Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH

Open Access

Feature-enhanced fusion of U-NET-based improved brain tumor images segmentation

Abdul Haseeb Nizamani[1], Zhigang Chen[1*], Ahsan Ahmed Nizamani[1*] and Kashif Shaheed[2]

Abstract

The field of medical image segmentation, particularly in the context of brain tumor delineation, plays an instrumental role in aiding healthcare professionals with diagnosis and accurate lesion quantification. Recently, Convolutional Neural Networks (CNNs) have demonstrated substantial efficacy in a range of computer vision tasks. However, a notable limitation of CNNs lies in their inadequate capability to encapsulate global and distal semantic information effectively. In contrast, the advent of Transformers, which has established their prowess in natural language processing and computer vision, offers a promising alternative. This is primarily attributed to their self-attention mechanisms that facilitate comprehensive modeling of global information. This research delineates an innovative methodology to augment brain tumor segmentation by synergizing UNET architecture with Transformer technology (denoted as UT), and integrating advanced feature enhancement (FE) techniques, specifically Modified Histogram Equalization (MHE), Contrast Limited Adaptive Histogram Equalization (CLAHE), and Modified Bi-histogram Equalization Based on Optimization (MBOBHE). This integration fosters the development of highly efficient image segmentation algorithms, namely FE1-UT, FE2-UT, and FE3-UT. The methodology is predicated on three pivotal components. Initially, the study underscores the criticality of feature enhancement in the image preprocessing phase. Herein, techniques such as MHE, CLAHE, and MBOBHE are employed to substantially ameliorate the visibility of salient details within the medical images. Subsequently, the UT model is meticulously engineered to refine segmentation outcomes through a customized configuration within the UNET framework. The integration of Transformers within this model is instrumental in imparting contextual comprehension and capturing long-range data dependencies, culminating in more precise and context-sensitive segmentation. Empirical evaluation of the model on two extensively acknowledged public datasets yielded accuracy rates exceeding 99%.

Keywords Feature based segmentation, Transformers, UNET, Magnetic resonance imaging, Image enhancement filters

Introduction

Brain tumors, notably characterized by their uncontrolled growth within the brain, represent a significant health challenge due to their complex origins and the severe impact they have on patients' lives and well-being. Among these, gliomas are particularly significant, constituting about 35% of all brain tumors [1]. They originate from glial cells and are known for their invasive nature, ranging from low-grade benign forms to highly malignant types like glioblastoma. The early detection of these tumors is critically important because of their high malignancy and the typically short survival time for

*Correspondence:
Zhigang Chen
czg@csu.edu.cn
Ahsan Ahmed Nizamani
ahsan.official@csu.edu.cn
[1] School of Computer Science and Engineering, Central South University, Changsha 410083, China
[2] Department of Multimedia Systems, Faculty of Electronics, Telecommunication and Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland

affected patients, underscoring the urgent need for effective diagnostic procedures [2].

In the realm of modern diagnostics, several methods are employed, including ultrasound imaging, CT scans, X-ray, and notably Magnetic Resonance Imaging (MRI). MRI stands out for its non-invasive nature, the detailed insights it offers without exposing patients to harmful ionizing radiation, and its exceptional ability to differentiate soft tissues, such as tumors [3]. The varied imaging sequences available with MRI enable physicians to gain a comprehensive understanding of the tumor's characteristics, making it an indispensable tool in the diagnosis of brain tumors. The significance of medical imaging in modern medical diagnostics cannot be overstated, as these images are crucial in visualizing the internal structures of the human body [4, 5]. Medical image processing, which encompasses detection, segmentation, registration, and fusion, is essential in this context. Currently, the focus of medical image segmentation is on images of various human organs, tissues, and cells, segmenting them into regions based on similarities or differences. Enhancing these techniques, especially in MRI, is vital in advancing our ability to accurately diagnose and effectively treat brain tumors, ultimately improving patient outcomes [6].

Over the past few years, the field of medical image segmentation has witnessed a continuous stream of research endeavors, leading to the development and proposition of numerous techniques and methods. These approaches encompass a diverse array, encompassing threshold-based segmentation, region-based segmentation, and edge detection-based segmentation methods. Notably, traditional machine learning techniques, including decision trees, random forests, and clustering algorithms, have demonstrated their effectiveness in achieving precise image segmentation. Nevertheless, these methods are inherently reliant on feature engineering, and their performance is inherently constrained by the limited expressiveness of the features they extract [7, 8].

In recent years, deep learning methods, especially those based on convolutional neural networks (CNNs), have demonstrated strong feature recognition capabilities. They have generally outperformed traditional machine learning methods in areas like medical image segmentation. Consequently, deep learning-based medical image segmentation methods have garnered increasing attention and application [9, 10]. In this domain, medical image segmentation is a cornerstone, facilitating the differentiation of distinct regions within images, including discerning between healthy tissues and anomalies [11]. Deep learning techniques, such as the Fully Convolutional Network (FCN) [12], Deep lab [13], and notably the UNET architecture [14], have been pivotal in enhancing the precision of this vital task. Yet, while UNET has found tremendous success, it isn't devoid of limitations, primarily its rigidity in adapting to datasets of different sizes and potential inefficiencies in leveraging skip connections.

UNET, a powerful tool in medical image segmentation, has notable limitations that must be considered when applying it in healthcare settings. One significant drawback is its substantial data appetite. UNETs require large and diverse datasets for training, which can be difficult to obtain, particularly for rare conditions. Moreover, deep learning models like UNET are susceptible to overfitting, especially when the training dataset is limited. This means that while they may perform exceptionally well on training data, their ability to generalize to new, unseen cases can be compromised [15]. The computational demands of UNETs can also be a hindrance, as they necessitate robust hardware resources, both for training and inference, making them less accessible to smaller healthcare facilities. Another critical limitation is the model's interpretability, or rather, the lack thereof. UNETs are often regarded as "black boxes," making it challenging to explain how they arrive at their decisions, a crucial concern in healthcare where transparent decision-making is imperative. Additionally, UNETs may struggle with precise boundary delineation, potentially producing slightly irregular object boundaries. Variations in image quality, acquisition devices, and protocols can also pose challenges for the model's robustness. Addressing these limitations is paramount for the successful integration of UNETs into clinical practice. The major contributions of this study in the field of medical image segmentation are as follows:

- Innovative Hybrid Model (FE1-UT, FE2-UT, and FE3-UT): Our new algorithm combines the UNET architecture with Transformers and feature enhancement techniques (MHE, CLAHE, and MBOBHE). The resulting hybrid models, named FE1-UT, FE2-UT, and FE3-UT, represent a significant advancement in the field of medical image segmentation.
- Improved Performance: The primary focus of this research is to enhance the accuracy of results of segmentation for brain tumors. By integrating Transformers into the UNET framework, the models gain the ability to understand context and capture long-range dependencies within the data. This contextual understanding significantly improves segmentation accuracy, especially in cases involving intricate anatomical structures and indistinct features.
- Feature Enhancement in Image Preprocessing: The study emphasizes the importance of feature enhancement during the image preprocessing stage. The use

of image quality method such as MHE, CLAHE, and MBOBHE enhances the visibility of critical details within medical images, ensuring better results of segmentation.
- Exceptional Accuracy: The models developed in this study achieve remarkable accuracy rates, exceeding 99%, on two publicly available datasets. This level of accuracy is a significant achievement in medical image segmentation and reflects the excellence of the proposed approach.

The remaining sections of the paper are structured as follows:

- Section II provides an in-depth comparison of our novel methods with existing approaches.
- In Section III, we offer a concise overview of the structure of our innovative techniques.
- Section IV is dedicated to discussing the experimental results, including comprehensive discussions and comparisons with established methodologies.
- Concluding the paper, we present our final remarks and conclusions in Section VI.

Related work

This section is structured into two main categories: segmentation methods based on Convolutional Neural Network (CNN)-UNET approaches and segmentation methods using Transformer-based techniques. This division allows for a more in-depth exploration of the specific techniques and approaches within these two prominent branches of medical image segmentation.

Image segmentation with CNN and UNET

The journey continued with the groundbreaking concept of Convolutional Neural Networks (CNNs) introduced by LeCun et al. [16], and his collaborators. Their work achieved remarkable success in recognizing handwritten digits, notably with the construction of the LeNet-5 network. As computing power continued to advance, CNNs garnered widespread attention from researchers, gaining prominence in various domains. CNNs found their application in image segmentation, excelling not only in segmentation tasks but also in related areas such as image classification and object detection [17]. They have emerged as one of the most influential algorithms in the realm of deep learning. In the domain of medical image segmentation, CNN-based research predominantly falls into two categories:

Image Block Classification: In this approach, the task of image segmentation is transformed into the classification of local image blocks, where each pixel's location within the image plays a crucial role. For instance, researchers like Arkapravo Chattopadhyay and Mausumi Maitra have devised CNN-based models for brain tumor segmentation [18]. These models make extensive use of both local and global image features, enhancing their segmentation capabilities. The incorporation of fully connected layers at the end of the model significantly accelerates network training.

Semantic Segmentation based on Fully Convolutional Networks (FCN): This approach predicts the class to which each pixel within an input image belongs, enabling pixel-level semantic segmentation. Notably, Long and his team introduced the concept of Fully Convolutional Networks (FCN), capable of pixel-wise classification through forward propagation. This technology transforms image input into image output, enabling end-to-end segmentation [19]. FCN-based semantic segmentation has attracted substantial research efforts, with novel techniques emerging to facilitate hierarchical feature learning, classification optimization, and the creation of dense predictions for entire images.

Furthermore, advanced 3D networks, inspired by U-net-like topologies, have been introduced to extract contextual information from adjacent slices within 3D volumes used extensively in clinical practice [20]. Notable examples include 3D U-net and V-net, which leverage context from neighboring slices to enhance segmentation accuracy. In recent years, FCN-based semantic segmentation has dominated the landscape of medical image analysis. A significant proportion of international competitions, approximately 70%, focus on this particular area. Consequently, this chapter will delve into the exploration of fully convolutional neural networks, with a primary focus on the research status of the U-Net model in the domain of medical image segmentation.

Image segmentation with transformers

Although Convolutional Neural Networks (CNNs) have been around for many years, it wasn't until the introduction of AlexNet that CNNs became the mainstream deep learning model in the field of computer vision. Since then, deeper and more effective deep network models have gradually been proposed, such as ResNet, GoogleNet, DenseNet, and others [21]. In addition to exploring network architectures, these studies also included improvements to CNN itself, such as the introduction of dilated convolutions and depth-wise separable convolutions. One of the primary advantages of CNNs compared to traditional machine learning methods is that CNNs extract richer and more expressive features, eliminating the need for manual feature engineering. In different application scenarios, selecting suitable handcrafted features can be challenging,

while CNNs do not require manual feature selection and can perform end-to-end feature extraction.

However, one limitation of CNNs is their local operation, meaning that they have limited receptive fields [22]. To address the problem of limited convolutional receptive fields, commonly used operations like dilated convolutions effectively increase the receptive field without reducing resolution. Dilated convolution, uses convolution kernels with different dilation rates to extract features at different scales, to some extent alleviating the limitations of standard convolution operations. Feature pyramidal pooling, on the other hand, uses different sizes of pooling combinations to obtain multiscale feature information, enhancing classification accuracy. In the field of medical image segmentation, despite the success of models based on Convolutional Neural Networks (CNNs) like U-Net, there are still limitations in terms of segmentation accuracy and granularity due to the complexity of medical images, difficulty in data labeling, and limited annotated data.

Researchers have proposed various variations of the U-Net model to address these limitations. For example, U-Net++introduced mesh-like connectivity by using denser skip connections to link different stages of features [23]. R2U-Net [24] ensured segmentation continuity by introducing recurrent convolution modules and Long Short-Term Memory (LSTM) networks. SA-U-Net incorporated spatial attention modules to suppress irrelevant areas of feature maps, enhancing classifier discriminative accuracy, and used Dropout layers to mitigate overfitting [25]. However, these variations are primarily focused on improving convolutional models and do not fundamentally address the lack of global information in convolutional features. These improved variant networks still struggle to handle long-range semantic interactions in CNNs.

The Transformer was initially introduced in natural language processing research and was first applied to computer vision tasks, such as ImageNet image classification [26], through models like ViT (Vision Transformer) [27], achieving unprecedented success. Transformers divide images into fixed-size image patches, project them to a specified dimension through linear projection, and represent them as token sequences, offering a novel segmentation approach. Transformers model global information without downsampling, allowing for global information modeling while maintaining image resolution [28]. This approach is a fresh approach to semantic segmentation. Without relying on operations like dilated convolutions and Feature Pyramid Networks (FPN) used in convolutional methods, the Transformer expands receptive fields and obtains feature responses from a global perspective.

Transformers, based on multi-layer self-attention and multi-layer perceptrons, achieved significant success in natural language processing [29]. ViT was the first successful application of Transformers in computer vision, outperforming many advanced models in image recognition tasks. However, ViT is more suitable for large datasets. Touvron et al. [30], and others [31] improved ViT's performance on small datasets through various training strategies. The Swin-UNET model, utilizing a pure Transformer U-shaped network architecture, achieved excellent results in liver image segmentation [32]. Due to the high computational complexity of core self-attention computations in Transformers, Swin Transformer introduced the concept of sliding windows, reducing parameter counts for calculating self-attention within each window and enabling communication between non-adjacent patches.

While Transformer structures may perform relatively poorly on medical image datasets with limited data, some researchers have made progress in applying Transformers to image processing with promising results. The SETR model proposed using Transformers exclusively for semantic segmentation and introduced context information dependencies at every stage, removing the previous limitations of relying on dilated convolutions and attention mechanisms to increase receptive fields [33]. TransUNET was the first model to combine Transformer and CNN in a U-shaped lightweight network for abdominal organ segmentation. It used conventional CNNs to extract low-level information, serialized feature maps in the last stage of the Encoder using patches to obtain tokens, and then obtained global information through Transformers [34]. The TransFuse model employed a dual-branch structure with Swin-Transformer and CNN for feature encoding, capturing both local information and global dependencies. It introduced the Bifusion module to fuse multiscale features, achieving state-of-the-art results in Polyp dataset segmentation. DS-TransUNET [35] used two different patch sizes for partitioning and introduced a dual-branch Swin Transformer to extract different scale feature representations. It proposed the TIF fusion strategy to combine the results of two different scales. In the Decoder stage, it also introduced Swin-Transformer to establish global dependencies during upsampling. The Medical Transformer model used Gated axial attention and decomposed global spatial attention into two axial directions [36], significantly reducing parameter counts. It also introduced the Local branch and Global branch to fuse global and local segmentation results [37]. However, these methods have several limitations. For instance, while Transformers can establish global context dependencies, they may disrupt the shallow features of the

convolutional network, which contain crucial local information for improving edge segmentation accuracy. Therefore, designing a more suitable fusion model that retains low-level information while establishing long-term dependencies is a key challenge to address.

Method

The initial phase of the feature-enhanced UNET-based Transformer (FE-UT) model involves enhancing image features through a series of preprocessing steps. These steps utilize Contrast-Limited Adaptive Histogram Equalization (CLAHE) [38], Modified Histogram Equalization (MHE) [39], and Modified Brightness and Contrast Enhancement (MBOBHE) [40] techniques. These image enhancement methods are applied to enhance the contrast and visibility of the input image, ensuring that it is well-prepared for subsequent analysis and processing. In Fig. 1, the comprehensive implementation strategy for all algorithms is visually presented, outlining the various steps involved in this process.

Image enhancement

The preprocessing stage incorporates the application of MHE, CLAHE, and MBOBHE to leverage image enhancement techniques aimed at augmenting the contrast and visibility of the input image prior to any subsequent analysis or processing. Each of these methods possesses unique characteristics and brings specific advantages to the enhancement process. All models of enhancements are described as follows:

a) MBOBHE method

MBOBHE operates with the explicit goal of simultaneously addressing three critical aspects of image enhancement: contrast enhancement, brightness preservation and detail preservation.

Hum et al. [40] have conducted extensive research to demonstrate the superior performance of MBOBHE in comparison to existing bi-Histogram Equalization methods. Both quantitative and qualitative results substantiate the effectiveness of MBOBHE, highlighting its ability to provide a holistic view of image enhancement. Notably, MBOBHE excels in striking the delicate balance between preserving image brightness, retaining intricate details, and enhancing contrast in the final enhanced images Figs. 2 and 3.

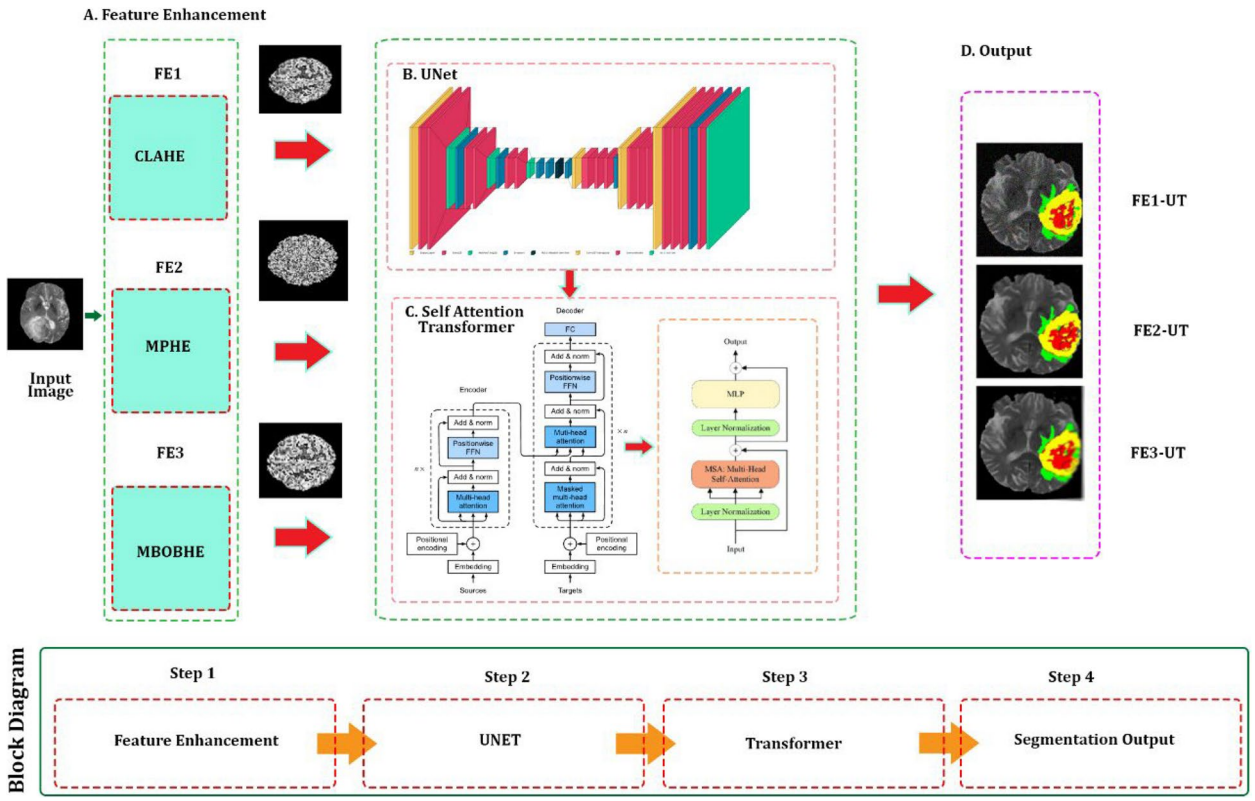

Fig. 1 Feature Enhanced Model for MRI segmentation using UNET and transformers

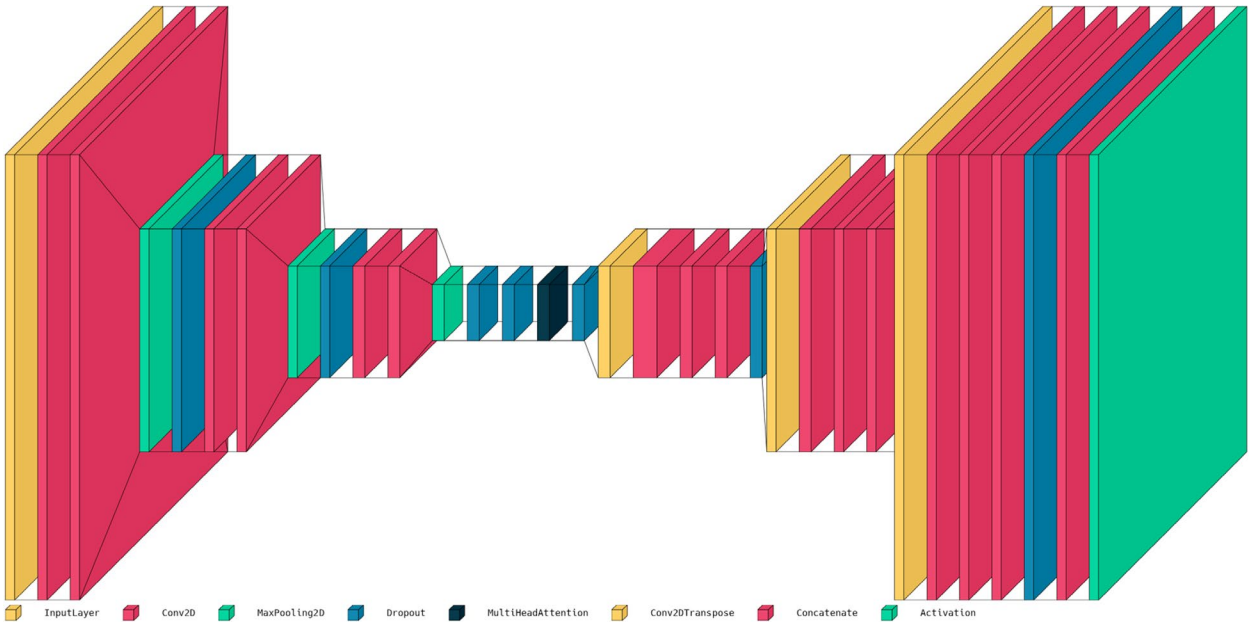

Fig. 2 Layered architecture of U-Net in this study

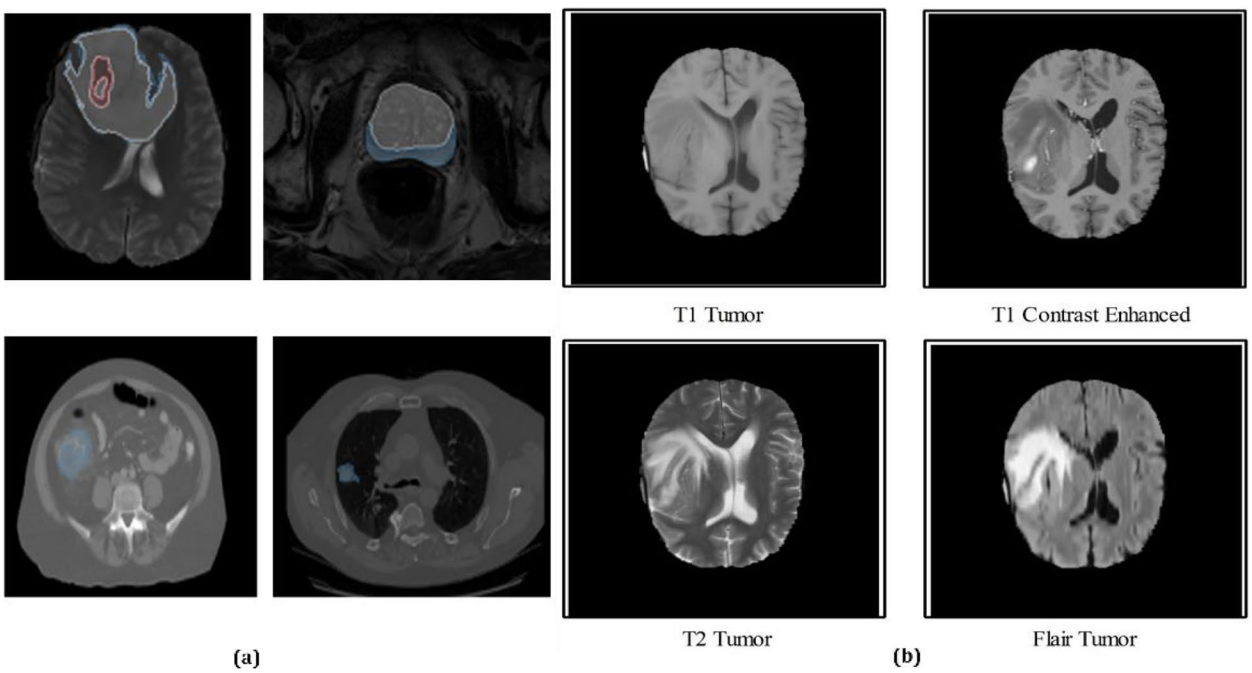

Fig. 3 Datasets used in this study (a) MSD dataset (b) BraTS dataset

b) Multipeak Histogram Equalization (MHE)

Multipeak Histogram Equalization, also known as Multi-Modal or Multi-Peak Histogram Equalization, is a variation of the traditional histogram equalization technique used in image processing and computer vision. Traditional histogram equalization aims to enhance the contrast of an image by redistributing the pixel intensity values in such a way that the resulting histogram is approximately uniform. However, this approach may not be suitable for images with multiple distinct intensity peaks or modes, as it can cause

unwanted artifacts and exaggerate the differences between the modes. Multipeak Histogram Equalization is designed to handle images with multiple intensity modes effectively. It does so by identifying the distinct peaks or modes in the histogram and then equalizing each mode separately. Here's a outline of the process:

- Compute the Histogram

  Let H be the histogram of the input image I, where H(i) represents the number of pixels with intensity i.
- Identify Intensity Modes:
  Detect the peaks or modes in the histogram.
  Divide the Image:
  Divide the input image I into subregions based on the identified modes.
- Apply Histogram Equalization:
  For each subregion, perform histogram equalization independently. Let's denote the subregions as I_1, I_2, ..., I_n, where n is the number of modes. Apply histogram equalization to each subregion as follows:

1. Compute the cumulative distribution function (CDF) for the subregion:

$$\text{CDF\_i(j)} = \frac{\text{sum (H(i) for all i from 0 to j)}}{\text{Total number of pixels in I\_i}} \quad (1)$$

2. Apply histogram equalization to the subregion:

$$\text{I\_i(j)} = \text{round (CDF\_i(I\_i(j))} * \text{(Number of intensity levels} - 1)) \quad (2)$$

3. Reconstruct the Image:

After equalizing all the sub regions, combine them to reconstruct the final equalized image.

Multipeak Histogram Equalization can be particularly useful for enhancing the contrast in images where different objects or regions have varying illumination conditions or intensity characteristics. By equalizing each mode separately, it preserves the relative differences between modes while enhancing the contrast within each mode.

iii) Contrast-limited adaptive histogram equalization (CLAHE)

Contrast-Limited Adaptive Histogram Equalization (CLAHE) is a widely used technique in image processing to enhance the contrast of an image while limiting the amplification of noise in flat or low-contrast regions. CLAHE is particularly useful when dealing with images that have uneven lighting conditions or regions with varying contrasts. The basic idea behind CLAHE is to divide the image into small tiles or blocks and perform histogram equalization within each tile. However, to prevent excessive amplification of noise, CLAHE also limits the contrast enhancement for each tile by clipping the histogram.

Here's an explanation of CLAHE along with equations:

Divide the image into tiles

Divide the input image I into non-overlapping tiles or blocks. Let's denote these tiles as I(x, y), where (x, y) represents the coordinates of the top-left corner of each tile.

Calculate the histogram for each tile

For each tile I(x, y), compute the histogram H(x, y) that represents the distribution of pixel intensities within that tile.

Clip the histogram

Apply contrast limiting by clipping the histogram. This is done by setting a predefined threshold T. If any bin in the histogram exceeds this threshold, the excess pixels are redistributed to other bins. The formula for this clipping is as follows:

$$\text{H\_clip}(x, y, i) = \min(\text{H}(x, y, i), \text{T}), \text{ for all i} \quad (3)$$

After clipping, normalize the histogram so that its sum remains the same

$$\text{H\_norm}(x, y, i) = \text{H\_clip}(x, y, i)/\text{sum}(\text{H\_clip}(x, y, j) \text{ for all j}) \quad (4)$$

Calculate the Cumulative Distribution Function (CDF)

Compute the cumulative distribution function (CDF) for each clipped and normalized histogram:

$$\text{CDF}(x, y, i) = \text{sum}(\text{H\_norm}(x, y, j) \text{ for all j from 0 to i}) \quad (5)$$

Apply histogram equalization

For each tile, map the pixel intensities using the CDF to perform histogram equalization:

$$I\_equalized(x, y, p) = round(CDF(x, y, I(x, y, p)) * (Number\ of\ intensity\ levels - 1)) \quad (6)$$

Reconstruct the image

Combine the equalized tiles to form the final CLAHE-enhanced image.

The key parameter in CLAHE is the contrast threshold (T). Adjusting this threshold will control the degree of contrast enhancement and noise amplification. A lower value of T results in stronger contrast enhancement but may increase noise, while a higher value of T reduces contrast enhancement but also limits noise amplification.

CLAHE is a powerful technique for enhancing local contrast in images and is commonly used in medical image processing and other applications where contrast is crucial. Its adaptability to local image content makes it a valuable tool for various image enhancement tasks.

Improved U-net segmentation

The U-Net architecture, renowned for its exceptional performance in medical science and bioinformatics image segmentation tasks, has garnered significant attention among researchers [41]. Its name is derived from the network's structural shape, which bears a resemblance to the letter "U." This architecture encompasses two fundamental paths:

Self-attention-transformer

Recent advancements have seen the incorporation of Transformers, which excel in capturing long-range dependencies and contextual information [42]. The Transformer blocks can be inserted at various points in the U-Net architecture to enhance feature extraction and segmentation performance. By attending to and aggregating information across the feature maps, Transformers contribute to a deeper understanding of image context, allowing for more context-aware and accurate segmentation.

Algorithm

This code implements a convolutional neural network (CNN) based on the U-Net architecture with an additional Transformer module for image segmentation tasks. Below is a detailed explanation of the code in points:

- Input Shape and Layers Initialization

The input shape of the images is defined as (240, 240, 4), indicating images with a resolution of 240x240 pixels and 4 channels.

The code initializes an input layer (inply) using the defined input shape.

- Encoder

Convolutional Layers: The input passes through a series of convolutional layers (conv1, conv2, conv3) with increasing filters (64, 128, 256) and 3x3 kernel size, followed by ReLU activation and same padding. This extracts essential features from the input image.

MaxPooling and Dropout: After each set of convolutional layers, max-pooling is applied to reduce spatial dimensions, and dropout is used for regularization to prevent overfitting.

- Transformer Module

Dropout: A dropout layer with a dropout rate of 0.1 is added to the output of the encoder (drop3).

Multi-Head Attention: The dropout output is fed into a Multi-Head Attention layer with 4 heads and a key dimension of 64. This layer captures complex patterns and long-range dependencies in the input features.

- Decoder

Convolutional Transpose Layers: The output from the Transformer module is passed through a series of transpose convolutional layers (tran1, tran2, tran3). These layers upsample the features to reconstruct the spatial dimensions of the image.

Concatenation: At each stage of the decoder, the upsampled features are concatenated with the corresponding features from the encoder to provide skip connections. This helps the network to retain detailed information from the encoder.

Convolutional Layers and Dropout: After concatenation, the features go through several additional convolutional layers (conv4, conv5, conv6) with ReLU activation and same padding. Dropout is applied after each convolutional layer for regularization.

- Output Layer

Convolutional Layer with Softmax Activation: The output from the decoder is passed through a 1×1 convolutional layer with 4 filters (for 4 segmentation

classes) and same padding. Softmax activation is applied to obtain the final segmentation probabilities for each class.

Experimental setting and results

Results with experimental settings are described in this section.

Evaluation metrics

This study employed a range of evaluation metrics to assess the performance of the model and the equations for these evaluation methods are as follows:

$$\text{Dice} = \frac{2 * \text{Tp}}{2 * \text{Tp} + \text{Fp} + \text{Fn}} \tag{7}$$

$$\text{Accuracy} = \frac{Tp + Tn}{Tp + Tn + Fn + Fp} \tag{8}$$

$$\text{Sensitivity} = \frac{Tp}{Tp + Fn} \tag{9}$$

$$\text{Specificity} = \frac{\text{Tn}}{\text{Tn} + \text{Fp}} \tag{10}$$

$$\text{Precision} = \frac{Tp}{Tp + Fp} \tag{11}$$

$$\text{Inter section over Union (IOU)} = \left(\frac{Tp}{Fp + Tp + Fn}\right) \tag{12}$$

$$\text{Kappa} = \frac{(\text{Tn} + \text{Fn})(\text{Tn} + \text{Fp}) + (\text{Fp} + \text{Tp})(\text{Fn} + \text{Tp})}{\text{Tp} + \text{Tn} + \text{Fn} + \text{Fp}} \tag{13}$$

$$\textit{Balanced Accuracy}(\text{BA}) = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{14}$$

Where;

FN (False Negative): This refers to cases where the model or classifier incorrectly predicted the negative class when the true class was actually positive. In other words, it's a situation where a positive instance is missed or not detected.

TP (True Positive): This indicates cases where the model correctly predicted the positive class when the true class was indeed positive. It represents accurate positive predictions.

FP (False Positive): FP refers to cases where the model incorrectly predicted the positive class when the true class was actually negative. In this situation, the model made a positive prediction when it should not have.

Dataset description

Datasets used in this study are BraTS dataset [43] and Medical Segmentation Decathlon (MSD) [44]. The Medical Segmentation Decathlon (MSD) dataset is a comprehensive collection of medical images and corresponding segmentation masks, designed for evaluating and developing medical image segmentation algorithms. It covers various imaging modalities and anatomical regions, making it versatile for different medical tasks. Researchers use it to benchmark and compare the accuracy of segmentation algorithms, making it a valuable resource in medical image analysis research for applications like organ segmentation, disease diagnosis, and treatment planning.

The BRATS (BraTS) 2020 dataset is a widely recognized collection of medical images designed for the evaluation and development of algorithms related to brain tumor segmentation and diagnosis. It contains multi-modal magnetic resonance imaging (MRI) scans, including T1-weighted, T1-weighted contrast-enhanced, T2-weighted, and FLAIR (Fluid Attenuated Inversion Recovery) images. The dataset provides annotations for brain tumor regions, including gliomas, making it invaluable for machine learning and deep learning research in the field of medical image analysis. Researchers and practitioners use the BRATS 2020 dataset to develop and benchmark segmentation and classification algorithms for brain tumor detection and analysis, contributing to advancements in neuro-oncology and medical imaging.

Experimental parameters setting

The hardware environment for this experiment includes a CPU with a clock speed of 3.40 GHz, an NVIDIA GTX 1070 GPU, and 16 GB of memory. This study models were constructed using TensorFlow and Keras as the backend framework. To optimize performance, the Adam optimizer, known for its robustness, was chosen. Furthermore, the preprocessing steps includes normalization to avoid complexities in preprocessing. For the dataset, we adopted a random split, allocating approximately 80% of the data to the training set and reserving the remaining 20% for the test set. Specific parameter configurations for all the algorithms employed in this research are provided in detail in Table 1.

Proposed methods segmentation performance evaluation

Our study three algorithms are compared by using different validation criteria. Balanced accuracy takes into

Table 1 Layer and parameter settings

Part	Layer Type	Layer Name	Output Shape	Number of Parameters
Encoder	Conv2D	conv2d	(None, 240, 240, 64)	2368
Encoder	Conv2D	conv2d_1	(None, 240, 240, 64)	36928
Encoder	MaxPooling2D	max_pooling2d	(None, 120, 120, 64)	0
Encoder	Dropout	dropout_1	(None, 120, 120, 64)	0
Encoder	Conv2D	conv2d_2	(None, 120, 120, 128)	73856
Encoder	Conv2D	conv2d_3	(None, 120, 120, 128)	147584
Encoder	MaxPooling2D	max_pooling2d_1	(None, 60, 60, 128)	0
Encoder	Dropout	dropout_2	(None, 60, 60, 128)	0
Encoder	Conv2D	conv2d_4	(None, 60, 60, 256)	295168
Encoder	Conv2D	conv2d_5	(None, 60, 60, 256)	590080
Transformer	Dropout	dropout_3	(None, 30, 30, 256)	0
Transformer	MultiHeadAttention	multi_head_attention	(None, 30, 30, 256)	263168
Transformer	Dropout	dropout_4	(None, 30, 30, 256)	0
Decoder	Conv2DTranspose	conv2d_transpose	(None, 60, 60, 256)	262400
Decoder	Concatenate	concatenate	(None, 60, 60, 512)	0
Decoder	Conv2D	conv2d_6	(None, 60, 60, 256)	1179904
Decoder	Conv2D	conv2d_7	(None, 60, 60, 256)	590080
Decoder	Conv2DTranspose	conv2d_transpose_1	(None, 120, 120, 128)	131200
Decoder	Concatenate	concatenate_1	(None, 120, 120, 256)	0
Decoder	Conv2D	conv2d_8	(None, 120, 120, 128)	295040
Decoder	Conv2D	conv2d_9	(None, 120, 120, 128)	147584
Decoder	Conv2DTranspose	conv2d_transpose_2	(None, 240, 240, 64)	32832
Decoder	Concatenate	concatenate_2	(None, 240, 240, 128)	0
Decoder	Conv2D	conv2d_10	(None, 240, 240, 64)	73792
Decoder	Conv2D	conv2d 11	(None, 240, 240, 64)	36928
Output	Conv2D	conv2d_12	(None, 240, 240, 4)	260
Output	Activation	activation	(None, 240, 240, 4)	0

account both the sensitivity and specificity of the segmentation results, making it a more reliable measure in situations where the class distribution is imbalanced, as often seen in medical imaging. This metric provides a balanced evaluation of how well the algorithm identifies both positive and negative regions within the image, ensuring that neither class is disproportionately favored in the assessment, which is crucial for accurate and fair evaluation of medical image segmentation models. For Dataset 1, FE1-UT is having higher balanced 98.64% which is higher then FE2-UT (98.53%) and FE3-UT(98.42%). Similarly, Precision is 98.19% for FE1-UT which is higher then FE2-UT (97.98%) and FE3-UT(97.85%). Higher precision of FE1-UT means that a greater proportion of the pixels or regions identified as belonging to a particular class (e.g., a specific anatomical structure or lesion) are indeed correct or true positives. Therefore, all algorithms higher precision shows that the segmentation algorithm produces fewer false positives and is better at correctly identifying the regions of interest in the medical image. This is particularly important in medical applications, where misclassifying or missing important structures can have serious clinical consequences. Another important metric is Recall, which is also 98.18% for FE1-UT while 97.90% for FE2-UT and 97.85% for FE3-UT. Higher recall means that the segmentation algorithm has correctly identified a greater proportion of the actual positive regions (e.g., important anatomical structures or abnormalities) within the image. Similar higher results are observed for FE1-UT in other dataset 2 as well in all metrics, which shows that CLAHE improvement has a better impact on image segmentation. Figures 4 and 5 shows the comparative performance of the proposed algorithms against different image segmentation metrics.

CLAHE enhances local contrast, which helps in better delineation of subtle details and boundaries in medical images. U-Net provides excellent spatial feature extraction and segmentation capabilities. The Transformer

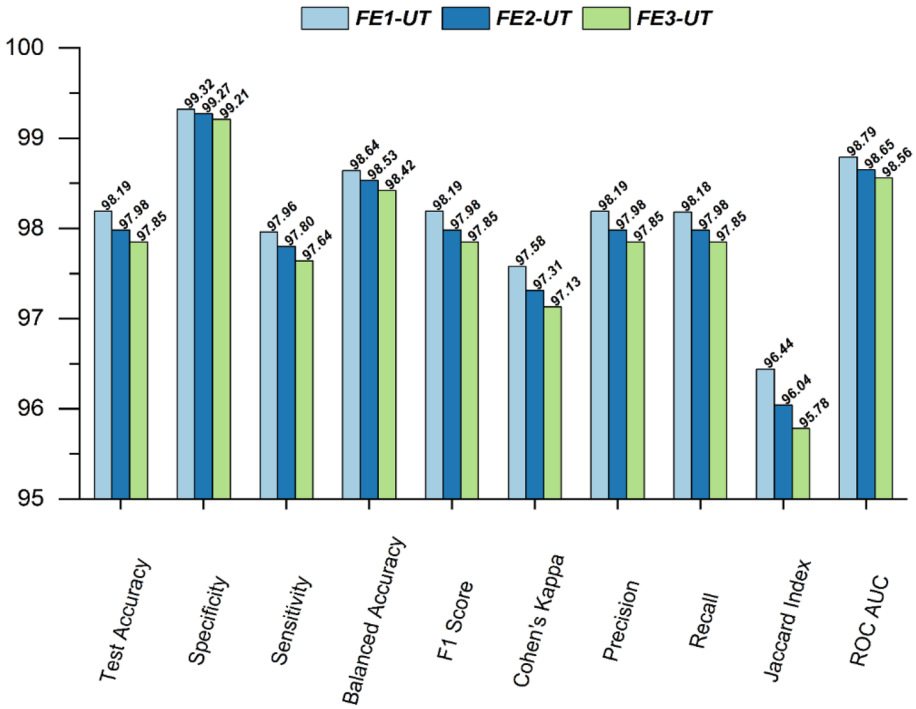

Fig. 4 Performance of proposed algorithms in MSD dataset

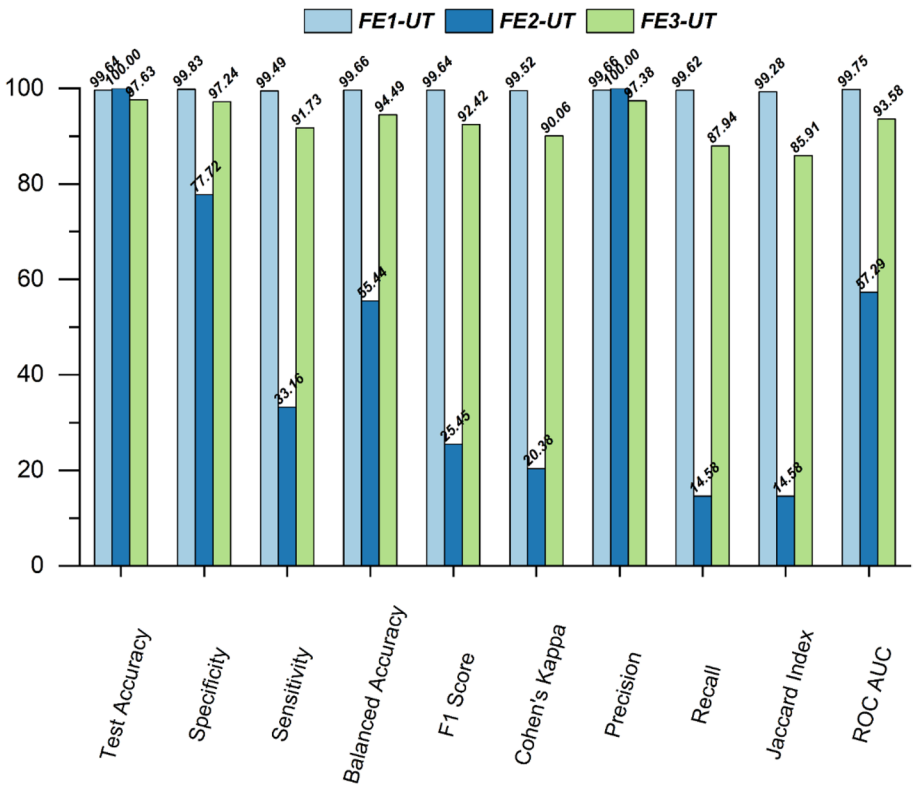

Fig. 5 Performance of proposed algorithms in BRaTS dataset

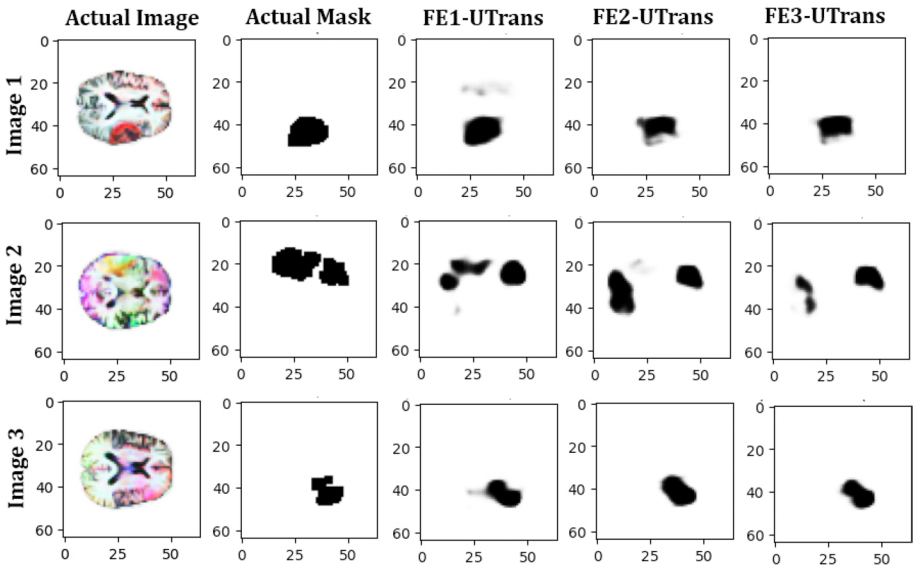

Fig. 6 Comparison of proposed methods visual in database 1

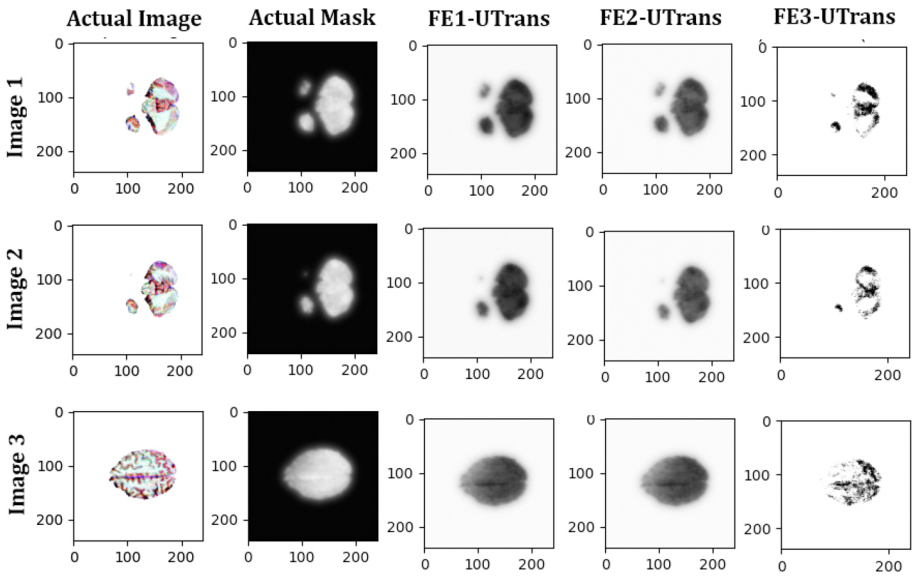

Fig. 7 Comparison of proposed methods visual in database 2

component can capture long-range dependencies, making it effective for tasks where context matters. The synergistic use of these components can improve segmentation accuracy, especially in complex and high-contrast medical images, enabling the model to handle a wider range of clinical scenarios and deliver superior results. Figures 6 and 7 provide a visual comparison of our proposed model with each dataset. It is evident that our model exhibits remarkable consistency with the ground truth in terms of feature extraction, closely resembling other existing methods. However, what sets our proposed hybrid model apart is its superior segmentation performance, with more distinct and clearly visible boundaries.

Traditional methods comparison with proposed methods

In medical image segmentation, like when using a U-Net architecture, sensitivity is an important evaluation metric because it measures the ability of the model to correctly identify positive instances (i.e., true positives) within the dataset. Sensitivity is also known as the true positive rate, recall, or hit rate, and it quantifies the model's ability to detect all instances of a particular class, typically

Table 2 Sensitivity comparison of different algorithms with proposed methods

Method	MSD	BRATS
UNET	45.21	90.19
Dense UNET	46.27	91.46
Att -UNET	50.08	85.24
UNET + +	49.0	89.06
UNET3 +	44.9	89.30
Trans-UNET	47.22	92.34
TransU2-UNET	47.45	93.85
UNET + CLAHE	97.12	98.36
UNET + MBOBHE	97.42	**99.50**
UNET + MPHE	97.63	77.72
FE1-UT	**97.96**	<u>99.49</u>
FE2-UT	97.8	91.73
FE3-UT	<u>97.64</u>	63.16

the presence of a specific medical condition or object of interest within an image. To validate the results of our proposed method against different state of the arts (SOTA) methods we used to compare the sensitivity of our proposed method with other algorithms. Table 2 shows the results of different methods with best results are bold while second best results are underline.

The loss function and accuracy graph with epoch is a vital visualization tool for monitoring the training progress and evaluating the performance of deep learning models in medical image segmentation. The loss function graph shows how the loss (e.g., Dice loss or cross-entropy loss) decreases as training progresses. A decreasing loss indicates that the model is converging and learning to produce better segmentations. In contrast, the accuracy graph measures the similarity between predicted and ground truth segmentations, which is crucial for assessing the model's performance. An increasing accuracy suggests that the model is improving in segmenting medical images accurately. These graphs help researchers and practitioners fine-tune models, detect overfitting or underfitting, and decide when to stop training, ensuring that the model achieves the desired level of segmentation accuracy for clinical applications. In Fig. 8, we present the average loss and accuracy graphs per epoch for both training and testing datasets.

Latest methods comparison with proposed models

We conducted a comparative analysis of our proposed models against recent studies in the field of MRI segmentation, including Zhang et al. [45], Nizamani et al. [46], and Huang et al. [47], which have demonstrated commendable performance in their recent research endeavors. Huang et al.'s model [47] is distinguished by its utilization of improved segmentation by using patch-based feature extraction. Nizamani et al.'s work [46] encompasses segmentation, clustering, and the application of CLAHE with UNET for feature extraction and tumor classification.

The results, as depicted in Tables 3 and 4 with training 70%, unveil that our proposed model exhibits superior performance when compared to another feature-based segmentation method. It's noteworthy that the other CLAHE does not perform optimally due to its limited efficiency in effectively segregating intricate datasets. Additionally, the studies by Huang et al. [48] and Aamir et al. [49] exhibit suboptimal performance, primarily due to their limited semantic understanding of complex data structures.

Ablation experiments

Additionally, we performed the ablation experiments by removing transformer and adding filters to UNET module directly and results are shown in Tables 5 and 6 that transformer addition with CLAHE is producing better results for FE1-UT method in both datasets and show the superiority of our method.

Discussion

The practical significance of our study underscores the promising advantages of harnessing sophisticated deep learning methods in the realm of medical image segmentation, with a particular focus on the analysis of brain tumor MRI scans. Nonetheless, it is imperative for researchers and healthcare professionals to remain cognizant of the study's limitations and proactively work towards mitigating them to ensure the secure and efficient integration of these techniques into clinical applications.

Practical applications

There are many practical implications of our proposed method:

- Brain Tumor Detection and Segmentation: The primary focus of the study is enhancing the precision of brain tumor MRI image segmentation. This technology can be deployed in clinical settings to assist radiologists and oncologists in accurately delineating tumor boundaries, which is crucial for treatment planning and monitoring disease progression [50–55].
- Tumor Volume Assessment: Accurate segmentation of tumors allows for precise measurement of tumor volumes over time. This is essential for tracking treatment response, assessing disease progression, and adjusting treatment strategies accordingly.

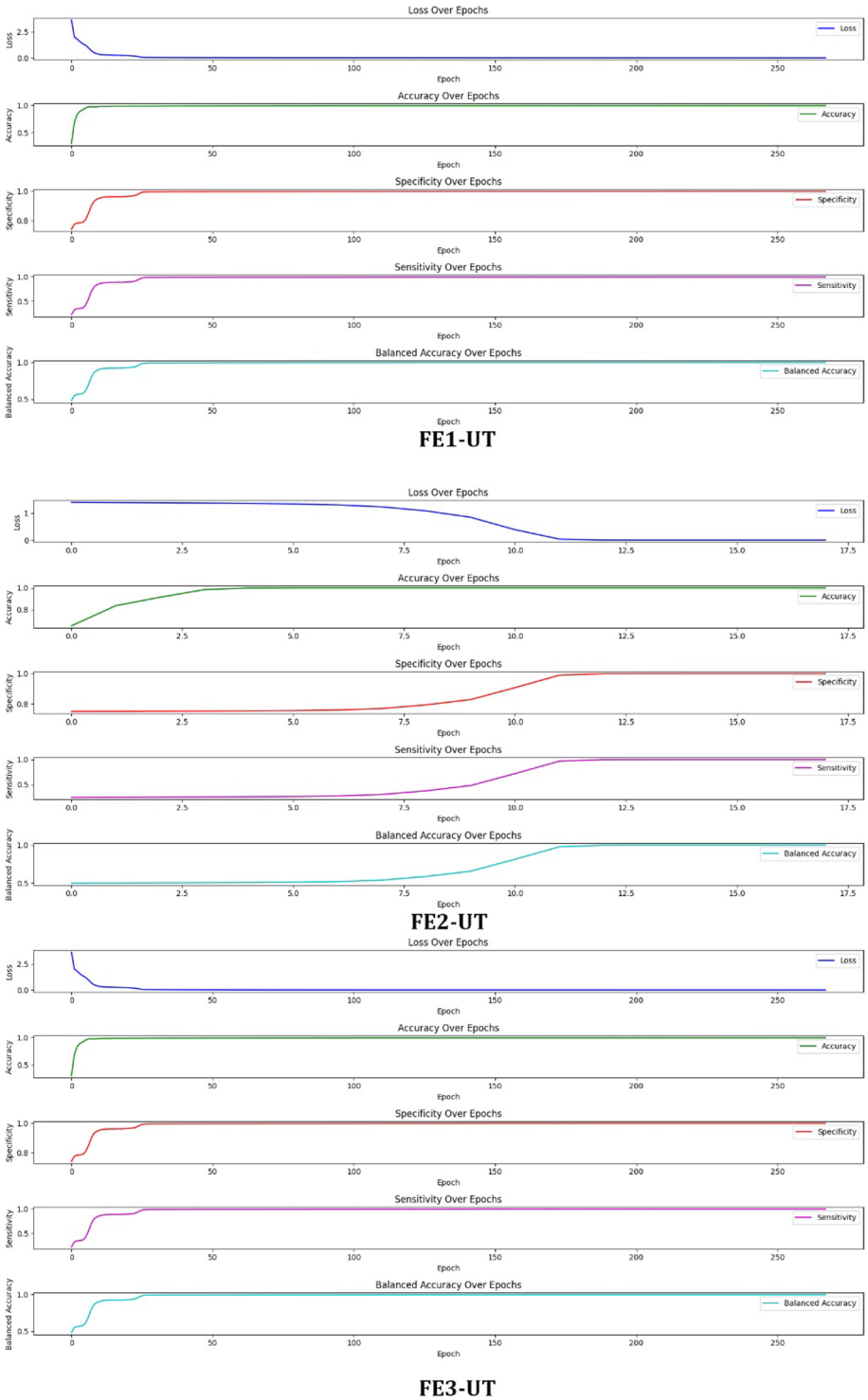

Fig. 8 Impact on performance with epochs

- Radiotherapy Planning: Medical image segmentation plays a vital role in radiotherapy planning. The technology can help radiation oncologists identify tumor regions and healthy tissues, enabling them to create treatment plans that deliver radiation therapy precisely to the affected area while sparing surrounding healthy tissue.

Table 3 Proposed methods comparison with latest methods in Brats dataset

Algorithm	Kappa	DSC	IoU	Accuracy	Balanced Accuracy
FE1-UT	0.669	0.670	0.571	0.806	0.799
FE2-UT	0.677	0.678	0.578	0.816	0.809
FE3-UT	0.702	0.703	0.6	0.846	0.838
Study [45]	0.6021	0.603	0.5139	0.7254	0.7191
Study [48]	0.6093	0.6102	0.5202	0.7344	0.7281
Study [49]	0.6318	0.6327	0.54	0.7614	0.7542

Table 4 Proposed methods comparison with latest methods in MSD dataset

Algorithm	Kappa	DSC	IoU	Accuracy	Balanced Accuracy
FE1-UT	0.544	0.549	0.413	0.799	0.651
FE2-UT	0.551	0.556	0.418	0.809	0.659
FE3-UT	0.571	0.577	0.433	0.838	0.683
Study [45]	0.4896	0.4941	0.3717	0.7191	0.5859
Study [48]	0.4959	0.5004	0.3762	0.7281	0.5931
Study [49]	0.5139	0.5193	0.3897	0.7542	0.6147

tumor removal while minimizing damage to healthy brain tissue.

- Disease Diagnosis and Staging: Beyond brain tumors, the methodology can be adapted to segment and analyze medical images for various other conditions, such as lung tumors, liver lesions, and cardiovascular diseases. This aids in disease diagnosis, staging, and treatment planning [56–59].
- Automated Diagnosis: The technology can be integrated into diagnostic systems to assist healthcare providers in making accurate and timely diagnoses. This can be especially valuable in situations where timely intervention is critical, such as stroke diagnosis.

Limitations

Our proposed algorithm is better in medical field but it has some limitations:

- Data Dependency: Deep learning models, including UNETs and Transformers, typically require substantial amounts of labeled data for training. In the medical field, obtaining large and diverse datasets can be challenging, especially for rare conditions or specific patient demographics. Limited data may hinder the model's generalizability and performance in diverse cases.

Table 5 Ablation experiments for Brats dataset

Model	Filter	Balanced Accuracy	F1 Score	Cohen's Kappa	Precision	Recall	Jaccard Index	ROC AUC
UNET	CLAHE	0.9890	0.9869	0.9826	0.9872	0.9867	0.9742	0.9912
FE1-UT		**0.9966**	**0.9964**	**0.9952**	**0.9966**	**0.9962**	**0.9928**	**0.9975**
UNET	MBOBHE	**0.9967**	**0.9956**	**0.9942**	**0.9956**	**0.9956**	**0.9913**	**0.9971**
FE2-UT		0.9449	0.9242	0.9006	0.9738	0.8794	0.8591	0.9358
UNET	MPHE	0.5544	0.2545	0.1873	1.0	0.1458	0.1458	0.5729
FE3-UT		0.5544	0.2545	**0.2038**	1.0	0.1458	0.1458	0.5729

Table 6 Ablation experiments for MSD dataset

Model	Filter	Balanced Accuracy	F1 Score	Cohen's Kappa	Precision	Recall	Jaccard Index	ROC AUC
UNET	CLAHE	0.9808	0.9735	0.9646	0.9735	0.9735	0.9483	0.9823
FE1-UT		**0.9864**	**0.9819**	**0.9758**	**0.9819**	**0.9818**	**0.9644**	**0.9879**
UNET	MBOBHE	0.9828	0.977	0.9693	0.977	0.977	0.955	0.9846
FE2-UT		**0.9853**	**0.9798**	**0.9731**	**0.9798**	**0.9798**	**0.9604**	**0.986**
UNET	MPHE	0.9842	0.978	0.9706	0.978	0.978	0.9569	0.9853
FE3-UT		**0.9842**	**0.9785**	**0.9713**	**0.9785**	**0.9785**	**0.9578**	**0.9856**

- Computationally Intensive: Training deep learning models, particularly those with extensive layers and parameters, can be computationally intensive. This may necessitate powerful hardware and longer training times, making it less accessible for smaller healthcare facilities with limited resources.
- Model Interpretability: Deep learning models, such as UNETs and Transformers, are often considered as "black boxes." It can be challenging to interpret the decision-making process of these models, which can be a critical concern in medical applications where transparency and interpretability are essential.
- Overfitting: Deep learning models are susceptible to overfitting, especially when dealing with small datasets. Overfit models may perform exceedingly well on the training data but generalize poorly to new, unseen cases. Regularization techniques and data augmentation are employed to mitigate this issue, but it remains a concern.
- Imaging Variability: Medical images can exhibit substantial variability due to differences in acquisition equipment, protocols, and conditions. The model's ability to handle such variability may be limited, potentially leading to decreased accuracy in real-world clinical settings.
- Clinical Validation: Although the model demonstrates high accuracy on publicly available datasets, its performance in a real clinical setting might differ due to variations in image quality, patient population, and clinical practices. Clinical validation and integration into healthcare systems are critical steps that must be addressed.
- Ethical and Privacy Concerns: The use of deep learning models in healthcare raises ethical and privacy concerns related to patient data security and consent. Proper data handling and adherence to ethical guidelines are essential when implementing such systems.
- Algorithm Bias: If the training data is not representative of the entire population, the model may exhibit bias, potentially leading to disparities in diagnosis and treatment recommendations.
- Deployment Challenges: Integrating deep learning models into clinical workflows and ensuring their seamless operation can be challenging. Healthcare institutions may require significant infrastructure and expertise for deployment and maintenance.

Conclusion

In conclusion, the precision of medical image segmentation is undeniably crucial in the modern healthcare landscape, significantly impacting diagnosis and treatment planning. Recent strides in deep learning have ushered in a new era by harnessing the capabilities of UNETs and Transformers to automate labor-intensive manual segmentation processes. However, despite these advancements, challenges persist, particularly when dealing with intricate anatomical structures and indistinct features, which can compromise accuracy.

Our study presents an innovative and effective solution to elevate the precision of brain tumor MRI image segmentation. We achieve this by seamlessly integrating UNET architecture with Transformers and incorporating feature improvement techniques, specifically MHE, CLAHE, and MBOBHE, to develop the high-performance image segmentation algorithms—FE1-UT, FE2-UT, and FE3-UT.

Our approach relies on three fundamental pillars. Firstly, we emphasize the significance of feature imrpovement during the image preprocessing stage. Through techniques like MHE, CLAHE, and MBOBHE, which employ contrast enhancement, we enhance the visibility of critical details within medical images. Secondly, our UT model is meticulously designed to enhance segmentation results through personalized layering within the UNET architecture. The incorporation of Transformers brings in contextual understanding and facilitates the capture of long-range dependencies in the data, thereby enabling more precise and context-aware segmentation.

The resulting model represents a comprehensive framework for achieving precise medical image segmentation, skillfully combining the power of UNETs, Transformers, and feature-enhanced filters. Our approach is not merely theoretical; it has been rigorously validated through experimental evaluations, which affirm its excellence in distinguishing complex brain tissues. In essence, our research makes a significant contribution to the ongoing transformation of healthcare practices. By pushing the boundaries of medical image segmentation and offering a highly accurate, automated solution for brain tumor MRI image segmentation, we are poised to enhance the quality of patient care, expedite diagnosis, and streamline treatment planning in the field of healthcare. The combination of deep learning, feature improvement, and advanced network architectures offers a promising path forward, potentially revolutionizing medical image analysis and improving patient outcomes.

Authors' contributions

Ahsan and haseeb worked for experiment. Chen and haseeb supervised. Haseeb,kashif and Ahsan prepared figures. All authors reviewed the manuscript.

Funding

This work was supported in part by the intelligent software and hardware system of medical process assistant and its application belong to "2030 Innovation Megaprojects"—New Generation Artificial Intelligence under Grant 2020AAA0109605".

Declarations

Competing interests
The authors declare no competing interests.

Received: 23 October 2023 Accepted: 22 November 2023
Published online: 06 December 2023

References

1. Zhang Z, Wang L, Zheng W, Yin L, Hu R, Yang B (2022) Endoscope image mosaic based on pyramid ORB. Biomed Signal Process Control 71:103261
2. Wu Y, Zhang L, Bhatti UA, Huang M (2023) Interpretable machine learning for personalized medical recommendations: A LIME-based approach. Diagnostics 13(16):2681
3. Bhatti UA, Huang M, Neira-Molina H, Marjan S, Baryalai M, Tang H, Bazai SU (2023) MFFCG–Multi feature fusion for hyperspectral image classification using graph attention network. Exp Syst Appl 229:120496
4. Zhuang, Y., Chen, S., Jiang, N., & Hu, H. (2022). An Effective WSSENet-Based Similarity Retrieval Method of Large Lung CT Image Databases. KSII Trans Internet Inform Syst. 16(7). https://doi.org/10.3837/tiis.2022.07.013
5. Zhuang, Y., Jiang, N., Xu, Y., Xiangjie, K., & Kong, X. (2022). Progressive Distributed and Parallel Similarity Retrieval of Large CT Image Sequences in Mobile Telemedicine Networks. Wireless Commun Mobile Comput. 2022. https://doi.org/10.1155/2022/6458350
6. Agravat RR, Raval MS (2021) A survey and analysis on automated glioma brain tumour segmentation and overall patient survival prediction. Arch Comput Methods Eng 28:4117–4152
7. Ranjbarzadeh R, Caputo A, Tirkolaee EB, Ghoushchi SJ, Bendechache M (2022) Brain tumour segmentation of MRI images: a comprehensive review on the application of artificial intelligence tools. Comput Biol Med 152:Article 106405
8. Bhatti UA, Tang H, Wu G, Marjan S, Hussain A (2023) Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence. Int J Intell Syst 2023:1–28
9. Jyothi P, Singh A.R. (2022). Deep learning models and traditional automated techniques for brain tumour segmentation in MRI: a review. Artif Intell Rev. 1–47.
10. Rao CS, Karunakara K (2021) A comprehensive review on brain tumour segmentation and classification of MRI images. Multimed Tool Appl 80(12):17611–17643
11. N. Sharma and L. M. Aggarwal (2010). Automated medical image segmentation techniques. Jmedical physics/Association of Medical Physicists of India. 35(1).
12. Krasteva V, Ménétré S, Didon J-P, Jekova I (2020) Fully convolutional deep neural networks with optimized hyperparameters for detection of shockable and non-shockable rhythms. Sensors 20(10):2875
13. Lu S, Liu S, Hou P, Yang B, Liu M, Yin L, Zheng W (2023) Soft Tissue feature tracking based on deep matching network. Comput Model Eng Sci 136(1):363–379. https://doi.org/10.32604/cmes.2023.025217
14. Sun, L., Zhang, M., Wang, B., Tiwari, P. (2023). Few-Shot Class-Incremental Learning for Medical Time Series Classification. IEEE J Biomed Health Informatics. https://doi.org/10.1109/JBHI.2023.3247861
15. Piccinini Gualtiero (2020) The First Computational Theory of Cognition: McCulloch and Pitts's "A Logical Calculus of the Ideas Immanent in Nervous Activity" P=107–C5.P91
16. LeCun Y, Bengio Y, Hinton G (2015) Deep Learn Nat 521(7553):436–444
17. Chua LO, Roska T (1993) The CNN paradigm. IEEE Trans Circuits Systems I Fundamental Theory Appl 40(3):147–156
18. Chattopadhyay A, Maitra M (2022) MRI-based brain tumour image detection using CNN based deep learning method. Neurosci Inform 2(4):100060
19. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. pp 3431–3440
20. Zeng, G., Yang, X., Li, J., Yu, L., Heng, P. A., & Zheng, G. (2017). 3D U-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3D MR images. In Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8. pp. 274–282. Springer International Publishing.
21. Zhang C, Benz P, Argaw D. M, Lee S, Kim J, Rameau F, Kweon I. S (2021) Resnet or densenet? introducing dense shortcuts to resnet. Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp 3550–3559
22. Jacobsen J. H, Van Gemert J, Lou Z, Smeulders A. W (2016) Structured receptive fields in cnns. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 2610–2619
23. Micallef N, Seychell D, Bajada CJ (2021) Exploring the u-net++ model for automatic brain tumor segmentation. IEEE Access 9:125523–125539
24. Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.
25. Wang H, Xu G, Pan X, Liu Z, Tang N, Lan R, Luo X (2022) Attention-inception-based U-Net for retinal vessel segmentation with advanced residual. Comput Electr Eng 98:107670
26. Tsipras, D., Santurkar, S., Engstrom, L., Ilyas, A., & Madry, A. (2020, November). From imagenet to image classification: Contextualizing progress on benchmarks. In International Conference on Machine Learning. pp. 9625–9635. PMLR.
27. Yin H, Vahdat A, Alvarez JM, Mallya A, Kautz J, Molchanov P (2022) A-vit: Adaptive tokens for efficient vision transformer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp 10809–10818
28. Chen Z, Zhu Y, Zhao C, Hu G, Zeng W, Wang J, Tang M (2021) Dpt: Deformable patch-based transformer for visual recognition. Proceedings of the 29th ACM International Conference on Multimedia. pp 2899–2907
29. Rendón-Segador, F. J., Álvarez-García, J. A., & Varela-Vaca, A. J. (2023). Paying Attention to cyber-attacks: A multi-layer perceptron with self-attention mechanism. Comput Secur. 103318.
30. Touvron H, Cord M, Jégou H (2022) Deit iii: Revenge of the vit. European Conference on Computer Vision. Cham, Springer Nature Switzerland, pp 516–533
31. d'Ascoli, S., Touvron, H., Leavitt, M. L., Morcos, A. S., Biroli, G., & Sagun, L. (2021). Convit: Improving vision transformers with soft convolutional inductive biases. In International Conference on Machine Learning. pp. 2286–2296. PMLR.
32. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2022) Swin-UNET: UNET-like pure transformer for medical image segmentation. European conference on computer vision. Springer Nature Switzerland, Cham, pp 205–218
33. Kiya H, Nagamori T, Imaizumi S, Shiota S (2022) Privacy-preserving semantic segmentation using vision transformer. J Imag 8(9):233
34. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., & Zhou, Y. (2021). TransUNET: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.
35. Lin A, Chen B, Xu J, Zhang Z, Lu G, Zhang D (2022) Ds-transUNET: Dual swin transformer u-net for medical image segmentation. IEEE Trans Instrum Meas 71:1–15
36. Ning, Y., Zhang, S., Xi, X., Guo, J., Liu, P., & Zhang, C. (2021, December). Cac-emvt: Efficient coronary artery calcium segmentation with multi-scale vision transformers. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1462–1467. IEEE.
37. Wang L, Pan L, Wang H, Liu M, Feng Z, Rong P, Peng S (2023) DHUNET: Dual-branch hierarchical global–local fusion network for whole slide image segmentation. Biomed Signal Process Control. 85:104976
38. Setiawan A. W., Mengko T. R., Santoso O. S, Suksmono A. B (2013) Color retinal image enhancement using CLAHE. International conference on ICT for smart society. pp 1–3
39. Tang JR, Isa NAM (2017) Bi-histogram equalization using modified histogram bins. Appl Soft Comput 55:31–43
40. Hum YC, Lai KW, Mohamad Salim MI (2014) Multiobjectives bihistogram equalization for image contrast enhancement. Complexity 20(2):22–36
41. Weng Y, Zhou T, Li Y, Qiu X (2019) Nas-UNET: neural architecture search for medical image segmentation. IEEE Access 7:44247–44257
42. Maziarka, Ł., Majchrowski, D., Danel, T., Gaiński, P., Tabor, J., Podolak, I., & Jastrzębski, S. (2021). Relative molecule self-attention transformer. arXiv preprint arXiv:2110.05841.

43. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Van Leemput K (2014) The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imag. 34(10):1993–2024
44. Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, Cardoso MJ (2022) The medical segmentation decathlon. Nat Commun 13(1):4128
45. Zhang Y, Chen J, Ma X, Wang G, Bhatti UA, Huang M (2024) Interactive medical image annotation using improved Attention U-net with compound geodesic distance. Expert Syst Appl 237:121282
46. NIZAMANI, A. H., Chen, Z., NIZAMANI, A. A., & Bhatti, U. A. (2023). Advance Brain tumor segmentation using feature fusion methods with deep U-Net model with CNN for MRI data. J King Saud Univ-Comput Inform Sci. 101793.
47. Huang S, Huang M, Zhang Y, Chen J, Bhatti U (2020) Medical image segmentation using deep learning with feature enhancement. IET Image Proc 14(14):3324–3332
48. Lou Z, Gong YQ, Zhou X, Hu GH (2018) Low expression of miR-199 in hepatocellular carcinoma contributes to tumor cell hyper-proliferation by negatively suppressing XBP1. Oncol Lett 16(5):6531–6539. https://doi.org/10.3892/ol.2018.9476
49. Huang A, Zhou W (2023) Mn-based cGAS-STING activation for tumor therapy. Chin J Cancer Res 35(1):19–43. https://doi.org/10.21147/j.issn.1000-9604.2023.01.04
50. Cao J, Chen C, Wang Y, Chen X, Chen Z, Luo X (2016) Influence of autologous dendritic cells on cytokine-induced killer cell proliferation, cell phenotype and antitumor activity in vitro. Oncol Lett 12(3):2033–2037. https://doi.org/10.3892/ol.2016.4839
51. Mao X, Chen Y, Lu X, Jin S, Jiang P, Deng Z, Zhu X, Cai Q, Wu C, Kang S (2023) Tissue resident memory T cells are enriched and dysfunctional in effusion of patients with malignant tumor. J Cancer 14(7):1223–1231. https://doi.org/10.7150/jca.83615
52. Li, M., Wei, J., Xue, C., Zhou, X., Chen, S., Zheng, L.,... Zhou, M. (2023). Dissecting the roles and clinical potential of YY1 in the tumor microenvironment. Front Oncol. 13. https://doi.org/10.3389/fonc.2023.1122110
53. Chen S, Zeng J, Huang L, Peng Y, Yan Z, Zhang A, Xu D (2022) RNA adenosine modifications related to prognosis and immune infiltration in osteosarcoma. J Transl Med 20(1):228. https://doi.org/10.1186/s12967-022-03415-6
54. Xu H, Wang H, Zhao W, Fu S, Li Y, Ni W, Xin Y, Li W, Yang C, Bai Y, Zhan M, Lu L (2020) SUMO1 modification of methyltransferase-like 3 promotes tumor progression via regulating Snail mRNA homeostasis in hepatocellular carcinoma. Theranostics 10(13):5671–5686. https://doi.org/10.7150/thno.42539
55. He B, Dai C, Lang J, Bing P, Tian G, Wang B, Yang J (2020) A machine learning framework to trace tumor tissue-of-origin of 13 types of cancer based on DNA somatic mutation. Biochimica et Biophysica Acta (BBA) - Mol Basis Dis 1866(11):165916. https://doi.org/10.1016/j.bbadis.2020.165916
56. Lin, Q., Xiongbo, G., Zhang, W., Cai, L., Yang, R., Chen, H., Cai, K. (2023). A Novel Approach of Surface Texture Mapping for Cone-beam Computed Tomography in Image-guided Surgical Navigation. IEEE J Biomed Health Inform. https://doi.org/10.1109/JBHI.2023.3298708
57. Yang S, Li Q, Li W, Li X, Liu A (2022) Dual-Level representation enhancement on characteristic and context for image-text retrieval. IEEE Trans Circuits Syst Video Technol 32(11):8037–8050. https://doi.org/10.1109/TCSVT.2022.3182426
58. Wang Y, Xu N, Liu A, Li W, Zhang Y (2022) High-order interaction learning for image captioning. IEEE Trans Circuits Syst Video Technol 32(7):4417–4430. https://doi.org/10.1109/TCSVT.2021.3121062
59. Xu, H., Van der Jeught, K., Zhou, Z., Zhang, L., Yu, T., Sun, Y.,... Lu, X. (2021). Atractylenolide I enhances responsiveness to immune checkpoint blockade therapy by activating tumor antigen presentation. J Clin Investig. 131(10). https://doi.org/10.1172/JCI146832

Publisher's Note