

Bleeding Classification and Segmentation in Capsule Endoscopy Images using Deep Learning

Gatreddi Srivardhan, Doni Kavya, Busayavalasa Venkata Sai Siri Chandana,
Erigala Pranav Deep, Yalagada Jaya Sri, Baisakh
GMR Institute of Technology, Rajam, Vizianagaram, India

Abstract: *Gastrointestinal (GI) bleeding is a common and potentially serious condition that requires timely and accurate diagnosis. Wireless Capsule Endoscopy (WCE) has emerged as a valuable tool for capturing images of the digestive tract without the need for invasive procedures. However, the manual review of these images is time-consuming and depends on the expertise of medical professionals. To improve efficiency and accuracy, we propose an automated deep learning-based approach for detecting and segmenting bleeding in WCE images. Our method leverages Swin Transformer for classifying bleeding and non-bleeding cases and employs UNet++ for precise segmentation of bleeding areas. We evaluate our approach using the WCEBleedGen dataset, achieving high accuracy in classification and precise segmentation in segmentation tasks. The integration of deep learning significantly boosts accuracy and efficiency, revolutionizing medical image analysis. Providing clear visualizations of bleeding regions, this method minimizes human error and streamlines the diagnostic process. The results demonstrate the potential of artificial intelligence in healthcare, making medical diagnostics more reliable and effective. Future research should focus on expanding the dataset diversity and integrating real-time processing capabilities to further enhance the system's reliability and speed. Continuous improvements in AI-driven methodologies will contribute to the evolution of automated GI diagnostics, offering even greater support to healthcare professionals in disease detection and treatment planning..*

Keywords: Wireless Capsule Endoscopy (WCE), Bleeding Detection, Deep Learning, Swin Transformer, Swin-Unet, UNet++, Medical Image Segmentation, Vision Transformers, AI in Healthcare

I. INTRODUCTION

Gastrointestinal (GI) bleeding is a serious medical condition that requires early detection and intervention to prevent complications. Traditional endoscopic techniques, such as colonoscopy and esophagogastroduodenoscopy (EGD), are effective in detecting bleeding but are invasive, costly, and require sedation. Moreover, these methods do not provide continuous monitoring, which limits their ability to detect intermittent bleeding. To address these limitations, Wireless Capsule Endoscopy (WCE) has emerged as a non-invasive and patient-friendly alternative, allowing continuous visualization of the GI tract. In WCE, the patient swallows a small capsule equipped with a camera that captures thousands of images as it travels through the digestive system. While this approach provides comprehensive imaging, the manual analysis of WCE images is highly labor-intensive and subjective, making automated bleeding detection a critical area of research. Deep learning offers a promising solution by enabling efficient and highly accurate classification and segmentation of medical images, reducing the diagnostic workload and improving early detection. Deep learning-based models, particularly Convolutional Neural Networks (CNNs), have been widely used for medical image analysis, including WCE images. However, CNNs often struggle with capturing long-range dependencies and contextual relationships, limiting their ability to accurately detect subtle bleeding patterns in endoscopic images. To overcome these challenges, recent advancements in Vision Transformers (ViTs) have been explored, offering superior performance in medical imaging tasks by utilizing self-attention mechanisms to process images more effectively. In this study, we propose a deep learning-based framework for GI bleeding detection and segmentation using Swin Transformer for classification and UNet++ for segmentation. Our approach leverages Swin Transformer's hierarchical feature extraction to improve classification accuracy and UNet++'s nested skip connections to enhance segmentation



precision. By integrating these two advanced models, we aim to provide an efficient, automated solution for GI bleeding detection in WCE images. The first step involves classifying WCE images into bleeding and non-bleeding categories. Swin Transformer, a hierarchical Vision Transformer, is chosen due to its ability to capture long-range dependencies and contextual features more effectively than CNNs. It processes WCE images in shifted windows, enabling multi-scale feature learning while maintaining computational efficiency. This improves the detection of bleeding regions, even in complex or low-contrast images. Once the bleeding images are identified, it is essential to precisely localize the bleeding regions for clinical decision-making. We employ UNet++, an advanced version of the UNet model, designed to improve segmentation accuracy through nested and dense skip connections. These connections enhance feature propagation and refine segmentation boundaries, ensuring better localization of bleeding areas. This is particularly useful for small and subtle bleeding regions, which might be missed by traditional segmentation models. We train and evaluate our model using the WCEBleedGen dataset, a curated dataset for GI bleeding detection. It contains bleeding images with corresponding bounding boxes and segmentation masks, along with non-bleeding images serving as negative samples for classification. To improve model generalization, data augmentation, normalization, and image resizing are applied during preprocessing. To assess the performance of our framework, we use standard evaluation metrics. For classification, we measure Accuracy, Precision, Recall, and F1-score to evaluate Swin Transformer's ability to distinguish between bleeding and non-bleeding images. For segmentation, we use Dice Coefficient and Intersection-over-Union (IoU) to determine how accurately UNet++ localizes bleeding regions. Experimental results show that our model outperforms traditional CNN-based methods, achieving higher classification accuracy and better segmentation precision. The use of Swin Transformer improves the recognition of bleeding patterns, while UNet++ ensures more precise and reliable segmentation of bleeding areas. Our proposed framework significantly enhances automated bleeding detection and segmentation in WCE images, addressing the key challenges of manual WCE analysis. By integrating Swin Transformer and UNet++, our model reduces the diagnostic workload, helping gastroenterologists analyze WCE images more efficiently. It also improves detection accuracy, leading to early intervention and better patient outcomes. Furthermore, precise bleeding localization assists in treatment planning and clinical decision-making. Overall, this study demonstrates the potential of AI-driven medical imaging solutions in revolutionizing WCE analysis. By automating the detection and segmentation of bleeding regions, our approach enhances clinical efficiency, reduces human errors, and contributes to improved healthcare outcomes.

1.1 Problem Statement:

Gastrointestinal (GI) bleeding is a serious medical condition that, if left undetected, can lead to severe complications, including anemia, shock, or even death. Wireless Capsule Endoscopy (WCE) has revolutionized GI diagnostics by enabling non-invasive, real-time visualization of the digestive tract. However, the manual review of WCE images is highly time-consuming, labor-intensive, and prone to human error, as it requires clinicians to examine thousands of frames per patient. Missed or delayed detection of bleeding can result in misdiagnosis and ineffective treatment.

1.2 Objectives

- Implement Swin Transformer to classify WCE images into bleeding and non-bleeding categories.
- Leverage the self-attention mechanism of Swin Transformer to improve feature representation and classification accuracy.
- Ensure the model generalizes well to different WCE datasets by implementing data augmentation and transfer learning techniques.
- Utilize Swin-Unet and UNet++ for precise segmentation of bleeding regions in WCE images.
- Optimize model parameters to reduce computational complexity and improve inference time for real-time applications.
- Develop a computer-aided diagnosis (CAD) system that assists healthcare professionals in analyzing WCE images.



II. LITERATURE SURVEY

Wireless Capsule Endoscopy (WCE) has rapidly emerged as a pivotal modality for non-invasive gastrointestinal (GI) diagnostics, particularly effective for the detection of obscure GI bleeding, which often evades conventional endoscopic techniques. The inherent challenges in WCE, including voluminous video data, heterogeneous visual patterns, and variable lighting conditions, necessitate the deployment of robust, scalable, and accurate automated bleeding detection algorithms. Brzeski et al. [1] propose an enhanced convolutional neural network (CNN) framework integrating visual feature maps with saliency attention to magnify bleeding-prone regions. Their model is architected to capture low-level color intensities and mid-level texture variations while leveraging contextual encoding layers to ensure discriminative feature learning across varying GI topographies.

In the realm of super pixel-based modeling, Liu et al. [2] devised a bleeding detection algorithm grounded in adaptive over segmentation techniques. By decomposing WCE frames into homogeneous super pixels using SLIC (Simple Linear Iterative Clustering), their method isolates compact perceptual units, within which chromatic and spatial consistency is maintained. These super pixels are then analyzed through a specialized feature extraction protocol that encodes hue variance, spatial entropy, and local contrast, effectively discriminating hemorrhagic tissues from non-pathological mucosa. The model leverages structural uniformity and boundary preservation to minimize intra-class ambiguities, outperforming baseline pixel-wise methods. Building upon region-based segmentation paradigms, Bchir et al. [3] introduced a multiple bleeding detection system that synergizes thresholding heuristics with supervised classification algorithms. Their approach employs an intelligent region-growing mechanism augmented by a probabilistic bleeding likelihood function, which iteratively classifies candidate regions based on red chromaticity features and spatial proximity metrics. The classifier is trained on a curated dataset encompassing diverse bleeding morphologies, facilitating generalization across diffuse, active, and residual bleeding types within the GI lumen.

Musha et al. [4], in a sweeping systematic review, dissect the landscape of computer-aided bleeding detection algorithms through a multifactorial lens. The review catalogs methodologies across a temporal axis, highlighting the evolutionary trajectory from rule-based classifiers to state-of-the-art transformer models. Notably, they emphasize the growing trend towards hybrid architectures that fuse CNN-based encoders with attention-guided decoders, as well as the increasing incorporation of domain adaptation techniques to mitigate dataset heterogeneity and annotation scarcity—two persistent bottlenecks in clinical AI. Rathnamala and Jenicka [5] advanced a GMM-super pixel-based bleeding detection framework that operationalizes probabilistic modeling for color distribution extraction within homogeneous regions. The pipeline is constructed upon a multi-resolution analysis scheme, enabling the system to accommodate both coarse and fine-grained bleeding manifestations. The Gaussian Mixture Model, optimized via expectation-maximization, learns bleeding-related chromatic clusters which are subsequently used to probabilistically label super pixels. This dual-layered statistical reasoning introduces robustness against inter-patient variability and image noise.

In a complementary systematic investigation, Musha et al. [6] reiterate the need for explainable AI systems in WCE diagnostics. They critique the interpretability gap in existing deep learning frameworks, particularly when deployed in high-stakes clinical decision-making scenarios. Their analysis stresses the exigency for attention visualization, confidence scoring, and radiologist-in-the-loop feedback mechanisms to foster trust and usability in real-world clinical settings. Their findings also highlight the nascent field of few-shot learning and synthetic data augmentation as potential frontiers for circumventing data scarcity issues. Ghosh and Chakareski [7] champion the application of deep transfer learning for intestinal bleeding detection, positing that domain-specific fine-tuning of large-scale pretrained models can substantially ameliorate the sample inefficiency endemic to medical imaging tasks. By adapting ImageNet-pretrained ResNet architectures to the WCE domain, they manage to retain generalized feature abstraction capabilities while incorporating domain-specific bleeding semantics via progressive fine-tuning. Their evaluation demonstrates superior classification accuracy under limited training conditions and high inter-class variance.

A landmark contribution to the domain of WCE benchmarking comes from Handa et al. [8], who introduced the WCEbleedGen dataset—a meticulously curated corpus of annotated WCE frames categorized across classification, detection, and segmentation tasks. The dataset encompasses a wide spectrum of bleeding appearances, camera angles,



and anatomical regions, rendering it an ideal testbed for algorithmic benchmarking and cross-model validation. Additionally, the inclusion of multi-label ground truth segmentation masks enables the training of both coarse-grained classifiers and fine-grained pixel-wise segmentation models, fostering advancements in supervised and weakly-supervised learning paradigms. In a seminal comparative study, Pogorelov et al. [9] conducted a methodical evaluation of color and texture-based features for bleeding detection. They underscore the limitations of relying solely on color metrics—particularly under variable lighting and compression artifacts—highlighting that bleeding may manifest across a diverse range of red tones, which can be visually confounded with non-pathological elements like food debris or bile. Their findings suggest that integrating texture descriptors such as Local Binary Patterns (LBP) and Gabor filters enhances model robustness, facilitating more consistent detection across complex GI environments.

Ghosh et al. [10] devised the CHOBS methodology, wherein statistical analysis of color histograms across block-wise image segments is employed to detect bleeding manifestations. Their approach benefits from both spatial and chromatic normalization, allowing the model to account for local color fluctuations while preserving region-specific semantics. This lightweight architecture is particularly well-suited for real-time implementation on embedded platforms used in capsule endoscopy hardware, demonstrating clinically acceptable inference latency and resource efficiency. Li and Meng [11] laid the foundational work in bleeding detection using morphological analysis and heuristic-based thresholding. Although simplistic by contemporary standards, their work is seminal in its delineation of red hue thresholds and geometric descriptors for bleeding region identification. Their handcrafted pipeline operationalized hue saturation thresholds alongside shape-based filtering to exclude false positives induced by vascular artifacts and luminal debris.

Alavala et al. [12] presented a bleeding detection architecture employing a fusion of Swin Transformer for hierarchical feature extraction and RTDETR for object localization. This dual-module framework exploits window-based attention mechanisms and dense spatio-temporal embeddings to capture both macro-level scene context and micro-level tissue anomalies. Their method achieves real-time inference capabilities while preserving high-resolution spatial accuracy, demonstrating potential for clinical deployment in autonomous diagnostic systems. Singh et al. [13] introduced ColonNet, a hybrid model architecture integrating DenseNet121 and U-Net. The DenseNet backbone captures deep hierarchical features from input WCE frames, which are then decoded via a U-Net-based segmentation head. This hybridization ensures that both global context and local detail are retained, enabling the model to delineate bleeding zones with surgical precision. The model also integrates skip connections and auxiliary loss functions to mitigate vanishing gradients and enhance convergence dynamics.

Lin et al. [14] proposed a two-stage bleeding localization approach rooted in modular decomposition of WCE imagery. Their divide-and-conquer model separates anatomical parsing from hemorrhagic inference, thereby isolating tissue-specific visual semantics prior to detection. This decomposition facilitates more accurate distinction between similar-looking pathological features and improves performance in complex anatomical landscapes marked by shadows, fluid residues, and peristaltic artifacts. Balasubramanian et al. [15] unveiled ClassifyViStA, a transformer-driven architecture incorporating segmentation-aware visual attention modules. The system employs a multi-head attention mechanism to extract context-aware embeddings while attending to bleeding-prone regions identified via an auxiliary segmentation head. This dual-channel architecture not only boosts classification precision but also facilitates model interpretability via attention-based visual saliency maps. Alawode et al.

[16] further push the boundary of transformer-based architectures by proposing a bleeding detection pipeline that integrates spatial transformers and temporal aggregation layers. Their model, capable of capturing long-range dependencies and sequential frame correlations, delivers heightened sensitivity in detecting transient and low-contrast bleeding instances that often elude conventional frame-wise classifiers.

Ghosh, Li, and Chakareski [17] engineered a semantic segmentation framework tailored for bleeding region delineation, incorporating dilated convolutions and pyramid pooling modules to capture multi-scale contextual information. Their model is optimized for edge preservation and small-object segmentation, enabling precise boundary localization of bleeding zones—an essential requirement for quantifying lesion severity and guiding therapeutic interventions. Bchir et al. [18], transitioning from handcrafted features to CNNs, developed a deep learning-based multi-bleeding detection model capable of processing complex WCE image patterns with variable bleeding intensity.



Their model incorporates dual-branch convolutional pathways—one focused on color-based features, the other on texture—to enhance robustness against color noise and anatomical confounds.

Rustam et al. [19] introduced a CNN-based classification model tailored to WCE imagery, leveraging extensive data augmentation and dropout regularization to prevent overfitting. Their pipeline emphasizes architecture simplicity while achieving state-of-the-art classification performance on publicly available WCE datasets, reinforcing the utility of lightweight models in constrained clinical settings. Rani et al. [20] presented a custom deep learning model designed for the classification of GI bleeding using optimized convolutional blocks and activation functions. Their method emphasizes class balance via weighted loss functions and employs stratified sampling to ensure representative training across minority classes. Their results indicate high recall and precision, particularly in detecting early-stage hemorrhagic lesions critical for timely clinical intervention. In synthesis, this corpus of literature encapsulates a comprehensive evolution from classical image processing techniques to cutting-edge deep learning and transformer-based frameworks. The overarching trend gravitates toward architectures that are not only performance-optimized but also interpretable, robust to real-world artifacts, and adaptable across diverse clinical scenarios. Continued research efforts in this domain are poised to deliver increasingly autonomous, accurate, and scalable bleeding detection systems that will fundamentally reshape GI diagnostics.

COMPARISION TABLE OF LITERTURE SURVEY

Table 1: comparison table

SNO	Title	Year	Merits	Demerits	Future Gaps
1	Visual Features for Improving Endoscopic Bleeding Detection Using CNNs [1]	2023	Incorporates visual saliency and deep convolutional layers, achieving high sensitivity in detecting bleeding regions with spatial attention.	Limited adaptability when tested on datasets with different lighting or anatomical diversity, suggesting low cross-domain robustness.	Requires comprehensive evaluation across multi-institutional datasets with domain generalization techniques
2	Feature Detection via Superpixel Segmentation [2]	2019	Superpixel segmentation isolates homogeneous regions, enhancing local feature representation and reducing computational load.	Performance degrades with irregular boundaries or presence of strong artifacts like bile or food residues.	Needs integration with deep neural networks for dynamic superpixel refinement and contextual interpretation.
3	Multiple Bleeding Detection in WCE [3]	2019	Employs probabilistic region analysis and supervised learning for multi-type bleeding detection, accommodating diffuse and concentrated bleeding.	Suffers from over-segmentation and classification errors in low contrast or shadowed areas.	Development of adaptive thresholding and uncertainty modeling for robust multi-instance predictions.
4	Systematic Review of Bleeding Detection Algorithms [4]	2023	Offers a panoramic synthesis of bleeding detection methods, from traditional to deep learning paradigms, aiding comprehensive understanding.	Lacks experimental validation or benchmarking framework for comparative performance.	Future works should implement benchmarking platforms with standardized evaluation metrics.
5	GMM-based Color	2021	Utilizes statistical modeling	Inconsistent	Demands adaptive



	Feature Extraction [5]		through Gaussian Mixture Models	performance under varied lighting and color distortion due to non-uniform illumination normalization and spatial awareness
10	CHOBBS: Histogram-based Detection [10]	2018	Utilizes block-level statistical color histograms, ensuring low complexity and fast inference suitable for embedded systems.	Does not generalize well to subtle or early-stage bleeding due to reliance on dominant color cues. Could benefit from hybridization with CNNs for better semantic representation of local patterns.
11	Classical Hue-Thresholding Approach [11]	2009	Pioneering technique that sets foundational baselines using simple hue and morphology-based detection rules.	Extremely limited under complex lighting, overlaps with non-bleeding red-colored structures. Should be restructured into preprocessing or assistive components in modern AI pipelines.
12	Swin Transformer + RTDETR Pipeline [12]	2024	Integrates hierarchical attention (Swin Transformer) with efficient object detection (RTDETR), offering high accuracy and speed.	Computationally expensive and requires significant hardware acceleration for real-time use. Future efforts should focus on model compression, quantization, and attention distillation.
13	ColonNet: DenseNet + UNet Hybrid [13]	2024	Combines deep feature extraction with pixel-level localization, achieving superior results in bleeding segmentation and classification.	Demands high memory and computation resources, especially during training with large images. Incorporating lightweight encoders or knowledge distillation can mitigate resource constraints.
14	Two-Stage Divide & Conquer Model [14]	2024	Separates anatomical parsing from bleeding inference, which significantly reduces false positives and enhances contextual reasoning.	Introduces latency due to multi-stage pipeline and additional computational overhead. Joint training of parsing and detection modules or real-time fusion strategies could improve efficiency.
15	ClassifyViStA with Visual Attention [15]	2024	Employs visual attention combined with segmentation masks, yielding both interpretability and precision in detection.	Performance highly depends on the availability of dense segmentation labels for training. Research into weakly-supervised attention mechanisms may reduce dependency on costly annotations.
16	Transformer-Based Bleeding Detection [16]	2024	Captures long-range dependencies and temporal transitions, allowing detection of transient bleeding episodes.	Inference latency and memory requirements are substantial for practical deployment. Future solutions should investigate sparse attention and edge-device optimization for real-time application.
17	Deep Semantic	2018	Effective for small-object	High computational Lightweight



	Segmentation with Pyramid Pooling [17]		segmentation with precise boundary mapping due to multiscale pyramid pooling and dilated convolutions.	complexity limits its integration into embedded or portable devices.	segmentation networks or adaptive feature pruning strategies could bridge this gap.
18	Deep Learning for Multiple Bleeding Types [18]	2023	Dual-branch CNN captures both color and texture aspects, increasing resilience to variations in rare or low- volume bleeding appearance.	Suffers from class imbalance, particularly for rare or low- volume bleeding patterns.	Class rebalancing through generative augmentation or synthetic data training should be explored.
19	CNN-Based Classification with Augmentation [19]	2021	Achieves reliable performance using a simple architecture reinforced by comprehensive data augmentation and dropout techniques.	Limited capability to handle sequential dependencies or subtle bleeding transitions.	Incorporation of temporal modeling or video-based feature learning would enhance continuity understanding.
20	GI Bleeding Classification via Custom CNN [20]	2024	Demonstrates high performance through tailored loss weighting and class-balanced sampling, improving minority class recall.	Requires frequent parameter tuning and architectural adjustments across different datasets.	Meta-learning strategies or adaptive hyperparameter frameworks could yield more flexible models.

III. SYSTEM REQUIREMENTS

To develop and implement the deep learning-based bleeding classification and segmentation framework, the following hardware and software requirements are needed:

3.1 Hardware Requirements:

The system should have a multi-core processor, with a minimum requirement of an Intel Core i5 (10th Gen) or AMD Ryzen 5, while an Intel Core i7/i9 (12th Gen) or AMD Ryzen 9 is recommended for faster computations. Since deep learning models require extensive computations, a dedicated GPU is essential. The minimum GPU requirement is NVIDIA GTX 1650 (4GB VRAM), but for faster training and inference, a NVIDIA RTX 3090 or A100 (24GB+ VRAM) is preferred. For smooth execution, the system should have at least 16GB of DDR4 RAM, though 32GB or more DDR5 RAM is recommended for handling large datasets efficiently. Additionally, a 256GB SSD with a 1TB HDD is the minimum storage requirement, but a 1TB NVMe SSD is ideal for faster data access.

3.2 Software Requirements:

The implementation requires Python 3.8 or later as the primary programming language. For deep learning, the system must have PyTorch 2.0+ or TensorFlow 2.8+ along with CUDA 11.8+ and cuDNN 8.0+ for GPU acceleration. Image processing will be handled using OpenCV and PIL (Pillow), while NumPy and Pandas will be used for data handling. To train and evaluate models, tools like PyTorch Lightning and TensorFlow-Keras are essential. The segmentation models will utilize MONAI and Albumentations for medical image processing. For data visualization, Matplotlib, Seaborn, and TensorBoard will be used. Model development and debugging can be done in Jupyter Notebook, VS Code, or PyCharm.



3.3 Additional Requirements:

It is also recommended to use Conda or Virtualenv for managing dependencies and ensuring a stable working environment. If cloud-based training is required, services like Google Colab Pro, NVIDIA DGX servers can be used. These requirements ensure the efficient training, inference, and evaluation of deep learning models for automated bleeding detection and segmentation in WCE images.

IV. METHODOLOGY

Dataset

The task at hand involves performing classification and segmentation on endoscopic images to detect bleeding vs. non-bleeding regions. The dataset used for this purpose is a custom collection of endoscopic images, consisting of 1309 images labeled as "bleeding" and 1309 images labeled as "non-bleeding," totaling 2618 images. Each image contains important visual information about the presence or absence of bleeding, making it suitable for binary classification tasks.

The dataset is split into two parts:

1. Classification Task: The classification task involves distinguishing between bleeding and non-bleeding images. This is achieved by analyzing the global visual patterns and characteristics of the images, such as color, texture, and overall structure.
2. Segmentation Task: The segmentation task involves identifying specific regions within the image that correspond to bleeding. This requires the model to not only classify the image as bleeding or non-bleeding but also to highlight the exact pixels or regions affected by bleeding.

The images in this dataset are annotated with binary labels: 0 for non-bleeding and 1 for bleeding. The segmentation masks provide pixel-level labels, allowing the model to learn spatial features relevant to detecting bleeding areas.

The dataset is processed with the following preprocessing steps:

- CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied to enhance the contrast of the images, improving the visibility of bleeding areas, especially in low-light conditions.
- Lab Color Space Transformation: The images are converted from the RGB color space to the Lab color space to separate the intensity (L) from the color channels (a and b). This aids in better feature extraction for both classification and segmentation tasks.

The dataset is used for training and evaluating models on two tasks:

1. Binary Classification: The model learns to classify images as either bleeding or non-bleeding.
2. Image Segmentation: The model learns to segment the image, identifying the exact regions that correspond to bleeding.

This dataset provides a comprehensive framework for training deep learning models, combining both global image classification and pixel-wise segmentation tasks. The evaluation metrics for these tasks include accuracy, IoU (Intersection over Union), and Dice coefficient, which assess the performance of the models in both distinguishing bleeding vs. non-bleeding images and accurately identifying bleeding regions in segmentation masks.

4.1 Classification task:

The classification task was performed on the Auto WCEBleedGen dataset, consisting of endoscopic images categorized into two classes: bleeding and non-bleeding. Each class contains a total of 1309 images, which were preprocessed to standardize input formats and enhance model training. The dataset was further divided into training, validation, and test sets. The splits were performed using stratified sampling to ensure that both classes were evenly distributed across each subset, mitigating any class imbalance issues.

The images were stored in directories labeled 'bleeding' and 'non-bleeding', and only image files with .jpg and .png extensions were considered during the dataset preparation process. The train-test-validation split was done in two stages: first splitting the dataset into training+validation and testing sets, and then further splitting the training+validation set into separate training and validation sets. The final split resulted in the following subsets:

- Training set: 80% of the dataset



- Validation set: 10% of the dataset
- Test set: 10% of the dataset

4.1.1 Data Preprocessing and Augmentation:

The images were loaded using Python's PIL library, where each image was converted to RGB format to ensure consistent color channels across the dataset. To prepare the data for input into the model, a series of preprocessing steps and augmentations were applied using PyTorch's torchvision.transforms module.

For the training set, several data augmentation techniques were applied to increase the variability of the input data, which helps improve the generalization ability of the model. These augmentations included:

- Random horizontal flips: To account for the different orientations of endoscopic images.
- Random rotations: To simulate various orientations of the camera and view.
- Random resizing and cropping: To simulate different zoom levels and scales of endoscopic views.
- Normalization: The pixel values were normalized using the mean and standard deviation values of the ImageNet dataset, since the Swin Transformer model was pre-trained on ImageNet. This normalization step ensures that the model's pre-trained weights can be effectively fine-tuned.

4.1.2 Model Selection and Architecture:

For the classification task, we utilized the Swin Transformer model (specifically, swin_tiny_patch4_window7_224 variant) from the timm library. The Swin Transformer is a state-of-the-art vision transformer model that has shown superior performance on various image classification tasks, particularly in medical image analysis, due to its ability to capture long-range dependencies across images efficiently.

We selected the Swin Tiny variant due to its smaller size, which balances efficiency and performance for a dataset of this scale. The model was initially pre-trained on the ImageNet dataset, and the weights from this pre-training were used to fine-tune the model on our endoscopic dataset. The final layer of the model was adjusted to have two output units, corresponding to the binary classification task: bleeding (class 1) and non-bleeding (class 0).

The Swin Transformer (Shifted Window Transformer) is a state-of-the-art model architecture for vision tasks, introduced to improve upon traditional Vision Transformers (ViTs). Unlike earlier ViT architectures that process images in a fully global manner (where every pixel interacts with every other pixel), the Swin Transformer is designed to be more computationally efficient by introducing local windows and hierarchical feature learning. The key innovation of the Swin Transformer lies in its shifted windowing mechanism, which allows it to capture both local and global contexts in a more computationally efficient way than traditional transformers.

4.1.3 Key Concepts Behind the Swin Transformer:

1. Window-based Self-Attention: The Swin Transformer divides the image into small non-overlapping windows (local patches), and within each window, it performs self-attention. This contrasts with the standard Vision Transformer, where attention is computed globally across the entire image.

o Local Attention: Within each window, attention is computed between the patches in that window. This reduces the computational complexity significantly as the attention mechanism is limited to a local region, rather than the entire image.

o Computational Efficiency: Traditional transformers compute attention between all patches, which has a time complexity of $O(N^2)O(N^2)O(N^2)$ where NNN is the number of patches. The window-based approach in Swin Transformer reduces this to $O(N \cdot M)O(N \cdot M)O(N \cdot M)$, where MMM is the window size, making it computationally more efficient.

2. Shifted Windowing Scheme: One of the primary innovations in the Swin Transformer is the shifted window scheme, which enhances the model's ability to capture long-range dependencies without incurring the high computational cost of global self-attention.



o In a typical self-attention mechanism, each token (patch) attends to every other token. This is computationally expensive for high-resolution images. In Swin Transformer, images are first divided into non-overlapping windows, and attention is calculated locally within each window.

o To enable interaction between adjacent windows, the windows are shifted between successive layers. This shift allows for cross-window connections, providing the model with a mechanism to learn long-range dependencies without the need for global attention.

o The window shift operation ensures that patches from different windows can interact with each other while maintaining efficient computation.

3. Hierarchical Feature Representation: Swin Transformer uses a hierarchical architecture, meaning the feature map size decreases at each stage, similar to how CNNs downsample feature maps. This allows Swin Transformer to capture features at multiple scales, which is essential for tasks like object detection and segmentation. Each layer of the Swin Transformer is designed to operate at progressively larger scales by increasing the patch size and reducing the spatial resolution.

o Patch Merging Layers: After a few stages, the Swin Transformer performs a patch merging operation, which reduces the spatial resolution while increasing the number of channels, similar to the pooling operations in CNNs. This operation allows the model to capture higher-level features as it progresses through deeper layers.

4. Transformer Encoder Blocks: The Swin Transformer architecture consists of multiple transformer encoder blocks. These blocks are repeated multiple times to create a deep architecture. Each transformer block contains:

o Window-based Multi-Head Self-Attention (W-MSA): Computes attention within the local windows.

o Shifted Window-based Multi-Head Self-Attention (SW-MSA): In the subsequent layer, the windows are shifted to allow interaction between adjacent windows.

o Feed-Forward Networks (FFN): After attention, the model applies a simple feed-forward network to process the features further.

o Layer Normalization and Residual Connections: To stabilize training and improve performance, Swin Transformer uses layer normalization and residual connections.

5. Efficient and Scalable: The key benefit of Swin Transformers over the original ViTs is their scalability and efficiency. The window-based self-attention significantly reduces the computational burden, and the hierarchical design allows the model to scale to higher resolutions. These attributes make Swin Transformers particularly suitable for vision tasks like image classification, object detection, and semantic segmentation.

4.1.4 Architecture of the Swin Transformer:

The Swin Transformer model consists of four stages, each operating at a different spatial resolution:

1. Stage 1:

o Initial patch embedding and the first set of attention layers.

o The image is divided into small patches, and local window-based attention is applied.

o The output of this stage is a set of features that are used as input to the next stage.

2. Stage 2:

o More window-based attention is applied, and the spatial resolution is reduced (via patch merging).

o This stage captures slightly higher-level features and continues the hierarchical processing.

3. Stage 3:

o The model continues to reduce the spatial resolution while increasing the depth of the feature maps.

o Attention is computed across larger windows to capture broader contextual information.

4. Stage 4:

o The final stage applies the deepest level of processing, where the model learns the most abstract features.

o The output from this stage is then passed through a classification head or other task-specific heads (for segmentation, detection, etc.).



Each of these stages is connected by a series of multi-head self-attention blocks, feed-forward layers, and residual connections. These stages allow the Swin Transformer to gradually build up features from local to global contexts while maintaining computational efficiency.

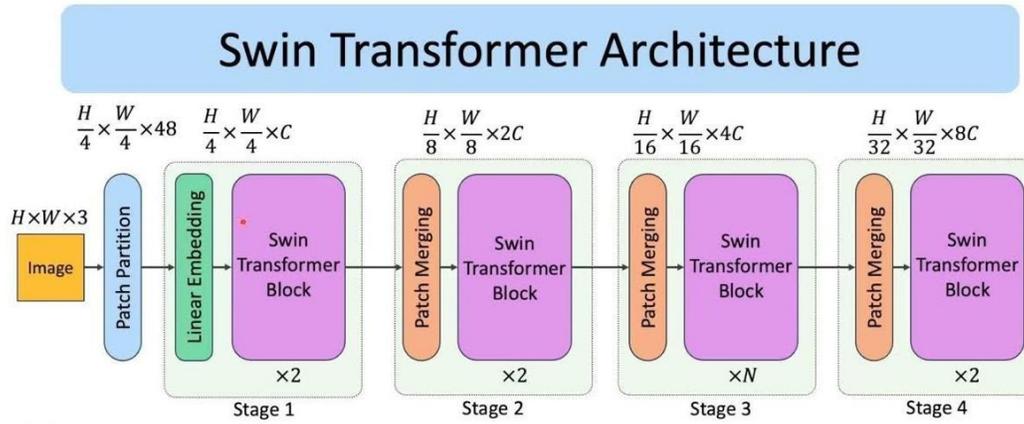


Figure 4.1.1 Swin-Transformer Architecture

In terms of fine-tuning, we froze the first two stages of the Swin Transformer to retain the learned features from the pre-training. The weights from these initial layers were kept fixed to reduce the computational cost and prevent overfitting on the smaller dataset. The later stages (3 and 4) were unfrozen to allow for further fine-tuning on our specific task.

Loss Function and Optimizer:

The model was trained using CrossEntropyLoss, which is appropriate for multi-class classification tasks, where the network outputs a probability distribution over the two classes. The AdamW optimizer was used with weight decay regularization to optimize the model parameters. The learning rate was set to $1e-4$, and weight decay was set to $1e-4$ to prevent overfitting.

To prevent the learning rate from decreasing too quickly, we employed the ReduceLROnPlateau scheduler, which reduces the learning rate by a factor of 0.5 if the validation loss does not improve after a set number of epochs.

4.1.5 Training and Validation:

The model was trained for 20 epochs, where during each epoch, the model's weights were updated based on the computed gradients. The training loop involved feeding the model with images in mini-batches, computing the loss using CrossEntropyLoss, performing backpropagation, and updating the model's parameters using the optimizer.

The performance was evaluated on the validation set after each epoch. The model's accuracy and loss were tracked during both training and validation. At the end of each epoch, the model's weights were saved if the validation accuracy improved compared to previous epochs, ensuring that the best performing model was retained.

4.2 Segmentation Using CBAM-Enhanced U-Net++:

The segmentation component of this project focuses on detecting bleeding areas in medical endoscopic images. To achieve this, we use an advanced neural network architecture called U-Net++ that is enhanced with CBAM (Convolutional Block Attention Module). This section provides a thorough explanation of the segmentation model architecture, the loss function, the dataset preparation, and the training process.

4.2.1 Model Architecture: Attention U-Net++

The model used for segmentation is based on U-Net++, a variant of the classic U-Net architecture, which is widely used for image segmentation tasks. The U-Net++ model is specifically designed to enhance the segmentation performance through nested skip pathways that provide dense connections between encoder and decoder layers. These pathways



allow for more effective feature reuse and better gradient flow during training, which are crucial for precise segmentation in medical images.

Incorporating CBAM (Convolutional Block Attention Module) into the U-Net++ model further enhances its performance by enabling the model to focus on relevant regions of the image. CBAM applies both channel attention and spatial attention, which help the model selectively highlight important features and suppress irrelevant ones.

1. Channel Attention: Channel attention works by analyzing the feature maps produced by the network and adjusting the importance of each channel. It does this by calculating the average and maximum pooled versions of the feature map and passing them through convolutional layers. The result is a set of attention weights that dynamically scale the feature maps, emphasizing more informative channels.

2. Spatial Attention: Spatial attention focuses on the spatial location of the features, identifying which regions of the image are important for the task at hand. This attention mechanism uses both average and maximum pooling to create an attention map that highlights important spatial areas (such as regions containing bleeding). This helps the network concentrate on the most relevant areas, thereby improving segmentation accuracy.

By applying both channel and spatial attention in the decoder layers of U-Net++, the model is better able to segment the bleeding regions of endoscopic images with higher accuracy.

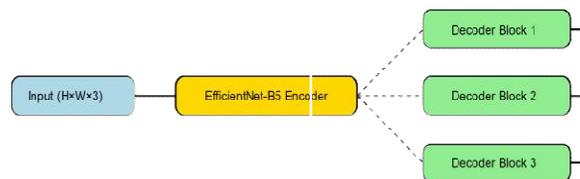


Figure 4.2.1 Attention UNet++ Architecture

Loss Function: Combined Loss

In segmentation tasks, especially in medical imaging, it is common to encounter class imbalance, where the foreground (e.g., bleeding regions) is much smaller than the background (e.g., non-bleeding regions). To address this, we use a combined loss function that incorporates both Dice Loss and Focal Loss.

1. Dice Loss: Dice Loss measures the overlap between the predicted segmentation mask and the ground truth mask. It is particularly well-suited for binary segmentation tasks where the goal is to distinguish between two classes (e.g., bleeding vs. non-bleeding). The loss value ranges from 0 to 1, with 1 indicating perfect overlap and 0 indicating no overlap.

2. Focal Loss: Focal Loss is a modification of the standard cross-entropy loss designed to address the issue of class imbalance. It focuses more on hard-to-classify examples (such as small or faint bleeding areas) by down-weighting the loss for well-classified pixels. This encourages the model to focus on the difficult cases, which are often the ones most critical in medical image analysis.

By combining these two losses, the model benefits from the accuracy-focused Dice Loss and the hard-example-focusing Focal Loss, which together help the model achieve better segmentation performance in imbalanced datasets.

Dataset and Data Augmentation:

The dataset for this project consists of endoscopic images along with their corresponding binary segmentation masks. Each image is labeled with pixel-wise annotations, where each pixel is marked as either part of the bleeding region or part of the non-bleeding region. These masks are used as the ground truth for training and evaluation.

To improve model generalization and prevent overfitting, we apply a series of data augmentation techniques using the Albumentations library. Data augmentation involves randomly applying transformations to the images and masks during training, such as:

- Horizontal flipping: Randomly flipping the image horizontally to simulate different perspectives.
- Rotation: Randomly rotating the image within a specified range to make the model invariant to rotation.



- Brightness and contrast adjustments: Randomly varying the brightness and contrast to simulate changes in lighting conditions.
 - Resizing: Standardizing the size of the images by resizing them to a fixed resolution (e.g., 256x256 pixels).
- These transformations help the model learn more robust features by exposing it to a variety of possible image variations, thereby improving its ability to generalize to unseen data.

Training Process:

The model is trained using the AdamW optimizer, which is a variant of the Adam optimizer with weight decay regularization. This optimizer is well-suited for training deep learning models as it adapts the learning rate based on the gradients of the loss function.

In addition to the AdamW optimizer, we use mixed precision training to speed up the training process and reduce memory usage. Mixed precision training involves using both 16-bit and 32-bit floating-point operations to compute gradients, which significantly improves computational efficiency without sacrificing model accuracy.

To further enhance the training process, we use a learning rate scheduler called ReduceLROnPlateau. This scheduler reduces the learning rate if the validation loss plateaus for a specified number of epochs, helping the model converge faster and avoid overfitting.

During training, the model's performance is evaluated using two key metrics for segmentation:

1. Dice Score: The Dice coefficient is used to measure the similarity between the predicted segmentation mask and the ground truth mask. A higher Dice score indicates better performance.
2. IoU (Intersection over Union): IoU measures the overlap between the predicted and true segmentation regions. It is a widely used metric for evaluating segmentation performance, especially in tasks involving binary classification.

The model is trained for a specified number of epochs, with the best performing model (the one with the lowest validation loss) saved for later use. The training loop involves feeding the images and masks through the network, computing the loss, and backpropagating the gradients to update the model's parameters. This process is repeated for several epochs until the model converges.

Evaluation and Model Saving:

After training, the model's performance is evaluated on a separate test set to assess its ability to generalize to new, unseen data. The best model (based on validation loss) is saved and can later be used for making predictions on new images.

Image Classification

1. Image Loading: The first step is to load the image that has been uploaded. This image is read and converted to the RGB color format to ensure it has three color channels.
2. Preprocessing: Once the image is loaded, it undergoes preprocessing:
 - o Resizing/Padding: The image is resized to a fixed size of 224x224 pixels, which is required by the classification model. If the aspect ratio of the image is not 1:1, padding is added to maintain the original proportions.
 - o Normalization: The pixel values of the image are normalized to ensure they align with the expected range for the pre-trained model (in this case, normalization typically centers the values around a mean and scales them based on standard deviation values that are specific to the pre-trained model).
3. Model Inference (Classification): The preprocessed image is then passed through the classification model (a Swin Transformer). This model outputs two values (logits), representing the likelihood of the image belonging to each class (e.g., "Bleeding" and "Non-Bleeding").
 - o Prediction: The class with the highest likelihood is chosen as the predicted class. The image is then classified as either "Bleeding" or "Non-Bleeding."

2. Image Segmentation (if Bleeding)

If the classification step predicts the image as "Bleeding," the next step is to perform segmentation on the image to highlight the bleeding region.



1. Letterbox Resizing: To process the image for segmentation, the image is resized while maintaining its aspect ratio, and padding is added to ensure that it fits the required size for the segmentation model. This ensures that the image dimensions are consistent while preserving the original proportions of the content in the image.
2. Segmentation Model Inference: The padded image is then passed through the segmentation model (an Attention U-Net++). This model generates a mask that highlights the regions of the image where bleeding is detected.
 - o The model generates a probability map indicating the likelihood that each pixel in the image is part of the bleeding region.
 - o A threshold is applied to convert these probabilities into a binary mask: pixels with a probability higher than the threshold are considered part of the bleeding region.
3. Post-Processing: After the mask is generated, it is post-processed to remove any padding that was added during resizing. The mask is cropped to the region that corresponds to the original content of the image, and then resized back to the original image dimensions.

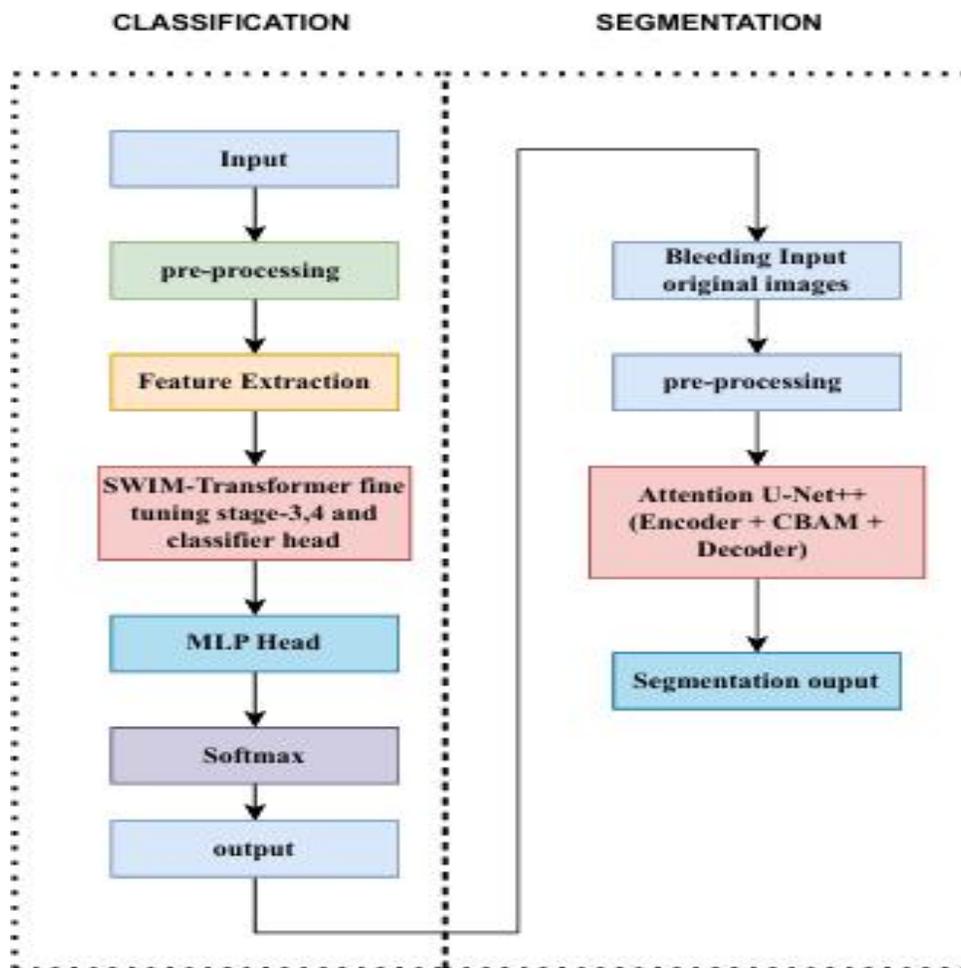


Figure 4.3.1 Proposed model workflow

3. Visualization of Results

Once the segmentation mask is generated, it can be overlaid on the original image to visualize the bleeding region:

1. Overlay Creation: The segmentation mask is used to create a red-colored overlay that highlights the bleeding region. This red mask is combined with the original image, allowing the user to see the exact area where bleeding is detected.



2. Displaying Results: The original image, along with the segmentation overlay and the binary segmentation mask, can be displayed. If the image was classified as "Non- Bleeding," the segmentation step is skipped

V. RESULTS

5.1 Result of Classification task:

The classification results demonstrate the effectiveness of different deep learning models in distinguishing between bleeding and non-bleeding images from the WCEBleedGen dataset. The Swin Transformer achieved the highest accuracy (99.43%) and F1-score (0.9943), indicating its superior feature extraction capabilities due to shifted window attention and hierarchical representation learning. ResNet50 followed closely with an accuracy of 99.02% and a perfect precision-recall balance, making it a robust alternative. DenseNet121 performed slightly lower (98.09% accuracy), but its recall (0.9771) suggests it effectively identifies positive cases while maintaining compact feature propagation. MobileViT achieved 97.33% accuracy, leveraging both CNN and ViT architectures, but its relatively lower recall (0.9580) indicates occasional false negatives. ViT Transformer, with the lowest accuracy (90.84%), struggled with generalization due to its reliance on patch-based tokenization without spatial inductive bias, leading to a trade-off between precision (0.8567) and recall (0.9809). Overall, Swin Transformer emerges as the best-performing model, demonstrating its suitability for medical image classification, particularly in endoscopic bleeding detection.

5.1.1 Comparison of Classification results:

Models	Accuracy	Precision	Recall	F1 Score
Swin- Transformer	99.43%	0.9924	0.9962	0.9943
Resnet50	99.02%	0.9962	0.9962	0.9962
Densenet121	98.09%	0.9846	0.9771	0.9808
Mobilevit	97.33%	0.9882	0.9580	0.9729
Vit Transformer	90.84%	0.8567	0.9809	0.9146

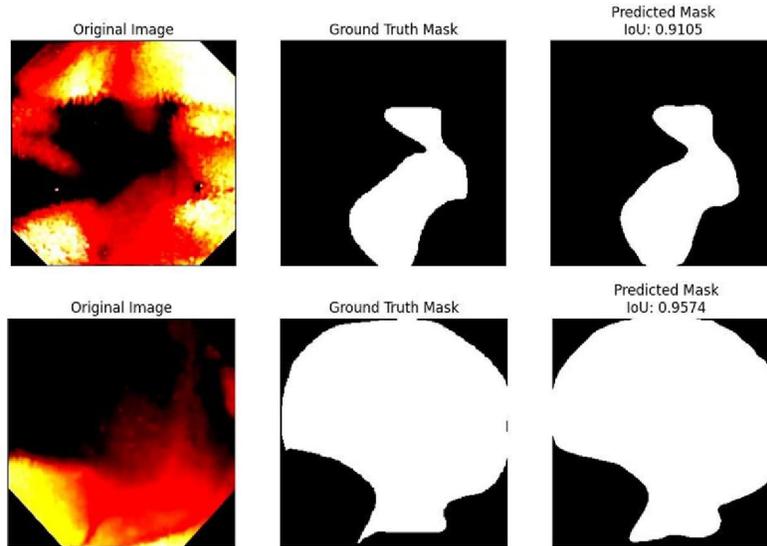
5.2 Result of Segmentation task:

The segmentation results highlight the performance of different models in accurately delineating bleeding regions in WCEBleedGen images. UNet++ achieved the highest Dice score (0.9750) and Intersection over Union (IoU = 0.8686), demonstrating its superior ability to capture fine-grained details by leveraging dense skip connections and an enhanced nested architecture. This indicates that UNet++ can effectively segment bleeding areas with minimal false positives and false negatives. In contrast, UNet exhibited significantly lower performance (Dice: 0.7023, IoU: 0.5478) due to its simpler architecture, which lacks the extensive feature refinement present in UNet++. DeepLabV3+, although designed for high-level semantic segmentation, underperformed (Dice: 0.6924, IoU: 0.5604), likely due to its reliance on dilated convolutions, which may not be optimal for fine-structured medical image segmentation. Overall, the results indicate that UNet++ is the most effective model for bleeding segmentation, making it well-suited for precise localization and quantification of bleeding regions in endoscopic images.

5.2.1 Comparing of Segmentation results:

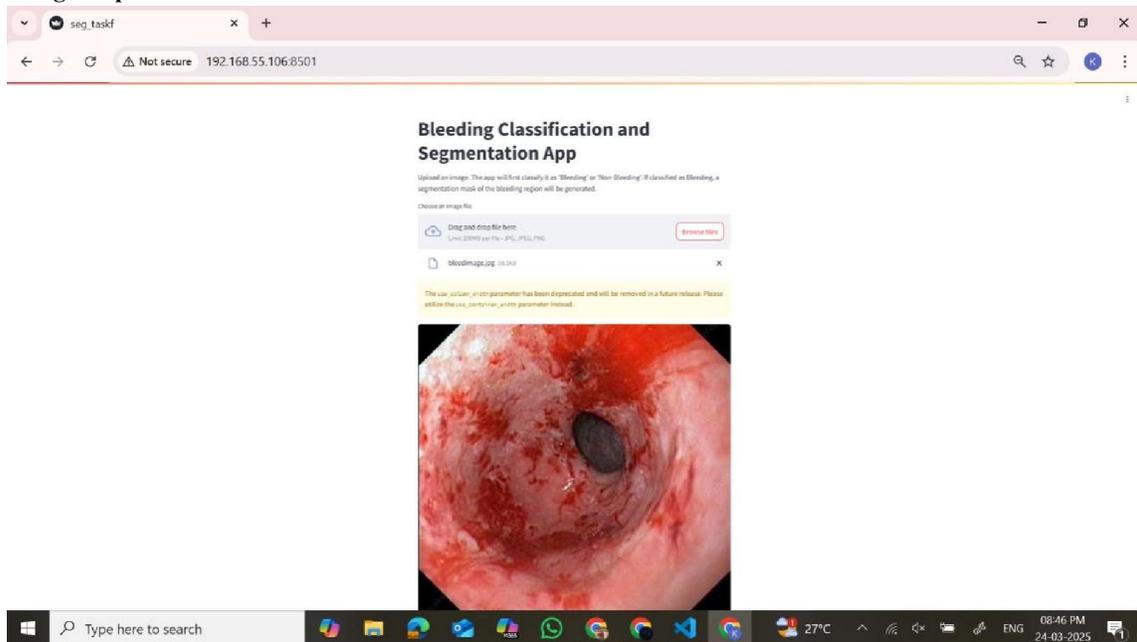
Models	Dice score	IoU
Unet++ with CBAM	0.9750	0.8686
unet	0.7023	0.5478
Deeplabv3+	0.6924	0.5604



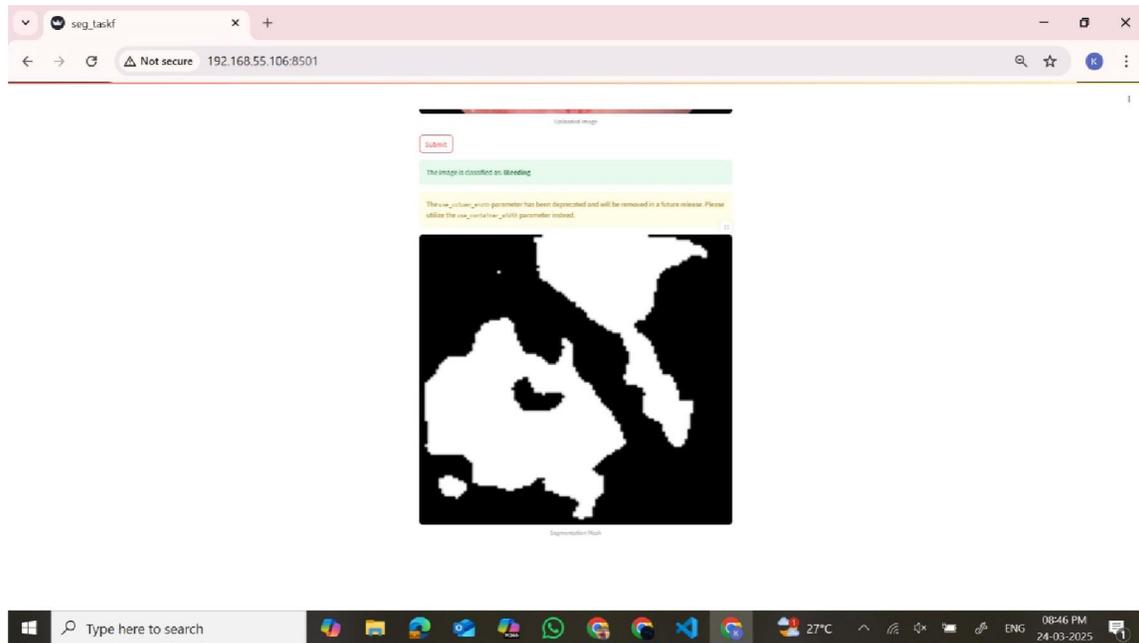


5.3 Output Screenshots

Bleeding sample:



Non-bleeding sample:



VI. CONCLUSION AND FUTURE SCOPE

In this study, we proposed an advanced deep learning-based framework for automated bleeding detection and segmentation in Wireless Capsule Endoscopy (WCE) images. Our approach integrates Swin Transformer for classification and Attention U-Net++ with CBAM for segmentation, ensuring high accuracy and precise localization of bleeding regions. The WCEBleedGen dataset was utilized, with enhanced preprocessing techniques such as CLAHE-based contrast adjustment and Gaussian blurring to improve image quality and feature extraction. Experimental results demonstrated the superior performance of Swin Transformer, achieving 99.43% accuracy in distinguishing bleeding from non-bleeding images, outperforming other models like ResNet50, DenseNet121, and ViT. Similarly, UNet++ with CBAM proved to be the most effective segmentation model, achieving a Dice score of 0.9750 and IoU of 0.8686, significantly outperforming traditional UNet and DeepLabV3+. The incorporation of CBAM and EfficientNet as the backbone further improved segmentation accuracy by refining feature selection. Our findings highlight the potential of deep learning in revolutionizing medical diagnostics by significantly reducing manual workload, enhancing diagnostic accuracy, and minimizing human error. The proposed AI-assisted system offers clear visualizations of bleeding regions, making it an invaluable tool for healthcare professionals in early detection and treatment planning of GI bleeding. For future work, we recommend expanding the dataset diversity, incorporating real-time processing capabilities, and exploring multi-modal approaches (e.g., combining WCE with clinical data) to further enhance system reliability and speed. With continuous advancements in AI-driven methodologies, automated GI diagnostics will become more efficient, accurate, and accessible, ultimately improving patient outcomes and advancing gastrointestinal healthcare.

REFERENCES

- [1] Brzeski, A., Dziubich, T., & Krawczyk, H. (2023). Visual Features for Improving Endoscopic Bleeding Detection Using Convolutional Neural Networks. *Sensors*.
- [2] Liu, Z., Hu, C., & Shen, Z. (2019, December). Research on a new feature detection algorithm for wireless capsule endoscope bleeding images based on superpixel segmentation. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (pp. 17441749). IEEE.



- [3] Bchir, O., Ben Ismail, M. M., & AlZahrani, N. (2019). Multiple bleeding detection in wireless capsule endoscopy. *Signal, Image and Video Processing*, 13, 121126
- [4] Musha, A., Hasnat, R., Mamun, A. A., Ping, E. P., & Ghosh, T. (2023). ComputerAided Bleeding Detection Algorithms for Capsule Endoscopy: A Systematic Review. *Sensors*.
- [5] Rathnamala, S., & Jenicka, S. (2021). Automated bleeding detection in wireless capsule endoscopy images based on color feature extraction from Gaussian mixture model superpixels. *Medical & biological engineering & computing*.
- [6] Musha, A., Hasnat, R., Mamun, A. A., Ping, E. P., & Ghosh, T. (2023). ComputerAided Bleeding Detection Algorithms for Capsule Endoscopy: A Systematic Review. *Sensors*.
- [7] Ghosh, T., & Chakareski, J. (2021). Deep transfer learning for automated intestinal bleeding detection in capsule endoscopy imaging. *Journal of Digital Imaging*
- [8] Handa, P., Dhir, M., Mahbod, A., Schwarzahans, F., Woitek, R., Goel, N., & Gunjan, D. (2024). WCEbleedGen: A wireless capsule endoscopy dataset and its benchmarking for automatic bleeding classification, detection, and segmentation.
- [9] Pogorelov, K., Suman, S., Azmadi Hussin, F., Saeed Malik, A., Ostroukhova, O., Riegler, M., ... & Goh, K. L. (2019). Bleeding detection in wireless capsule endoscopy videos—Color versus texture features. *Journal of applied clinical medical physics*.
- [10] Ghosh, T., Fattah, S. A., & Wahid, K. A. (2018). CHOBS: Color histogram of block statistics for automatic bleeding detection in wireless capsule endoscopy video. *IEEE journal of translational engineering in health and medicine*.
- [11] Li, B., & Meng, M. Q. H. (2009). Computeraided detection of bleeding regions for capsule endoscopy images. *IEEE Transactions on biomedical engineering*.
- [12] Alavala, S., Vadde, A. K., Kancheti, A., & Gorthi, S. (2024). A Robust Pipeline for Classification and Detection of Bleeding Frames in Wireless Capsule Endoscopy using Swin Transformer and RTDETR.
- [13] Singh, A., Prakash, S., Das, A., & Kushwaha, N. (2024). ColonNet: A Hybrid Of DenseNet121 And UNET Model For Detection And Segmentation Of GI Bleeding.
- [14] Lin, Y. F., Qiu, B. C., Lee, C. M., & Hsu, C. C. (2024). Divide and Conquer: Grounding Bleeding Areas in Gastrointestinal Image with TwoStage Model.
- [15] Balasubramanian, S., Abhishek, A., Krishna, Y., & Gera, D. (2024). ClassifyViStA: WCE Classification with Visual Understanding through Segmentation and Attention.
- [16] Alawode, B., Hamza, S., Ghimire, A., & Velayudhan, D. (2024). TransformerBased Wireless Capsule Endoscopy Bleeding Tissue Detection and Classification.
- [17] Ghosh, T., Li, L., & Chakareski, J. (2018, October). Effective deep learning for semantic segmentation based bleeding zone detection in capsule endoscopy images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*.
- [18] Bchir, O., Alkhudhair, G. A., Alotaibi, L. S., Almhizea, N. A., Almuhan, S. M., & Alzeer, S. F. (2023). Deep Learningbased Multiple Bleeding Detection in Wireless Capsule Endoscopy. *International Journal of Advanced Computer Science and Applications*.
- [19] Rustam, F., Siddique, M. A., Siddiqui, H. U. R., Ullah, S., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Wireless capsule endoscopy bleeding images classification using CNN based model. *IEEE access*.
- [20] Rani, K. A., Gowrishankar, S., & Amrutesh, A. (2024, March). Gastrointestinal Bleeding Classification in Capsule Endoscopy Images using Deep Learning. In *2024 4th International Conference on Data Engineering and Communication Systems (ICDECS)* (pp. 16). IEEE.

