Improving U-Net Performance in Medical Image Segmentation Using Intuitionistic Fuzzy Set: A Case Study on Video Capsule Endoscopy Polyp Images

Chih-Ping Yen [1,*], Tai-Ping Hsing [2], Chen-Yu Li [3], Kai-Lun Huang [4]

[1, 2, 3] Department of Information Management, Central Police University, Taoyuan 33304, Taiwan, R.O.C.

[4] 6th Investigation Corps, Criminal Investigation Bureau, Taipei 110055, Taiwan, R.O.C.

* Corresponding Author
E-mail address: peter@mail.cpu.edu.tw

Abstract

Medical image segmentation plays a pivotal role in aiding diagnosis and clinical decision-making. However, Video Capsule Endoscopy (VCE) images often present significant challenges for accurate segmentation due to inherent low contrast, uneven illumination, and blurry lesion boundaries. To address these issues and enhance segmentation performance, this study proposes a 5-channel input tensor feature enhancement strategy based on Intuitionistic Fuzzy Set (IFS) theory, aiming to optimize the performance of the U-Net model in semantic segmentation tasks for VCE polyp images. Our method extracts features such as intensity, hue, saturation, and texture from image pixels, and subsequently calculates their membership and non-membership degrees. These fuzzy features are then combined with the original RGB three channels to construct a five-channel input tensor, serving as the input for the U-Net model. We evaluated the model using the publicly available Kvasir-Capsule dataset. Experimental results demonstrate that, compared to the baseline model utilizing only original RGB input, the model employing IFS-enhanced input achieved significant improvements across key metrics including Overall Accuracy (0.8261 vs. 0.8064), Recall (0.9304 vs. 0.9179), Dice Similarity Coefficient (0.8668 vs. 0.8636), and Intersection over Union (IoU) (0.7771 vs. 0.7648). This validates that IFS can effectively handle uncertainty and fuzzy information within images, substantially enhancing the model's ability to capture true lesions, particularly demonstrating an advantage in reducing missed diagnoses. This research provides a feasible and robust direction for improving the precision of medical image segmentation.

Keywords: Deep Learning, Medical Image Segmentation, U-Net, Intuitionistic Fuzzy Set (IFS), Video Capsule Endoscopy (VCE).

1. Introduction

Medical image analysis plays a pivotal role in modern clinical medicine, providing essential visual information for early disease diagnosis, precise treatment planning, and objective prognosis

assessment. Medical imaging encompasses various modalities, including endoscopic imaging, X-ray, computed tomography (CT), positron emission tomography (PET), nuclear medicine imaging, ultrasound, and magnetic resonance imaging (MRI). Through automated processing and analysis of medical images, healthcare professionals can more efficiently identify lesions and assess disease progression, thereby enhancing the accuracy and efficiency of medical decision-making **[1-2]**. Currently, medical image analysis tasks are primarily categorized into two main types: classification and segmentation. Image classification aims to determine the presence of specific lesions within an image or to assign images to predefined categories, such as identifying whether an image contains a tumor [3]. Image segmentation, in contrast, is more refined, with its objective being to precisely delineate specific regions of interest (such as organs, tumors, or lesions) from the background at a pixel level. This is indispensable for quantifying lesion size, morphological analysis, and guiding interventional procedures [4].

In recent years, with the rapid advancements in deep learning technologies, particularly the rise of Convolutional Neural Networks (CNNs), the automation level of medical image analysis has significantly improved [5]. CNNs have demonstrated powerful capabilities in handling complex image data and have been widely applied to both classification and segmentation tasks in medical imaging. For classification tasks, models can automatically learn high-level features from images to distinguish between normal and abnormal cases; for segmentation tasks, networks like U-Net [6], with their unique encoder-decoder architecture and skip connections, have become the standard in medical image semantic segmentation, achieving breakthrough progress in the precise delineation of various anatomical structures and lesions [7].

However, despite the immense success of deep learning in medical image analysis, the inherent complexities of certain imaging modalities continue to pose challenges. This study will focus on polyp segmentation analysis in Video Capsule Endoscopy (VCE) images as a case study. VCE technology is increasingly prevalent in gastrointestinal disease screening and diagnosis due to its non-invasive nature, ease of operation, and ability to provide panoramic views of the digestive tract [8]. Nevertheless, intrinsic image quality issues in VCE, such as uneven illumination, low contrast, motion blur, limited field of view, and interference from food debris or air bubbles, significantly complicate the automated identification and precise segmentation of polyps. These issues often lead to a decline in model performance when dealing with ambiguous boundaries and minute lesions against complex backgrounds [9].

To overcome these challenges and further enhance the accuracy and robustness of polyp segmentation in VCE images, this study introduces Intuitionistic Fuzzy Set (IFS) theory as a novel data pre-processing strategy. IFS is capable of simultaneously capturing membership (μ), non-membership (ν), and hesitation (π) degrees within data, thereby providing a more comprehensive quantification of inherent fuzziness and uncertainty in images [10]. This offers potential advantages for handling the complex characteristics of VCE images. We propose an IFS-based 5-channel image pre-processing method, designed to provide the U-Net model with richer and more robust

feature representations against uncertainty. This method expands the original Red-Green-Blue (RGB) images into a 5-channel input tensor, comprising the original RGB channels along with the final membership (μ_{final}) and final non-membership (ν_{final}) channels computed via IFS theory. We will input this enhanced data into a U-Net model and conduct comprehensive experimental evaluations on the publicly available Kvasir-Capsule dataset (which includes polyp images professionally annotated by the Jha research team [11-12]).

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 elaborates on our proposed framework; Section 4 presents and analyzes the experimental results; finally, Section 5 concludes this study.

2. Related Work

2.1. Medical Image Segmentation

Medical image segmentation is a fundamental and crucial task in Computer-Aided Diagnosis (CAD) and treatment planning systems. Its primary objective is to precisely delineate specific regions of interest (such as organs, tumors, lesions, or tissue structures) from complex backgrounds within medical images, thereby assisting clinicians in accurate lesion identification, size measurement, localization, and evaluation of their relationship with surrounding tissues [13]. However, compared to natural images, medical images present unique challenges, such as low resolution, poor contrast, internal inconsistencies, and scattered target region distributions, while simultaneously demanding more stringent requirements for the accuracy and stability of segmentation results [4].

Early medical image segmentation methods were primarily based on traditional image processing techniques. These approaches included Thresholding, Region Growing, Edge Detection, and methods based on Active Contour Models (ACM) [14]. While these techniques achieved some success in specific applications, they generally relied on handcrafted features and were highly sensitive to image quality issues (e.g., noise, uneven illumination). Their robustness and generalization capabilities were often limited when dealing with highly heterogeneous and complex real-world medical images. Subsequently, Machine Learning (ML) methods, such as Support Vector Machines (SVM) and Random Forests classifiers, were introduced into the field of medical image segmentation to improve segmentation performance by learning more complex pixel-level features [15]. Nevertheless, the complexity of feature engineering and computational efficiency remained bottlenecks when these methods processed large volumes of high-dimensional image data.

In recent years, the advent of Deep Learning (DL) technologies, particularly the widespread application of Convolutional Neural Networks (CNNs), has revolutionized the field of medical image segmentation. CNNs, with their powerful automatic feature extraction capabilities, can learn multi-layered, rich, abstract feature representations directly from raw data, significantly enhancing segmentation accuracy and robustness [16]. Among these, Fully Convolutional Networks (FCN) laid the foundation for end-to-end pixel-level segmentation, while U-Net [6] and its numerous

variants, owing to their unique encoder-decoder architecture and skip connections, have proven particularly effective in medical image segmentation tasks, becoming the gold standard in the field. Despite the remarkable achievements of deep learning methods, further exploration is still needed when facing challenges such as the scarcity of large-scale high-quality annotated data, model generalization capability, computational resource consumption, and result interpretability [4].

2.2. Video Capsule Endoscopy (VCE) Polyp Segmentation

Polyps are common lesions in the digestive tract, and if not discovered and resected promptly, they may develop into malignant tumors. VCE, as a non-invasive examination method, offers convenience for gastrointestinal tract inspection. However, its vast image sequences lead to inefficient and fatiguing manual screening. Therefore, developing automated VCE polyp detection and segmentation systems holds significant clinical importance [17]. VCE image analysis faces multiple complexities, making automated polyp segmentation even more challenging. These challenges include: the inherent diversity in polyp morphology (significant variations in size, shape, color, and texture), the complexity of the imaging environment (such as uneven lighting within the intestine, lens reflections, motion blur), and interference from artifacts (e.g., air bubbles, food residues, fluid occlusion) [18]. Furthermore, some flat or diminutive polyps have indistinct visual features and blurred boundaries, making them highly susceptible to confusion with normal bowel wall tissue [19].

Early VCE polyp segmentation research primarily focused on methods based on handcrafted features. These methods commonly utilized image color features (e.g., red component analysis) combined with morphological operations to identify potential polyp regions [20], or employed texture analysis (such as Local Binary Patterns, LBP [21]) and shape features to differentiate lesions. Although these traditional methods achieved some success on specific and controlled datasets, they generally lacked robustness to the diversity of images in real clinical environments, struggling to cope with the aforementioned complex challenges, leading to insufficient segmentation precision and generalization ability [22].

In recent years, with the maturation of deep learning technologies, deep learning-based methods have become the mainstream trend for VCE polyp segmentation, achieving significant progress [PolypReview]. Researchers commonly adopt Convolutional Neural Networks (CNNs) as backbone networks and integrate various semantic segmentation architectures, such as Fully Convolutional Networks (FCN) [23], U-Net [24] and its numerous variants, or Mask R-CNN, among others [25]. These models, through an end-to-end learning approach, can automatically extract high-level, semantically rich features from VCE images. To further enhance model performance in complex VCE images, researchers have also explored multi-scale feature fusion, attention mechanisms, and advanced data augmentation strategies [26]. Despite these advancements, the VCE field continues to face challenges, such as the scarcity of high-quality pixel-level annotated data, the generalization ability of models on unseen data, and the precise

identification and boundary definition of minute or flat polyps [18]. Therefore, how to further enhance the model's ability to process ambiguous and uncertain information in VCE images is a pressing issue that current research needs to address.

2.3. Applications of Intuitionistic Fuzzy Sets (IFS) in Image Processing

In recent years, fuzzy sets have become a focus of artificial intelligence (AI) research due to their excellent interpretability. Related practical applications include industrial control, security, medical diagnosis, human-computer interaction (HCI), reinforcement learning, education, and finance [27]. IFS theory, proposed by Atanassov [10] in 1986, is an extension of classical fuzzy set theory. Unlike fuzzy sets, which only consider the membership degree, IFS simultaneously introduces the non-membership degree and the hesitation degree, enabling it to describe and handle uncertainty and imprecision in data more comprehensively and flexibly. In the field of image processing, IFS has been widely applied to various tasks [28]:

1) Image Enhancement: IFS is utilized to adjust image brightness and contrast, thereby enhancing details. For instance, some methods employ IFS operators to manage pixel fuzziness, consequently improving image quality, particularly when dealing with medical images that often exhibit issues like blurriness and low contrast [29-31].
2) Image Segmentation and Edge Detection: IFS shows potential in capturing image boundaries and regional information. Researchers leverage IFS to construct fuzzy relations for clustering or distinguishing image pixels, leading to more robust segmentation and edge detection in noisy environments [32-34].
3) Image Denoising and Feature Extraction: IFS can be used to build adaptive filters and extract features more robust to uncertainty by analyzing pixel properties, thereby identifying and quantifying fuzzy information in images [35-37].

Despite the unique advantages demonstrated by IFS in the aforementioned image processing tasks, most existing studies treat IFS as an independent image processing step or combine it with traditional machine learning methods [38]. Few studies have systematically explored deeply integrating IFS theory as an input preprocessing step for deep learning models, especially for medical image segmentation tasks. Traditional deep learning models typically process raw RGB or grayscale images directly, failing to fully leverage IFS's capability to quantify inherent uncertainty and fuzziness in images [39]. This study aims to fill this research gap by introducing an IFS-based multi-channel image preprocessing strategy, which inputs original images along with aggregated membership and non-membership information extracted by IFS into a U-Net model, with the aim of significantly enhancing the model's performance and robustness in the complex and challenging task of VCE polyp segmentation.

3. Methodology

This study's framework aims to enhance the automated semantic segmentation performance

of VCE polyps by combining IFS image enhancement techniques with a deep learning model. As illustrated in Figure 1, the entire process can be divided into two main stages: "IFS Image Enhancement" and "Performance Evaluation." In the "IFS Image Enhancement" stage, raw images from the KvasirCapsule-SEG dataset are first loaded (Subsection 3.1) and subsequently converted to the HSV color model (Subsection 3.2). Following this, core intuitionistic fuzzy image processing is performed, where the original RGB image is combined with the final membership and non-membership channels, computed via IFS theory, to construct 5-channel input tensor, which is then split into training images and testing images (Subsection 3.3). The process then moves to the "Performance Evaluation" stage, where the training images are fed into a U-Net model for training, yielding a trained U-Net model (Subsection 3.4). Finally, this trained model is applied to the testing images, and various performance metrics are calculated (Subsection 3.5) to quantitatively evaluate the effectiveness of the proposed method and derive the final research results.

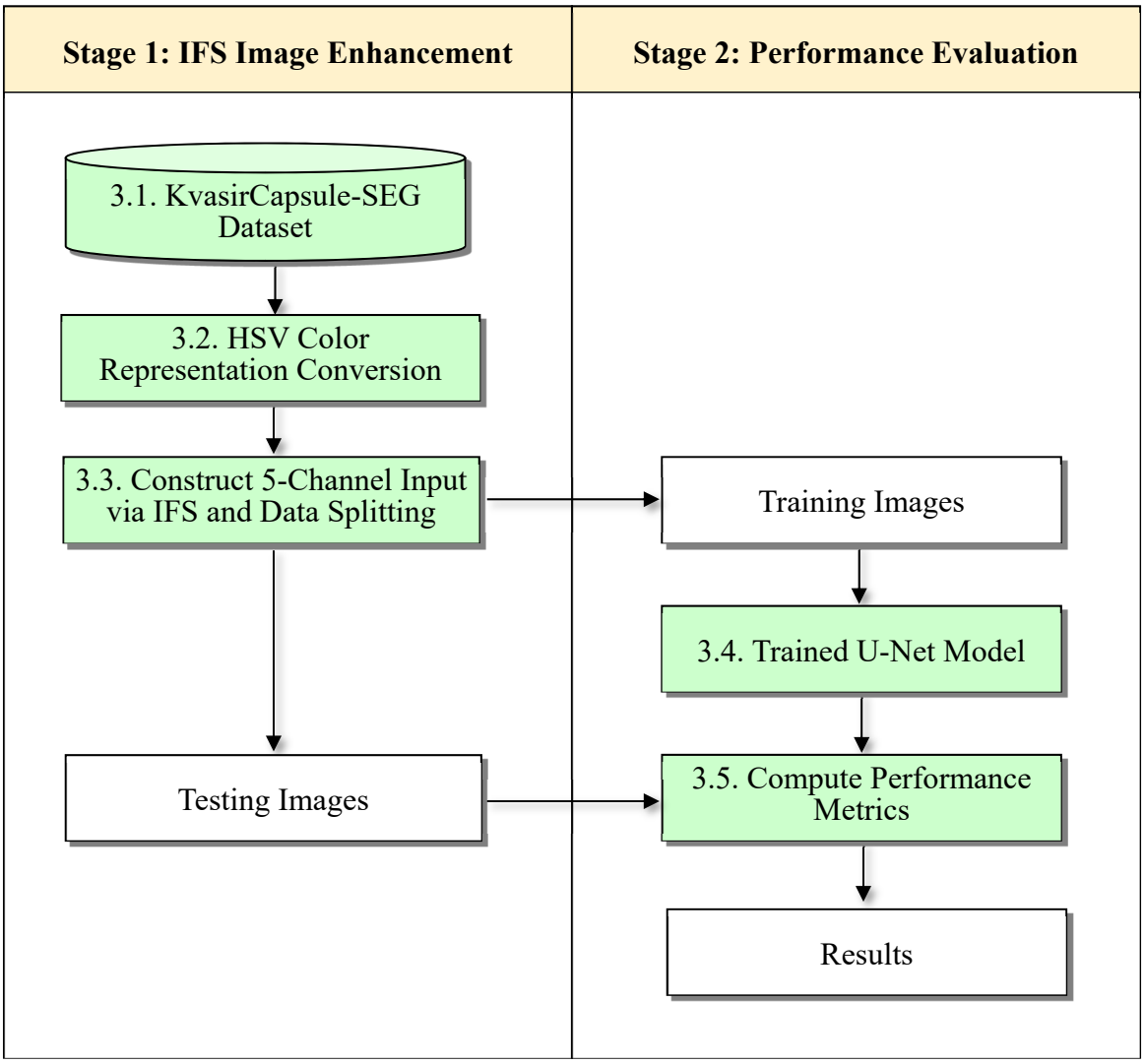


Figure 1: Workflow of the proposed IFS image enhancement and performance evaluation process for VCE image segmentation using the U-Net model.

3.1. KvasirCapsule-SEG Dataset

This study aims to address the problem of polyp semantic segmentation in VCE images. For this purpose, we primarily referenced and utilized the existing polyp class images from the Kvasir-Capsule dataset [11]. Kvasir-Capsule is a publicly available VCE image dataset, with its original version comprising 44,228 medically verified frames, covering 13 annotated anomalies and findings. Notably, within its rich content, only 55 frames explicitly contain polyps. To support accurate polyp segmentation tasks, we employed the KvasirCapsule-SEG dataset, which includes these 55 already annotated polyp images. KvasirCapsule-SEG provides pixel-level ground truth masks for these polyp images, with these masks having been completed by the Jha research team under the guidance of an experienced gastroenterologist, ensuring their medical accuracy [12]. This professionally annotated data, provided by a third-party team, forms the core foundation for the training and evaluation of our polyp semantic segmentation model in this study. Figure 2 presents examples of polyp images from the KvasirCapsule-SEG dataset and their corresponding ground truth masks.

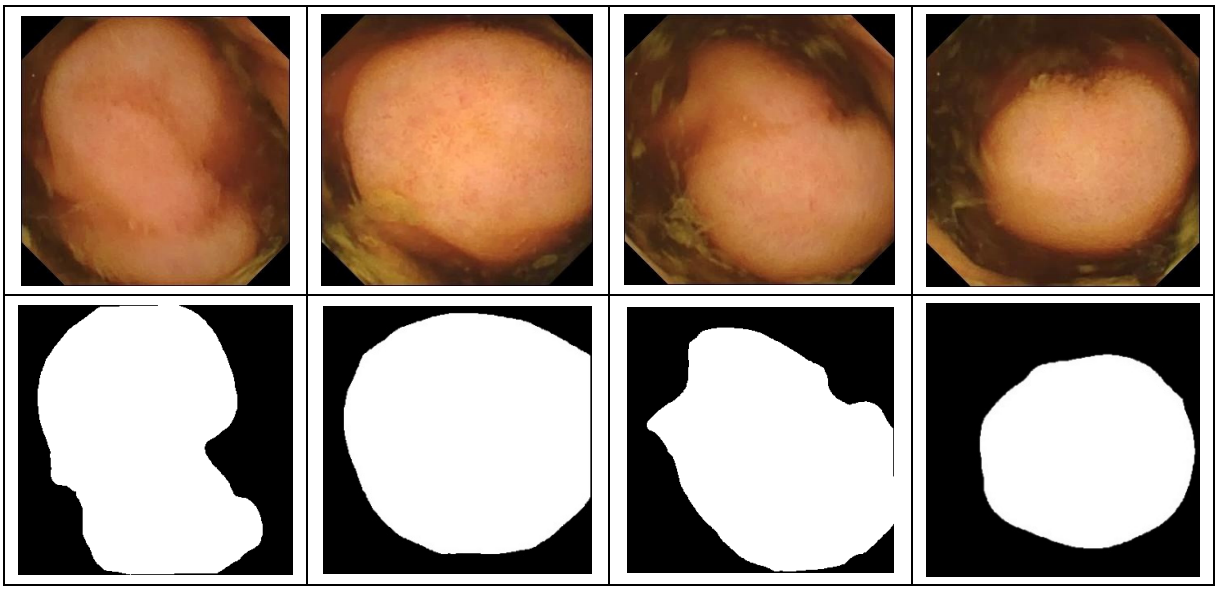


Figure 2: Examples from the "KvasirCapsule-SEG" dataset [12] used for VCE polyp segmentation in this study. The first row displays original RGB images, while the second row presents their precise ground truth masks, where white pixels explicitly define the polyp regions.

3.2. HSV Color Representation Conversion

To enhance polyp discriminability in VCE images, original RGB images are converted to HSV color space. This conversion, which separates chromatic (H, S) from brightness (V) and aligns with human perception [40], is crucial for mitigating adverse effects of variable illumination, reflections, and low contrast common in VCE. The Polyps exhibit distinct color attributes in HSV, making them more prominent. By stabilizing color representation, HSV facilitates robust feature extraction for subsequent IFS enhancement, effectively aiding membership degree assignment and ultimately

improving polyp segmentation accuracy.

3.3. Construct 5-Channel Input via IFS and Data Splitting

This stage represents the core innovation of this study, aiming to construct a 5-channel input for the U-Net model that quantifies image uncertainty. The specific steps are as follows:

1) **Feature Extraction**: For each HSV-converted image, we extract four core features for every pixel: Intensity, Hue, Saturation, and Texture. This Texture feature aims to capture the local spatial heterogeneity of the image. The calculation process is as follows: First, the color image is converted into a grayscale version. Subsequently, for each pixel in the grayscale image, we define a 7×7 local sliding window centered at that pixel. Within this window, the statistical standard deviation of all pixel intensity values is computed. This standard deviation value is then assigned to the central pixel as its texture feature. By repeating this operation for all pixels in the image, a complete texture feature map is ultimately obtained, as shown in Figure 3. This method quantifies local spatial intensity variations, effectively reflecting the surface roughness, smoothness, or uniformity of image regions.

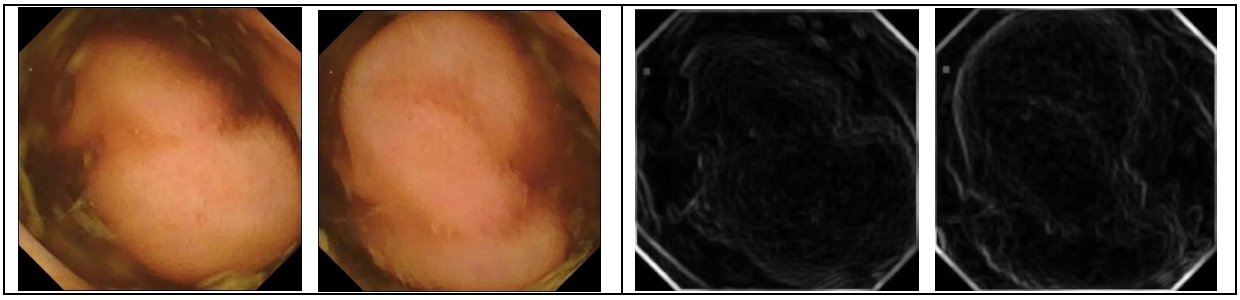


Figure 3: Visualization of texture features extracted from VCE polyp images. The two images in the left panel show the original VCE polyp images, and the two images in the right panel show their corresponding texture feature maps.

2) **Individual IFS Calculation**: For each pixel and each extracted feature k (Intensity, Hue, Saturation, Texture), we calculate its individual membership degree ($\mu_k(x)$) and non-membership degree ($v_k(x)$).

- For Intensity (I), Saturation (S), and Texture (Tex) features: We utilize Sigmoid functions to define their respective membership and non-membership degrees, describing the degree of feature value relevance to polyp attributes. The member function is defined as in Equation (1), and the non-member function is defined as in Equation (2):

$$\mu_k(x) = \frac{1}{1 + e^{-a_k(x - T_k)}} , \tag{1}$$

$$v_k(x) = \frac{1}{1 + e^{b_k(x-U_k)}} \ . \tag{2}$$

Where x represents the normalized value of the feature, and k denotes the feature type (I, S, Tex). Parameters a_k and b_k control the steepness of the Sigmoid curve, while T_k and U_k are the threshold center points. These parameters are adjusted based on the typical distribution ranges of polyps for each feature (e.g., polyps are generally brighter, more saturated, or have more complex textures) and empirical knowledge, such that $\mu_k(x)$ takes high values in regions characteristic of polyps, and $v_k(x)$ takes high values in regions characteristic of non-polyps.

- For the Hue (H) feature: As polyps typically concentrate within a specific reddish-to-pinkish hue range, we employ a Gaussian function to precisely capture its membership degree. The membership function is defined in Equation (3):

$$\mu_H = e^{-\frac{(x-\mu_{H_{center}})^2}{2\sigma^2_{H\_width}}} \ , \tag{3}$$

where x represents the normalized hue value (typically between 0 and 1), μ_{H_center} is the ideal central hue value for polyps in the HSV color space (e.g., the Hue value corresponding to the red-pinkish area), and σ_{H_width} controls the width of the acceptable range around this central hue. The non-membership degree for Hue, $v_H(x)$, is directly defined as $1-\mu_H(x)$.

- Validity Check and Processing: Although the μ and v for individual features are defined by the above functions, the sum of the final aggregated membership degree (μ_{final}) and final aggregated non-membership degree (v_{final}) (i.e., $\mu_{final} + v_{final}$) might exceed 1 after weighted aggregation of multiple features. To ensure the results comply with the validity rules of IFS ($\mu + v \leq 1$), we perform the following normalization on the variables μ_{final} and v_{final}. If $\mu_{final} + v_{final} > 1$, then the variables are rescaled as shown in Equation (4)-(5):

$$\mu_{final} \leftarrow \frac{\mu_{final}}{\mu_{final} + v_{final}} \ , \tag{4}$$

$$v_{final} \leftarrow \frac{v_{final}}{\mu_{final} + v_{final}} \ . \tag{5}$$

Based on the aforementioned Individual IFS Calculation, we visualize one sample from

the "KvasirCapsule-SEG" dataset, as shown in Figure 4.

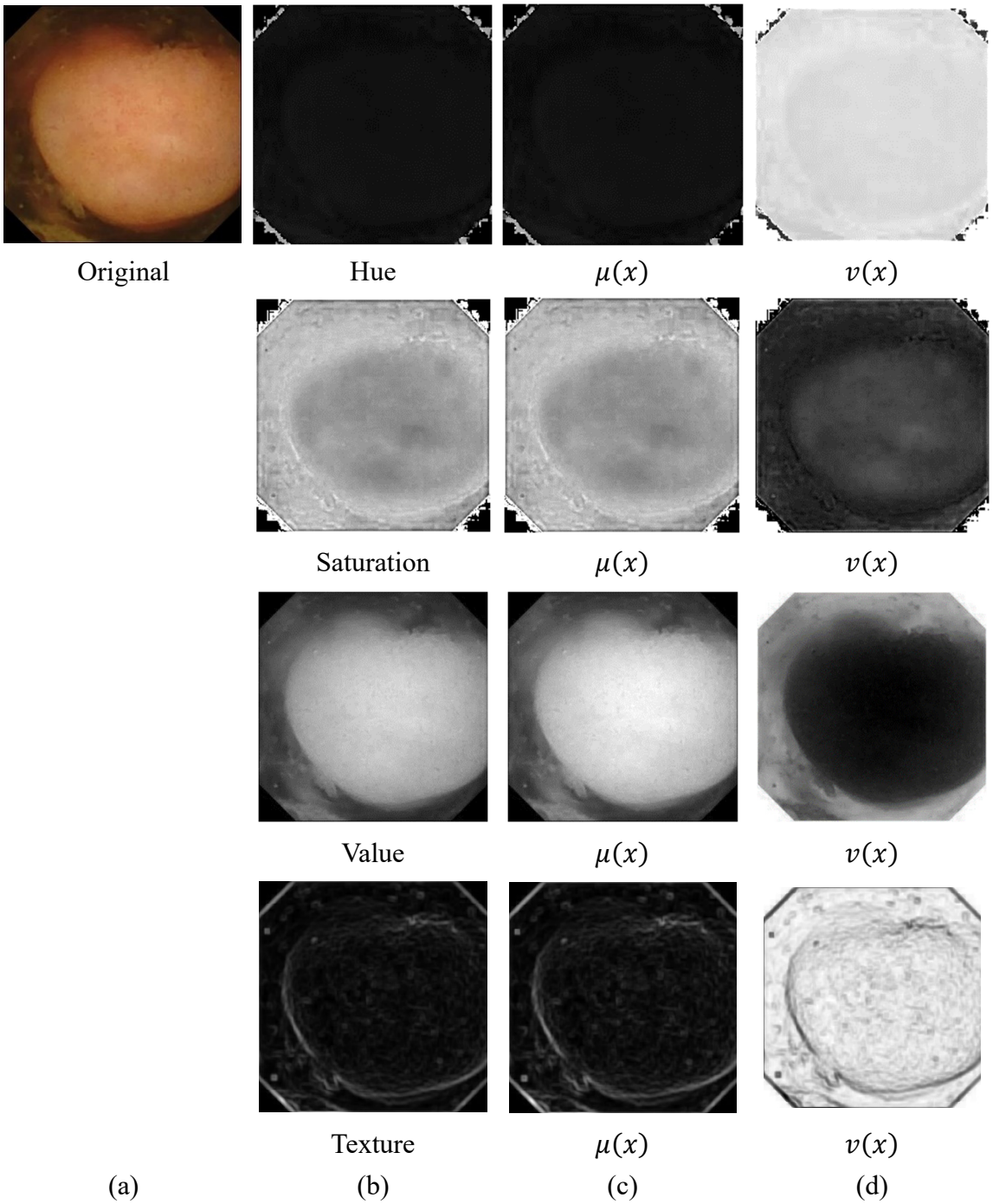


Figure 4: Visualization of the IFS image processing derived from a sample in the "KvasirCapsule-SEG" dataset [12]. This figure illustrates: (a) the original VCE polyp image, (b) its Hue, Saturation, Value (Intensity) channel, and Texture images, (c) the corresponding membership image (μ), (d) the corresponding non-membership image (v).

3) **Weighted Aggregation**: For each pixel, the individual membership degrees (μ_I, μ_H, μ_S, μ_{Tex}) and non-membership degrees (v_I, v_H, v_S, v_{Tex}) from the four features (Intensity, Hue, Saturation, Texture) are combined through a weighted aggregation process. This generates the final aggregated membership degree (μ_{final}) and final aggregated non-membership degree (v_{final}) for that pixel, as shown in Equation (6)-(7):

$$\mu_{final} = w_I \cdot \mu_I + \ w_H \cdot \mu_H + w_S \cdot \mu_S + w_{Tex} \cdot \mu_{Tex} \ , \quad (6)$$

$$v_{final} = w_I \cdot v_I + \ w_H \cdot v_H + w_S \cdot v_S + w_{Tex} \cdot v_{Tex} \ , \quad (7)$$

where w_I, w_H, w_S, w_{Tex} represent the weights for Intensity, Hue, Saturation, and Texture features, respectively. These weights reflect the relative importance of each feature in distinguishing polyp from non-polyp regions, and their sum is typically normalized to 1 (i.e., $w_I+w_H+w_S+w_{Tex}=1$). The setting of these weights can be based on empirical knowledge or determined through optimization methods. This study considers that all four features have their contributions, so the initial weight value is evenly divided, that is, $w_I=w_H=w_S=w_{Tex}=0.25$. Future research will explore optimizing weight distribution, exploring different research topics, and automated methods.

4) **5-Channel Input Tensor Construction**: This final preprocessing step aims to integrate the raw visual information with IFS-derived fuzzy uncertainty data into a unified tensor for the U-Net model. Specifically, the original 3-channel RGB image (which has been consistently resized to spatial dimensions of 336×336×3 during preprocessing) serves as the base. Subsequently, the previously generated single-channel μ_{final} map (336×336×1) and v_{final} map (336×336×1) are concatenated with the RGB image along their channel dimension. This operation ultimately yields a 5-channel input tensor with spatial dimensions of 336×336 and a depth of 5 (e.g., its dimension is 336×336×5). This composite tensor integrates the rich semantic information from the original image with the explicitly quantified uncertainty and belongingness information derived from the IFS, thereby providing a more comprehensive input representation for robust segmentation. The detailed construction of this 5-channel Input Tensor is illustrated as shown in Figure 5 and Algorithm 1.

Although hesitation degree (π) is an element of IFS, its information is implicitly contained within μ_{final} and v_{final}. To simplify the U-Net input and ensure data consistency across channels, we normalize μ_{final} and v_{final} so their sum is 1, thus not including π as an independent channel.

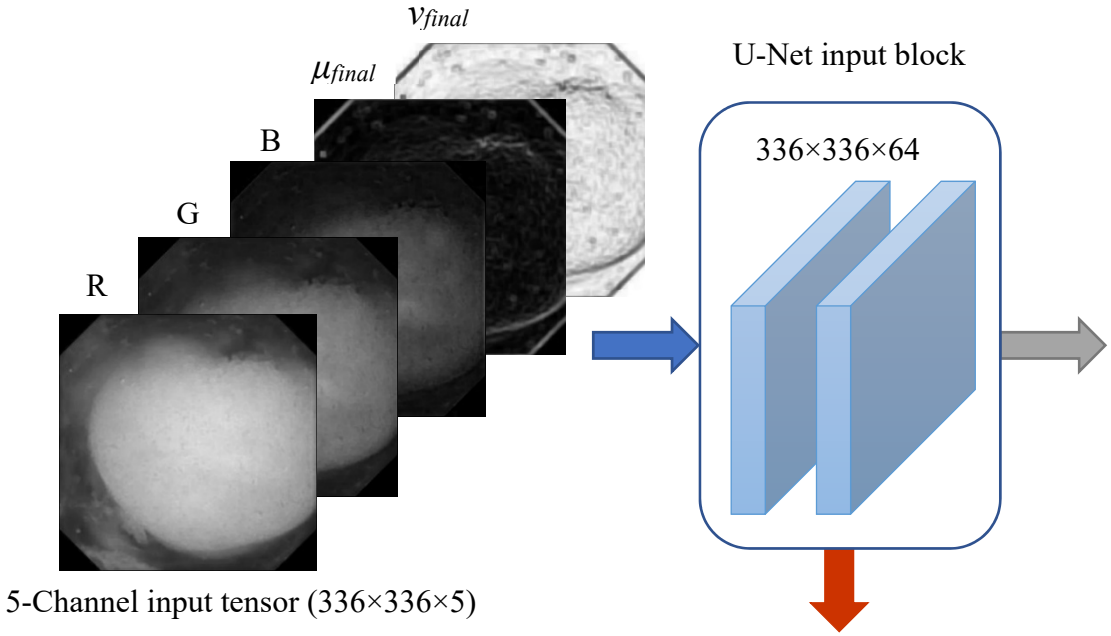


Figure 5: Illustration of the 5-channel input tensor construction for the U-Net model. The original RGB channels are concatenated with the IFS derived final aggregated membership degree (μfinal) and non-membership degree (νfinal) maps. This forms a 336×336×5 input, which is then fed into the U-Net's initial convolutional layers.

Algorithm 1: 5-Channel Input Tensor Construction

Input: RGB image $I_{RGB} \in \mathbb{R}^{H\times W\times 3}$

Output: 5-Channel $I_{5ch} \in \mathbb{R}^{H\times W\times 5}$

Step 1: Feature Extraction

Convert RGB to HSV and grayscale:

$I_{\text{HSV}} \leftarrow$ RGB_to_HSV(I_{RGB})

$I_{\text{Gray}} \leftarrow$ RGB_to_Grayscale(I_{RGB})

Extract feature maps:

$I_{\text{H}}, I_{\text{S}}, I_{\text{V}} \leftarrow$ split_channels(I_{HSV})

$I_{\text{Tex}} \leftarrow$ Compute_Texture(I_{Gray}) // e.g., local std

Step 2: Fuzzification and Weighted Aggregation

For each pixel $p \in (x, y) \in H\times W$:

Compute intuitionistic fuzzy values:

$u_H, v_H \leftarrow$ Fuzzify_Hue($I_H[p]$)

$u_S, v_S \leftarrow$ Fuzzify_Sat($I_S[p]$)

$u_V, v_V \leftarrow$ Fuzzify_Val($I_V[p]$)

$u_T, v_T \leftarrow$ Fuzzify_Tex($I_T[p]$)

Aggregate using weights w_H, w_S, w_V, w_T :

$u_{agg}(p) = w_H \cdot u_H + w_S \cdot u_H + w_V \cdot u_V + w_T \cdot u_T$

$v_{agg}(p) = w_H \cdot v_H + w_S \cdot v_H + w_V \cdot v_V + w_T \cdot v_T$

Step 3: Channel Concatenation

Merge channels:

$I_{5ch} \leftarrow$ merge_channels(R, G, B, $u_{agg}(p)$, $v_{agg}(p)$)

Return I_{5ch}

5) **Data Splitting**: To establish a standardized evaluation process and ensure a fair comparison of model performance, the constructed 5-channel data, along with their corresponding ground truth masks, are subsequently split into training and testing sets. This study adopts the holdout validation method for dataset splitting [41]. Specifically, for each image class, 80% of the data is randomly selected as the training set, while the remaining 20% is used as the testing set.

3.4. Trained U-Net Model

For the task of VCE polyp semantic segmentation, this study involved training two independent U-Net models. This dual-model approach was adopted to rigorously assess the impact of different input data on segmentation performance. Specifically, one U-Net model was trained using the original VCE images, while the second model was trained leveraging the 5-channel input tensors derived from the IFS preprocessing method.

To ensure a fair and consistent comparative analysis, both U-Net models shared the same network architecture and utilized identical training parameters, including the chosen loss function, learning rate, batch size, and total number of epochs. Training was performed within a supervised learning framework, where each model learned to map its respective input to the corresponding ground truth segmentation masks. Upon the successful completion of their training phases, the optimized model weights and associated training histories for both U-Net instances were separately saved. This systematic approach facilitates comprehensive performance evaluation and comparative analysis during the subsequent testing phase.

3.5. Compute Performance Metrics

Following the training of both U-Net models as described in Section 3.4, the final step involves a rigorous quantitative assessment of their segmentation performance. This evaluation is conducted on the unseen "Testing images" subset, along with their corresponding ground truth masks, which were strictly held out during the training phase. To comprehensively evaluate the models' efficacy in VCE polyp semantic segmentation and facilitate a direct comparison between the model trained on original images and the model trained on IFS-enhanced 5-channel inputs, a suite of widely

recognized performance metrics is computed. To establish a clear understanding of these metrics, we first define the components of the confusion matrix in the context of polyp segmentation: True Positives (TP) are pixels correctly identified as polyps; True Negatives (TN) are pixels correctly identified as non-polyps; False Positives (FP) are non-polyp pixels incorrectly identified as polyps; and False Negatives (FN) are polyp pixels incorrectly identified as non-polyps. These metrics include [42]:

1) **Accuracy**: Overall pixel-wise accuracy, representing the proportion of correctly classified pixels (both polyp and non-polyp) in the image. Its calculation formula is shown in Equation (8):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

2) **Precision**: Measures the proportion of true positive predictions among all positive predictions made by the model. Its calculation formula is shown in Equation (9):

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

3) **Recall / Sensitivity**: Measures the proportion of true positive predictions among all actual positive instances in the ground truth. Its calculation formula is shown in Equation (10):

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

4) **Dice Similarity Coefficient (DSC) / F1-score**: This metric quantifies the spatial overlap between the predicted segmentation mask and the ground truth mask, providing a robust measure of accuracy, especially for imbalanced segmentation tasks like polyp detection. Its calculation formula is shown in Equation (11):

$$\text{DSC} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{11}$$

5) **Intersection over Union (IoU) / Jaccard Index**: IoU measures the ratio of the intersection area to the union area of the predicted and ground truth segmentation masks. It is closely related to DSC and provides another perspective on segmentation quality. Its calculation formula is shown in Equation (12):

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{12}$$

The computed values for these metrics provide a quantitative basis for comparing the performance of the U-Net model trained with IFS-enhanced inputs against the model trained with original RGB inputs, thereby objectively assessing the contribution of the IFS preprocessing method to accurate polyp segmentation.

4. Experimental Results and Discussion

4.1. Experimental Environment

All experiments in this study were conducted on a single 64-bit Windows 11 personal computer, equipped with an Intel Core i9-13900 CPU (5.6 GHz) and 32 GB of RAM. Crucially, all computations, including data loading, preprocessing, U-Net model training, and performance evaluation, were performed in a CPU-only environment, specifically utilizing MATLAB R2024a with its Deep Learning and Image Processing Toolboxes. This setup was chosen to simulate typical user environments lacking high-performance computing resources, thereby ensuring the practical feasibility and robustness of our proposed IFS-enhanced methods under general computing conditions. This provides a realistic reference for future deployment in resource-limited settings.

4.2. Performance Comparison of Models Based on IFS Approach

This study aims to evaluate the contribution of the IFS method (5-Channel input tensor) in enhancing the segmentation performance of U-Net models. We compared models trained under two different image processing approaches: standard processing without the IFS method (Non-IFS) and processing integrated with the IFS method. Table 1 summarizes the performance of the models under these two approaches across key segmentation metrics.

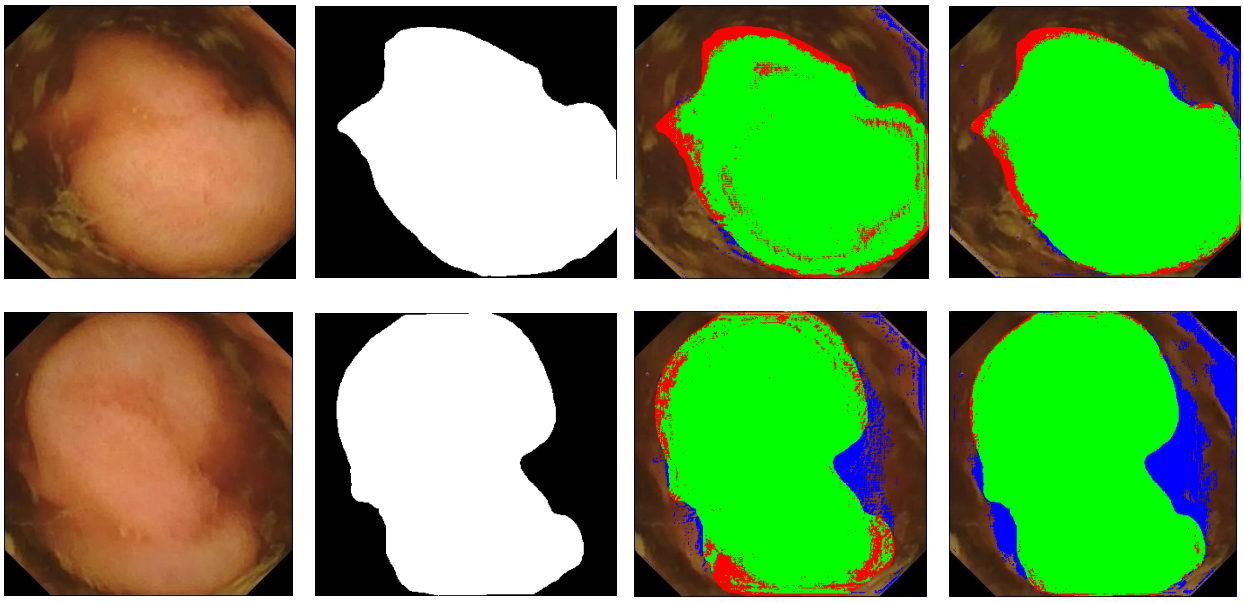
Table 1: Comparison of Segmentation Performance of Non-IFS and IFS in U-Net Models

Methods	Evaluation metrics (%)				
	Accuracy	Precision	Recall	DSC	IoU
Non-IFS	0.8064	**0.8346**	0.9179	0.8636	0.7648
IFS	**0.8261**	0.8113	**0.9304**	**0.8668**	**0.7771**

Based on the results in Table 1, although models trained using both methods demonstrated good segmentation capabilities, the model integrated with the IFS method exhibited superior performance over the traditional Non-IFS method in several key metrics. This is particularly evident in its improved ability to handle uncertainty and fuzziness. We present the following

analysis:

1) Accuracy: The IFS model (0.8261) achieved higher overall pixel accuracy than the Non-IFS model (0.8064). This demonstrates the effectiveness of the IFS method in improving the model's global classification ability, possibly by better handling pixel assignments in image boundaries or ambiguous regions.
2) IoU (Intersection over Union): The IFS model's IoU for the lesion class (0.7771) also surpassed that of the Non-IFS model (0.7648). IoU directly reflects the degree of overlap between the predicted segmentation and the ground truth lesion area. The higher IoU of IFS indicates that its segmentation boundaries and shapes are more accurately aligned with the true lesions, which may be attributed to IFS's ability to handle fuzzy boundaries. Figures 6 (c) and (d) visually demonstrate the performance of both methods in terms of segmentation result differences, where the IFS method (d) produces fewer red (false negatives) and blue (false positives) regions compared to the Non-IFS method (c), consistent with its higher value.



(a) (b) (c) (d)

Figure 6: Comparison of U-Net model segmentation results using Non-IFS and IFS methods. (a) Two test samples from the “KvasirCapsule-SEG” dataset. (b) The ground truth corresponding to samples (a). (c) The difference map of the semantic segmentation result using the Non-IFS method. Green regions indicate true positives (TP), red regions denote false negatives (FN), and blue regions represent false positives (FP). (d) The difference map of the segmentation result after applying the IFS method.

3) Precision: The Non-IFS model (0.8346) slightly outperformed the IFS model (0.8113) in terms of precision. Precision measures the proportion of pixels predicted as lesions that are actually lesions. This suggests that the Non-IFS model might be more "conservative" or "precise" when

predicting lesions, resulting in relatively fewer false positives (misclassifying non-lesions as lesions).

4) Recall: The IFS model (0.9304) significantly outperformed the Non-IFS model (0.9179) in recall. Recall is a crucial metric for evaluating a model's ability to avoid missed detections. The higher recall of the IFS model indicates its enhanced capability to comprehensively capture all actual lesion regions in an image, which is vital for applications like medical image diagnosis where missed diagnoses can have severe consequences.
5) DSC: The DSC of the IFS model (0.8668) was slightly higher than that of the Non-IFS model (0.8636). The DSC is the harmonic mean of precision and recall; the marginal advantage of IFS suggests it achieves a better balance between the two, providing more robust overall performance.

In summary, this study demonstrates that integrating the Intuitionistic Fuzzy Set (IFS) method significantly enhances the performance of U-Net models in lesion segmentation tasks compared to traditional Non-IFS methods. Although Non-IFS has a slight advantage in precision, the IFS model exhibits superior performance in overall accuracy, recall, F1-score, and most importantly, IoU for the lesion class. Particularly in medical image applications like lesion segmentation, recall is often more critical than precision, as missing (failing to detect) a true lesion can have more serious consequences than misclassifying a non-lesion area as a lesion. The higher recall of the IFS model strongly supports its value in practical applications. This highlights that the IFS method, through its capacity to handle uncertainty and fuzziness, provides the U-Net model with more robust discriminative power, making it more effective in identifying and precisely localizing lesions, thereby underscoring the potential value of IFS in this type of medical image analysis.

4.3. Analysis of Error Types and Visual Results

In the previous subsection, we compared the quantitative performance of the Non-IFS and IFS models. To gain a deeper understanding of the performance differences between the two models and their advantages and limitations in practical segmentation, this subsection will focus on analyzing the types of errors they produce (False Positives, FP, and False Negatives, FN), combined with visual results. As shown in Figure 6 (c) and (d), the difference maps intuitively present the model's error regions: red represents False Negatives (FN), meaning the parts of the true lesion that the model failed to detect; blue represents False Positives (FP), meaning regions that the model incorrectly identified as lesions but are actually non-lesions. Green regions indicate correctly identified True Positives (TP). Related analysis is as follows:

1) Analysis of False Negatives (FN): Observe the distribution and quantity of red regions in Figure 6 (c) and (d). The Non-IFS model (c) generally shows more red regions compared to the IFS model (d), especially at the edges or in complex shapes of the lesion. This corroborates the advantage of the IFS model in terms of Recall. By better handling boundary uncertainties, IFS effectively reduces instances where true lesions are misclassified as background, thereby

lowering the risk of missed diagnoses. This is particularly crucial in medical image analysis, as missing a lesion can lead to severe consequences.

2) Analysis of False Positives (FP): Observe the distribution and quantity of blue regions in Figure 6 (c) and (d). The blue regions of the Non-IFS model (c) may be fewer than or similar to those of the IFS model (d), or even slightly fewer, which is consistent with the slight advantage of the Non-IFS model in Precision. False positives typically occur in background regions with textures or colors similar to the lesion, or due to overfitting by the model. Although IFS generally performs better overall, its slightly lower precision indicates it might produce slightly more false alarms in some cases. However, in lesion segmentation, the priority is usually to avoid missed diagnoses, which makes the improvement in recall offered by IFS more valuable.
3) Analysis of Lesion Region Completeness (True Positives, TP): Both methods effectively identify the true lesion parts within the main body of the lesion (green regions). However, the green regions in the IFS model are typically more complete and closer to the true lesion's boundaries, further supporting its higher IoU value.

Through the visual analysis of error type difference maps, we can more clearly observe the superior ability of the IFS method in reducing false negatives, thereby improving the detection rate of lesions. Although Non-IFS shows a slight advantage in controlling false positives, the overall performance improvement brought by IFS in ensuring higher recall and IoU makes it a superior choice for lesion segmentation tasks. This also re-emphasizes the importance of ensuring all potential lesions are detected within the context of medical diagnosis.

5. Conclusion

This study aims to explore the enhancing effect of the Intuitionistic Fuzzy Set (IFS) method within the U-Net model for medical image segmentation performance, using VCE polyp image segmentation as a case study for in-depth evaluation. By comparing the performance of a traditional Non-IFS method with an IFS-integrated model, we successfully validated the effectiveness of IFS in addressing inherent challenges in medical image processing. Experimental results clearly demonstrate that, compared to the Non-IFS method, the IFS-integrated U-Net model exhibited significant superiority across multiple key evaluation metrics. Specifically, the IFS model achieved higher scores in Overall Accuracy, Recall, Dice Similarity Coefficient, and Intersection over Union (IoU) for the lesion class. Although the Non-IFS model showed a slight advantage in Precision, considering the clinical nature of VCE polyp detection, Recall is particularly crucial in such lesion segmentation tasks, as it directly relates to minimizing the risk of missed diagnoses. The IFS method, by leveraging its ability to handle image boundary fuzziness and uncertainty, significantly reduced false negatives (FNs), ensuring that more true polyps could be successfully identified by the model.

The main contribution of this research lies in proposing and empirically validating an input

feature enhancement strategy based on Intuitionistic Fuzzy Set theory, which, by constructing a five-channel input tensor, effectively overcomes challenges in VCE images and significantly improves the U-Net model's performance in accurate and robust polyp segmentation, particularly in achieving high recall. In summary, this study confirms the effectiveness of integrating the Intuitionistic Fuzzy Set method into the U-Net segmentation framework, providing a more robust and reliable strategy for medical image analysis in lesion detection and segmentation. These findings offer new directions for future research on precise segmentation of complex and ambiguous lesions. The suggested research directions include: integrating fuzzy logic more deeply into the model architecture, such as developing fuzzy convolution layers, to enhance its intrinsic robustness. Secondly, adaptive fuzzy logic should be explored, allowing the model to automatically learn optimal parameters to improve its generalization ability. Furthermore, this method can be extended to other medical imaging modalities, such as MRI or CT, to verify its versatility. Finally, combining with explainability tools like SHAP (SHapley Additive exPlanations) will help in understanding how fuzzy features influence decisions, thereby accelerating its application in clinical practice.

References

[1] N. B. Smith, A. Webb, Introduction to Medical Imaging: Physics, Engineering and Clinical Applications, Cambridge Texts in Biomedical Engineering, Cambridge University Press, Dec. 2010.

[2] P. Kumari, J. Chauhan, A. Bozorgpour, B. Huang, R. Azad, D. Merhof, “Continual learning in medical image analysis: A comprehensive review of recent advancements and future prospects,” Medical Image Analysis, Available online Jul. 2025.

[3] B. Raghuram, B. Hanumanthu, “Brain tumor image identification and classification on the internet of medical things using deep learning,” Measurement: Sensors, Vol. 30, Dec. 2023, 100905.

[4] Q. Xia, H. Zheng, H. Zou, D. Luo, H. Tang, L. Li, B. Jiang, “A comprehensive review of deep learning for medical image segmentation,” Neurocomputing, Vol. 613, Jan. 2025, 128740.

[5] H. Xiao, Z. Yang, T. Liu, S. Liu, X. Huang, J. Dai, “Deep learning for medical imaging super-resolution: A comprehensive review,” Neurocomputing, Vol. 630, May 2025, 129667.

[6] O. Ronneberger, P. Fischer, T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, pp. 234-241.

[7] J. Shao, S. Chen, J. Zhou, H. Zhu, Z. Wang, M. Brown, “Application of U-Net and optimized clustering in medical image segmentation: A review,” CMES-Computer Modeling in Engineering and Sciences, Vol. 136, Issue 3, Mar. 2023, pp. 2173-2219.

[8] P. Collins, “Video-capsule endoscopy – current and future role,” Medicine, Vol. 52, Issue 5,

May 2024, pp. 317-319.

[9] T. Singh, S. Jha, N. Bhatt, P. Handa, N. Goel, S. Indu, " Machine learning methodologies in video capsule endoscopy-based bleeding analysis: A systematic review of progress and prospects (2008–2024)," Engineering Applications of Artificial Intelligence, Vol. 159, Part B, Nov. 2025, 111659.

[10] K. T. Atanassov, "Intuitionistic fuzzy sets," Fuzzy Sets and Systems, Vol. 20, Issue 1, Aug. 1986, pp. 87-96.

[11] P. H. Smedsrud et al., "Kvasir-capsule, a video capsule endoscopy dataset," Springer Nature Scientific Data, 2021.

[12] D. Jha, N. K. Tomar, S. Ali, M. A. Riegler; H. D. Johansen, D. Johansen, "NanoNet: Real-time polyp segmentation in video capsule endoscopy and colonoscopy," 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Jun. 2021.

[13] P. M. de Azevedo-Marques, A. Mencattini, M. Salmeri, R. M. Rangayyan, Medical Image Analysis and Informatics: Computer-Aided Diagnosis and Therapy, CRC Press, Nov. 2017.

[14] C. L. Chowdhary, D.P. Acharjya, "Segmentation and feature extraction in medical imaging: A systematic review," Procedia Computer Science, Vol. 167, 2020, pp. 26-36.

[15] B. J Erickson, P. Korfiatis, Z. Akkus, T. L Kline, "Machine learning for medical imaging," National Center for Biotechnology Information, National Library of Medicine, Vol. 37, No. 2, Mar 2017, pp. 505-515.

[16] M. Aljabri, M. AlGhamdi, "A review on the use of deep learning for medical images segmentation," Neurocomputing, Vol. 506, Sep. 2022, pp. 311-335.

[17] P. Szczypiński, A. Klepaczko, M. Pazurek, P. Daniel, "Texture and color based image segmentation and pathology detection in capsule endoscopy videos," Computer Methods and Programs in Biomedicine, Vol. 113, Issue 1, Jan. 2014, pp. 396-411.

[18] F. Phillips, S. Beg, "Video capsule endoscopy: pushing the boundaries with software technology," Translational Gastroenterology and Hepatology, Vol. 6, Jan. 2021.

[19] S. Yuan, Q. Liu, Y. Liu, et al., "Characterization of video capsule endoscopy manipulation capabilities in low-viscosity liquids under a highly nonlinear magnetic field from a single magnet," 2024 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dec. 2024.

[20] C. Spada, G. Spera, M. Riccioni, A. Tringali, et al., "M2A patency capsule prior to video capsule endoscopy in patients with morphological small bowel strictures," Gastrointestinal Endoscopy, Vol. 59, Issue 5, Apr. 2004.

[21] J.-H. Oh, J. Muthukudage, W. Tavanapong, J. Wong, P. C. de Groen, S. J. Tang, "Abnormal image detection in endoscopy videos using a filter bank and local binary patterns," Neurocomputing, Vol. 144, 20 November 2014, pp. 70-91.

[22] P. Sullivan, S. Gupta, P. D. Powers, N. B. Marya, "Artificial intelligence research and development for application in video capsule endoscopy," Gastrointestinal Endoscopy

Clinics of North America, Vol. 31, Issue 2, Apr. 2021, pp. 387-397.

[23] L. Pinto, I. N. Figueiredo, P. N. Figueiredo," Reducing reading time and assessing disease in capsule endoscopy videos: A deep learning approach," International Journal of Medical Informatics, Vol. 195, Mar. 2025, 105792.

[24] A. Kanakatte, A. Ghose, "Precise bleeding and Red lesions localization from Capsule Endoscopy using Compact U-Net," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Nov. 2021.

[25] N. Li, N. Zhao, Z. Xiao, J. Lu, "Image segmentation of lesions in wireless capsule endoscopy based on Mask R-CNN," Applied Mathematics, Modeling and Computer Simulation, Feb. 2022.

[26] F. Ahamed, R. Islam, Nahiduzzaman, et al., "Automated colorectal polyps detection from endoscopic images using MultiResUNet framework with attention guided segmentation," Human-Centric Intelligent Systems, Vol. 4, 2024, pp. 299–315.

[27] K. Zhang, T. Shao, Y. Sun, et al., "Interpretable research of fuzzy methods: A literature survey," Information Fusion, Vol. 126, Part A, Feb. 2026, 103524.

[28] T. Chaira, "Chapter 9: Application of Fuzzy/Intuitionistic Fuzzy Set in Image Processing," Fuzzy Set and Its Extension: The Intuitionistic Fuzzy Set, John Wiley & Sons, Mar. 2019.

[29] Maheshkumar C.V., David Raj M., Saraswathi D., "On generalized Sugeno's class generator and parametrized intuitionistic fuzzy approach for enhancing low-light images," Applied Soft Computing, Vol.172, Mar. 2025, 112865.

[30] J. R. Jebadass, P. Balasubramaniam, "Color image enhancement technique based on interval-valued intuitionistic fuzzy set," Information Sciences, Vol. 653, Jan. 2024, 119811.

[31] C. Selvam, R. J. J. Jebadass, D. Sundaram, L. Shanmugam, "A novel intuitionistic fuzzy generator for low-contrast color image enhancement technique," Information Fusion, Vol. 108, Aug. 2024, 102365.

[32] H. Cui, Z. Xie, W. Zeng, R. Ma, et al., "Intuitionistic fuzzy local information C-means algorithm for image segmentation," Information Sciences, Vol. 681, Oct. 2024, 121205.

[33] Kumar S.N, Kannadhasan S, S. Zafar, H A. Kumar, "ROI extraction in corona virus (COVID 19) CT images using intuitionistic fuzzy edge detection," Heliyon, Vol. 10, Issue 6, Mar. 2024, e27798.

[34] T. Chaira, "Accurate segmentation of leukocyte in blood cell images using Atanassov's intuitionistic fuzzy and interval Type II fuzzy set theory," Micron, Vol. 61, Jun. 2014, pp. 1-8.

[35] V.P. Ananthi, P. Balasubramaniam, "A new image denoising method using interval-valued intuitionistic fuzzy sets for the removal of impulse noise," Signal Processing, Vol. 121, Apr. 2016, pp. 81-93.

[36] Z. Jia, J. Qiao, M. Chen, "Multi-level discrimination index for intuitionistic fuzzy coverings and its applications in feature selection," Expert Systems with Applications, Vol. 263, 5 Mar.

2025, 125735.

[37] A. Patel, S. Jana, J. Mahanta," Construction of similarity measure for intuitionistic fuzzy sets and its application in face recognition and software quality evaluation," Expert Systems with Applications, Vol. 237, Part B, 1 Mar. 2024.

[38] Y. Arora, S.K. Gupta, "Intuitionistic fuzzy twin proximal SVM with fuzzy hyperplane and its application in EEG signal classification," Applied Soft Computing, Vol. 163, Sep. 2024, 111816.

[39] P. Melo-Pinto, P. Couto, H. Bustince, E. Barrenechea, et al., "Image segmentation using Atanassov's intuitionistic fuzzy sets," Expert Systems with Applications, Vol. 40, Issue 1, Jan. 2013, pp. 15-26.

[40] R. C. Gonzalez, R. E. Woods, Digital Image Processing 4/e, Pearson FT Press, 2017.

[41] K. Pal, B. V. Patel, "Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Mar. 2020.

[42] M. E. Klontzas, A. Stanzione, A. Meddeb, A. Demircioğlu, et al., "Evaluation metrics in medical imaging AI: fundamentals, pitfalls, misapplications, and recommendations," European Journal of Radiology Artificial Intelligence, Vol. 3, Sep. 2025, 100030.