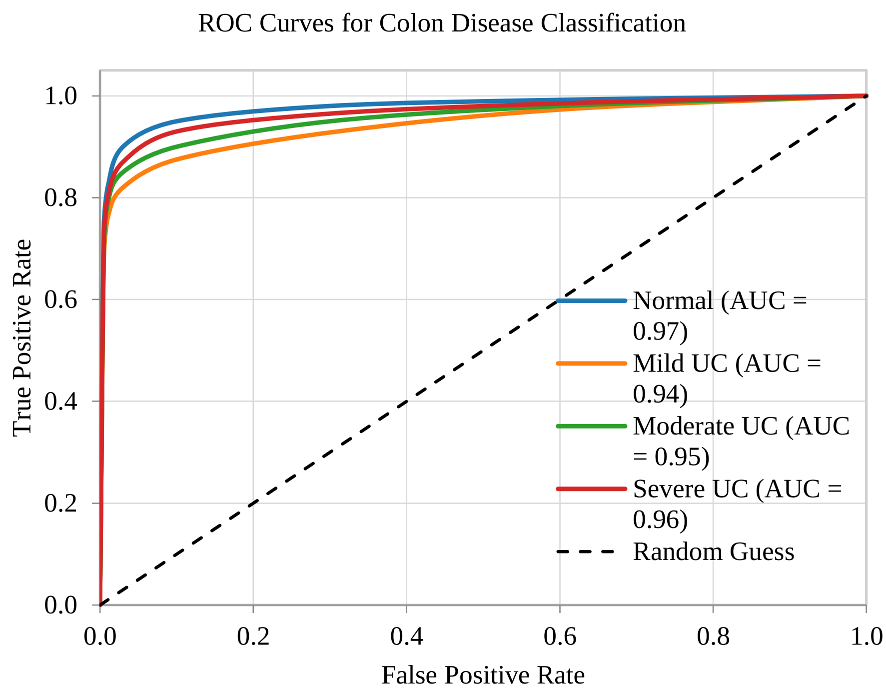

Fig. 9 ROC curve

5.5. Accuracy and Loss over Epochs

Figure 10 depicts curves that rise smoothly, which suggests that the model learns continuously as it trains. Beginning at 70%, accuracy climbs to more than 96% on sets. The increase, with a negligible gap between the two lines, demonstrates great model generalization without overfitting. Optimization with the AO, which has adaptive learning rate capability, was used, with categorical cross-entropy loss, which is best suited for MCC scenarios. The robust convergence behaviour confirms that the architecture as well as the training approach are well-matched. Monitoring accuracy enables the evaluation of the extent to which the model separates Normal, Mild UC, Moderate UC, and Severe UC. Under medical imaging scenarios, stable improvement in validation accuracy is critical to guarantee performance on unseen clinical data. This plot confirms the selected training configuration to favourably support disease stage identification. The reduction in loss in 50+ epochs is shown in Figure 11. The two loss curves will exhibit a consistent negative slope, starting with a high curve (~1.2), and will tend to reach 0.1. This gradual decrease is indicative of this model cutting errors in a manner that the training procedure optimizes Adam, which is best applied to sparse gradients and noisy data. CCE is the loss function used to assess the fit of predicted probabilities to true class labels.

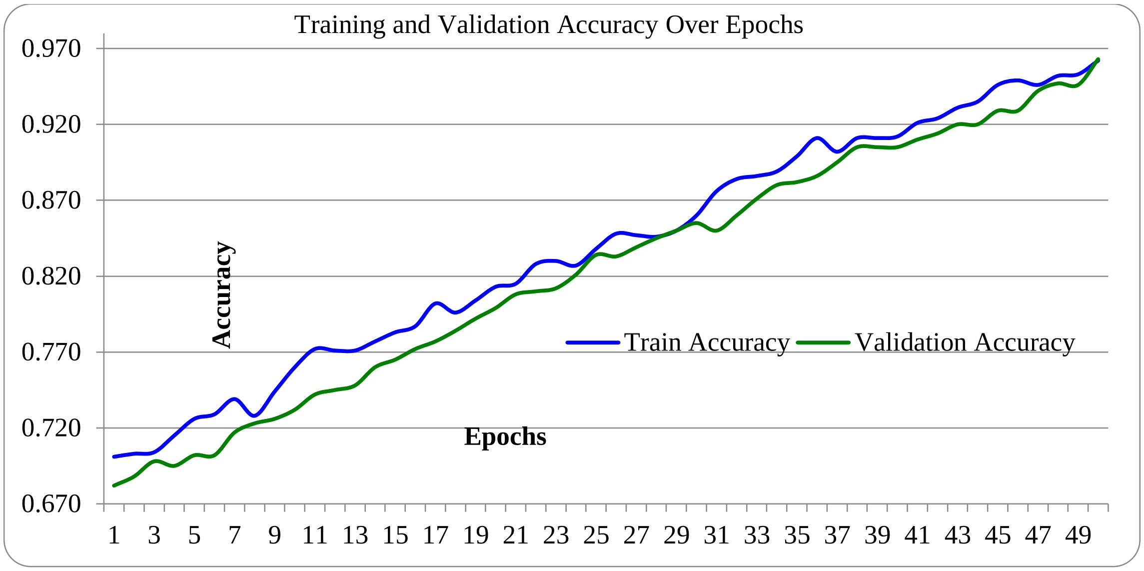

Fig. 10 Accuracy over epochs

As the training continues, the small difference between training and validation loss ensures that the model is not overfitting but picking up on generalized patterns. In deep learning, the convergence pattern indicates good feature learning, particularly for fine-grained medical image classification tasks. The loss plot supports the accuracy plot by ensuring the robustness of the learning process. It guarantees researchers that the model has been well-trained to facilitate correct and stable predictions in all colon disease severities.

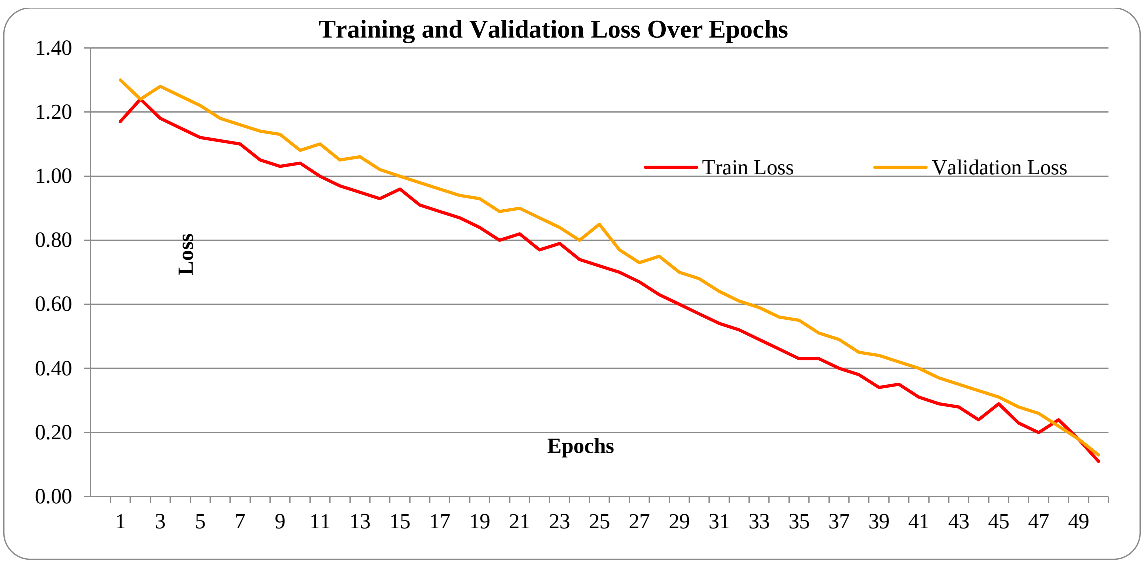

Fig. 11 Loss over epochs

5.6. Comparison with Various Models

The given comparative analysis proves that the proposed hybrid model can be successfully used to significantly improve the performance of conventional machine learning and standalone deep learning methods in the field of ulcerative colitis classification. Conventional models like Decision Tree were only able to get an accuracy of 83.6, which does not imply a high level of ability to capture the complex texture and structural variations that are found in medical pictures, as shown in Figure 12.

Table 5. Model comparison

Models	Accuracy
Decision Tree [36]	83.6
Xception [14]	92.4
DNN [23]	93.75
ResNet-50 (Baseline CNN) [37]	95
KNN [36]	93.33
Proposed model	97.5

On the same note, KNN achieved 93.33, which is better but does not present strong feature abstraction. Xception and DNN are the highest accuracy deep learning models with 92.4 and 93.75, respectively, as the results indicate a higher representation learning than classical models. ResNet-50 baseline CNN is even more effective because it extracts hierarchical convolutional features, but it is a single-architecture model and therefore might not resolve all the contextual dependencies.

The proposed hybrid framework, on the other hand, has an accuracy of 97.5% due to the combination of complementary local, global, and handcrafted features.

The high enhancement of this aspect shows the efficiency of multi-level feature fusion and dimensionality optimization, which proves that the suggested model is better than traditional and standalone deep learning models, as shown in Table 4.

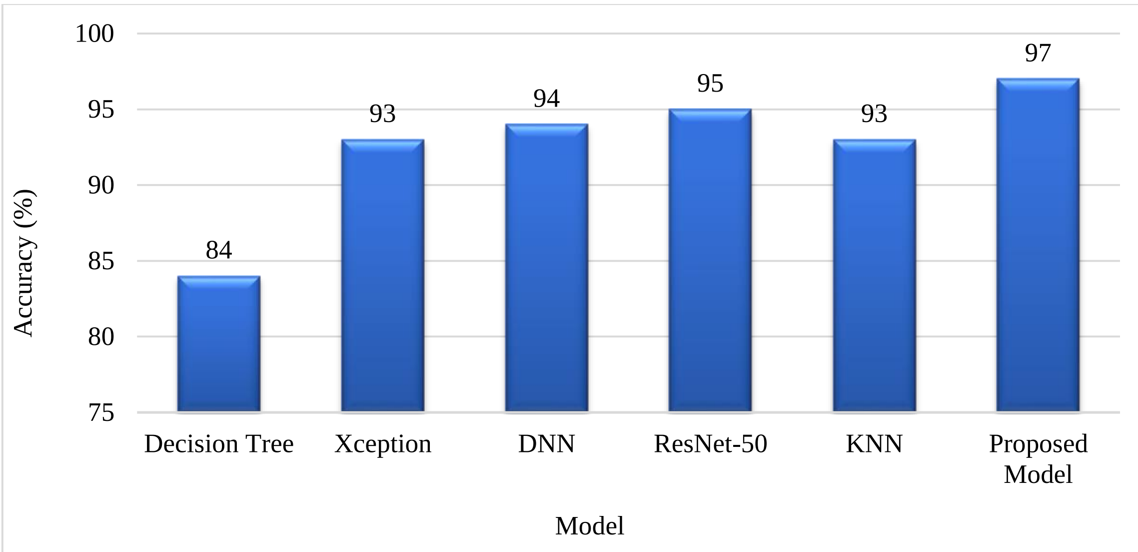

Fig. 12 Accuracy comparison of classification models

6. Discussion

The experimental findings prove that the designed hybrid ConvNeXtSwin Transformer model is more effective than traditional machine learning and single deep learning systems. Combining the deep local features, global contextual modeling, and handcrafted texture features can substantially increase the discriminative ability on severity classification of ulcerative colitis. PCA and feature fusion approach decrease redundancy and maintain critical diagnostic data, which results in better generalization abilities. Although the dataset size (800 images) is quite small, stratified sampling, data augmentation techniques, and regularization techniques contributed to the reduction of overfitting and enhanced robustness.

Table 6. Performance metrics

Study	Accuracy	Precision	Recall	F1-Score
Sutton et al. 2022. [30]	87.50%	N/R	N/R	N/R
Shah et al. 2024. [31]	90.00%	N/R	N/R	N/R
Jahagirdar et al. 2023. [32]	91.50%	N/R	87%	N/R
Ahmed et al. 2025.[33]	93.00%	N/R	N/R	N/R
Nie& Zhang 2025[34]	80.56%	N/R	N/R	N/R
Margapuri2025. [35]	~70%	N/R	N/R	73.10%
Proposed	**97.50%**	**97.50%**	**96.80%**	**98.10%**

** N/R = Not Reported in the original paper.*

Comparison of performance parameters of the proposed work with previous studies is shown in Table 6. The fact that the proposed framework has high accuracy (97.5%), recall (96.8%), and F1-score (98.1) is a sign that the proposed framework is effective in capturing micro-level abnormalities of the mucosal area and macro-level structural patterns. These results indicate that hybrid architectures are capable of producing a more reliable clinical decision support than single backbone networks. Despite the great performance of the proposed model, it has a number of limitations. No external validation was conducted, and the level of computational complexity of the hybrid model may act as an obstacle to implementing it in a resource-constrained clinical setting.

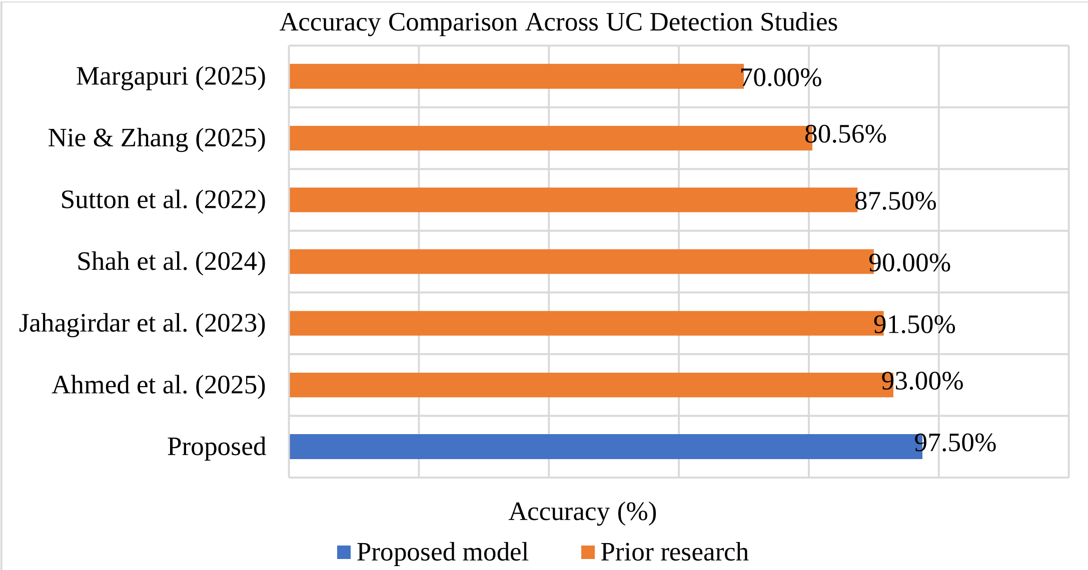

Fig. 13 Comparison of accuracy between UC detection research

Accuracy Figure 13 displays the comparison with past UC Detection Studies.

The study employed a small number of the WCE pictures (800), which may pose a constraint to generalization. The filtered collection of data may not be a true reflection of clinical variability in the actual contexts.

7. Conclusion

A hybrid ConvNeXt-Swin Transformer model to detect ulcerative colitis and predict the severity will be presented in this work. To obtain high classification with the proposed method, the use of local convolutional representations, global contextual modeling transformer, and handcrafted representations of texture is employed with 97.5%. The fusion strategy and dimensionality reduction assist in promoting additional efficiency and robustness. The findings suggest that hybrid deep learning architectures may be used to create potentially effective decision-support systems to be used in gastrointestinal tract diagnostics.

With additional validation and optimization, the presented system will be a significant contribution to automated clinical screening and monitoring of diseases. The future research will involve the validation of the model on bigger multi-center data sets to increase the generalizability of the study. Distribution shift resistance can be enhanced by means of integration of cross-dataset evaluation and domain adaptation methods. Pruning and quantization are lightweight model optimization methods that will be examined to allow clinical applications in real-time. Furthermore, explainable AI (XAI) will be included (Grad-CAM) to achieve greater interpretability and clinical trust. The extension to multi-disease gastrointestinal classification and severity grading is also a fruitful way.

References

[1] Marek Vebr et al., "A Narrative Review of Cytokine Networks: Pathophysiological and Therapeutic Implications for Inflammatory Bowel Disease Pathogenesis," *Biomedicines*, vol. 11, no. 12, pp. 1-52, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Chan Hyung Lee et al., "Animal Models of Inflammatory Bowel Disease: Novel Experiments for Revealing Pathogenesis of Colitis, Fibrosis, and Colitis-Associated Colon Cancer," *Intestinal Research*, vol. 21, no. 3, pp. 295-305, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Christian Philipp Selinger, Konstantina Rosiou, and Marco V. Lenti, "Biological Therapy for Inflammatory Bowel Disease: Cyclical Rather than Lifelong Treatment?," *BMJ Open Gastroenterology*, vol. 11, no. 1, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Catherine Le Berre, Silvio Danese, and Laurent Peyrin-Biroulet, "Can We Change the Natural Course of Inflammatory Bowel Disease," *Therapeutic Advances in Gastroenterology*, vol. 16, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Samina Khan et al., "Factors Influencing the Quality of Life in Inflammatory Bowel Disease: A Comprehensive Review," *Disease-a-Month*, vol. 70, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Rasmus Rankala et al., "Costs of Medications Used to Treat Inflammatory Bowel Disease," *Scandinavian Journal of Gastroenterology*, vol. 59, no. 1, pp. 34-38, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ryan Ungaro et al., "A Treat-to-Target Update in Ulcerative Colitis: A Systematic Review," *The American Journal of Gastroenterology*, vol. 114, no. 6, pp. 874-883, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Asif Hassan Syed et al., "Advances in Inflammatory Bowel Disease Diagnostics: Machine Learning and Genomic Profiling Reveal Key Biomarkers for Early Detection," *Diagnostics*, vol. 14, no. 11, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Iolanda Valentina 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, pp. 1-9, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Claudia Diaconu et al., "The Role of Artificial Intelligence in Monitoring Inflammatory Bowel Disease-The Future Is Now," *Diagnostics*, vol. 13, no. 4, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Hassaan Malik et al., "Multi-Classification Deep Learning Models for Detection of Ulcerative Colitis, Polyps, and Dyed-Lifted Polyps Using Wireless Capsule Endoscopy Images," *Complex & Intelligent Systems*, vol. 10, pp. 2477-2497, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Amit Das et al., "Deep Learning for Classification of Inflammatory Bowel Disease Activity in Whole Slide Images of Colonic Histopathology," *The American Journal of Pathology*, vol. 195, no. 4, pp. 680-689, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Venkat Margapuri, "Diagnosis and Severity Assessment of Ulcerative Colitis Using Self Supervised Learning," *2025 IEEE Symposium on Computational Intelligence in Health and Medicine Companion (CIHM Companion)*, Trondheim, Norway, pp. 1-5, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Sabrina Gloria Giulia Testoni et al., "Artificial Intelligence in Inflammatory Bowel Disease Endoscopy," *Diagnostics*, vol. 15, no. 7, pp. 1-44, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jaehyuk Lee, and Eunchan Kim, "Multi-Task Deep Learning Framework for Enhancing Mayo Endoscopic Score Classification in Ulcerative Colitis," *Digital Health*, vol. 11, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[16] Zeynep Özdemir, Hacer Yalim Keles, and Ömer Özgür Tanriöver, "CLoE: Curriculum Learning on Endoscopic Images for Robust MES Classification," *IEEE Access*, vol. 14, pp. 30441-30454, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[17] Mehwish Ahmed, Molly L. Stone, and Ryan W. Stidham, "Artificial Intelligence and IBD: Where are We Now and Where Will We Be in the Future?," *Current Gastroenterology Reports*, vol. 26, pp. 137-144, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Gian Eugenio Tontini et al., "Artificial Intelligence in Gastrointestinal Endoscopy for Inflammatory Bowel Disease: A Systematic Review and New Horizons," *Therapeutic Advances in Gastroenterology*, vol. 14, pp. 1-16, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Ryan W. Stidham et al., "Artificial Intelligence-Enabled Clinical Trials in Inflammatory Bowel Disease: Automating and Enhancing Disease Assessment and Study Management," *Gastroenterology*, vol. 169, no. 3, pp. 432-443, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Pieter Sinonquel et al., "Artificial Intelligence and its Impact on Quality Improvement in Upper and Lower Gastrointestinal Endoscopy," *Digestive Endoscopy*, vol. 33, no. 2, pp. 242-253, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jacob Broder Brodersen et al., "Artificial Intelligence-assisted Analysis of Pan-enteric Capsule Endoscopy in Patients with Suspected Crohn's Disease: A Study on Diagnostic Performance," *Journal of Crohn's and Colitis*, vol. 18, no. 1, pp. 75-81, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Gary R. Lichtenstein et al., "Increased Lifetime Risk of Intestinal Complications and Extraintestinal Manifestations in Crohn's Disease and Ulcerative Colitis," *Gastroenterology & Hepatology*, vol. 18, no. 1, pp. 32-43, 2022. [Google Scholar] [Publisher Link]

[23] Vipul Jairath et al., "Novel Outcomes in Inflammatory Bowel Disease," *Journal of Crohn's and Colitis*, vol. 19, no. 4, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] WCE Curated Colon Disease Dataset Deep Learning, Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning

[25] Kaiming He et al., "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[26] Haijig Sun et al., "An Improved Medical Image Classification Algorithm Based on Adam Optimizer," *Mathematics*, vol. 12, no. 16, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Stefania Barburiceanu, Romulus Terebes, and Serban Meza, "3D Texture Feature Extraction and Classification using GLCM and LBP-based Descriptors," *Applied Sciences*, vol. 11, no. 5, pp. 1-25, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[28] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[29] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016. [Google Scholar] [Publisher Link]

[30] Reed T. Sutton et al., "Artificial Intelligence Enabled Automated Diagnosis and Grading of Ulcerative Colitis Endoscopy Images," *Scientific Reports*, vol. 12, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Syed Abdullah Shah et al., "A Hybrid Approach of Vision Transformers and CNNs for Detection of Ulcerative Colitis," *Scientific Reports*, vol. 14, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[32] Vinay Jahagirdar et al., “Diagnostic Accuracy of Convolutional Neural Network Based Machine Learning Algorithms in Endoscopic Severity Prediction of Ulcerative Colitis: A Systematic Review & Meta-Analysis,” *Gastrointestinal Endoscopy*, vol. 98, no. 2, pp. 145-154, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[33] Md. Nasim Ahmed et al., “A Triple Pronged Approach for Ulcerative Colitis Severity Classification using Multimodal, Meta, and Transformer Based Learning,” *Scientific Reports*, vol. 15, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[34] Boying Nie, and Gaofeng Zhang, “Ulcerative Severity Estimation Based on Advanced CNN-Transformer Hybrid Models,” *Applied Sciences*, vol. 15, no. 13, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[35] Venkat Margapuri, “Diagnosis and Severity Assessment of Ulcerative Colitis Using Self Supervised Learning,” *2025 IEEE Symposium on Computational Intelligence in Health and Medicine Companion (CIHM Companion)*, Trondheim, Norway, pp. 1-5, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[36] Yuanming Yang et al., “Biomarkers Prediction and Immune Landscape in Ulcerative Colitis: Findings based on Bioinformatics and Machine Learning,” *Computers in Biology and Medicine*, vol. 168, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[37] Zhengxuan Qiu et al., “VIGIL: Vision-Language Guided Multiple Instance Learning Framework for Ulcerative Colitis Histological Healing Prediction,” *arXiv preprint*, pp. 1-12, 2025. [CrossRef] [Google Scholar] [Publisher Link]