Received: November 28, 2023. Revised: January 17, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

High-Resolution Model for Segmenting and Predicting Brain Tumor Based on Deep UNet with Multi Attention Mechanism

Hanaa M. Al Abboodi[1]* **Amera W. Al-funjan[2]** **Ahmed Aldhahab[1]**
Wafaa M. Salih Abedi[3] **Alaa H. Abdullah[1]** **Ameer Hamza Siraj[4]**

[1]*Department of Electrical Engineering, University of Babylon, Iraq*
[2]*College of Education for Pure Sciences, University of Babylon, Babylon, Iraq*
[3]*General Education Department, City University Ajman, UAE*
[4]*University of Udine, Udine, Italy*
* Corresponding author's Email: hanaa.ali@uobabylon.edu.iq

Abstract: The research presents a new technique for segmenting brain tumors using the UNet framework enhanced with an attention mechanism. By incorporating attention processes that selectively emphasize prominent aspects while recording comprehensive contextual information, our strategy overcomes the challenges of brain tumor delineation. The suggested UNet-attention model is intended to outperform traditional segmentation techniques regarding precision and clinical applicability. Integrating spatial and channel attention processes into the UNet design is one of our study's significant achievements. The spatial attention mechanism's focus improves the capacity of the model to differentiate the mechanism's tumor and non-tumor areas. Also, incorporating contextual clues from multi-scale hierarchies allows for a thorough comprehension of visual properties. The discrete wavelet transform has been applied as a feature extraction method to enhance the model performance regarding time and memory consumption. A wide range of datasets is evaluated in-depth, proving our UNet-attention model's superiority. Advanced deep learning is made possible by combining attention processes and contextual data to delineate tumors precisely and clinically. Many evaluation criteria involving dice scores, accuracy, mean IoU, sensitivity, specificity, and Hausdorff distance have been applied to evaluate our model performance in different aspects. The model attained a dice coefficient of 0.9971. The model's specificity of 0.9988 is particularly noteworthy, demonstrating its exceptional ability to identify regions without tumors accurately. The model also achieved 0.9986 accuracies, 0.9142 mean IoU, Hausdorff distance (mm) 3.48. These evaluation values were obtained for applying our model on flair images from BraTS 2020.

Keywords: BT detection, Intensity normalization, DCT, Deep-learning, UNet-attention, Segmentation.

1. Introduction

Tumors grow when aberrant cells divide and multiply out of control, and they may disrupt normal tissue or organ function [1, 2]. Different tumors have different causes, architectures, and groups of cells. The cerebrum is the primary site for developing early-stage malignancies, notwithstanding the potential for secondary tumors to metastasize to the brain from other anatomical regions [3]. Medical imaging has become an essential and inseparable element of modern medicine. Various imaging modalities constitute a broad collection of medical imaging techniques [4]. Imaging modalities like magnetic resonance imaging MRI are essential to brain tumor diagnosis by aiding physicians in their identification.

In studying brain tumors (BT), diagnosis is crucial because of the significant fatality rate associated with the disease [5].

Segmentation is vital in medical image processing, particularly in predicting brain tumors (BT). It involves isolating the tumor area from the backdrop, reducing the brain's tumor prediction process's computational complexity and the memory requirements of this process. It is now possible to segment 3D MRI images automatically, and this is

being achieved through applying different frameworks of convolutional neural networks (CNN) [6, 7]. Labeled 3D MRI scans from the BraTS (brain tumour and segmentation) collection have recently made detecting and segmenting brain tumors in scientific literature easier. [8]. Due to its robust characteristics that integrate the features of extractor and classifier, deep learning architecture and CNN are utilized to mitigate these challenges. The utilization of CNNs for developing computer-aided design (CAD) solutions is presently under consideration by many individuals. The effectiveness and astounding outcomes of CNN-based CAD applications are astonishing. Recent studies have explored the possibility of using CNN to categorize histological images of various tumors. The studies also deal with problems involving several classes. Tissue characterization is formulated and discussed in one publication as a part of lung disease diagnosis. Several obstacles must be overcome while developing and deploying automated CAD [9].

Many factors affect CNN's performance. The first one that involves a CNN's ability to extract features is the structure of its layers. Second, the amount of training data available has a role in CNN's capacity to generalize the input. As previously stated, the precision of CNN in classifying test data is contingent upon the factors involved. Hence, a thorough examination of many prospective methodologies for addressing the challenges associated with categorization is being conducted. Improving classification accuracy by combining texture features with CNN features and well-established classifiers is an alternate approach to dealing with concrete problems.

In our work, the CNN-based attention concept has been employed. Attention directs focus toward some visual field regions as necessary instead of encoding an image into a fixed vector. Attention enables the image feature to adapt and change as required, leading to more detailed and extensive descriptions for complex images. Our study adopted spatial and channel attention to handle brain tumor segmentation and prediction challenges. These mechanisms compute the spatial attention map and refine feature maps through the channel attention mechanism using global pooling and element-wise operations. In addition, our suggested method extracts texture features from the input dataset obtained by applying a 2D discrete wavelet transform. Generally, for the best texture feature representation, many methods utilize texture feature extraction algorithms, including wavelet transform characteristics, to reduce the dimensionality of the input data and computing complexity while simultaneously improving the resolution of brain tumor segmentation. The texture features obtained from applying 2D DWT, integrated with a sophisticated UNet framework, incorporate spatial and channel attention mechanisms to produce a high-resolution model for segmenting and predicting brain cancers using the BraTS 2020 dataset modalities.

Our paper is organized as follows: The most state-of-the-art and literature reviews are demonstrated in section 2. Deep learning and UNet attention with the suggested framework for the deep Unet multi-attention technique to identify the tumors in brain segments are explained in section 3. The dataset used to test and validate our suggested model is described in section 4. The methodology of our proposed model with all related theoretical and mathematical details is demonstrated in section 5 and its subsections. The suggested model results and evaluation details are explained in section 6. The segment prediction results are shown in section 7, and the comparative analysis is determined in section 8. The conclusion is defined in section 9.

2. Literature review

Medical image processing relies particularly on segmentation, especially for separating the tumor region from its surroundings, decreasing computational complexity and memory demands. Then, developing an ability to segment and predict 3D MRI images automatically. Due to the diversity of tumor types and sizes and the complexity of the brain's architecture, brain tumors are challenging. Several studies have been conducted in recent years to understand better how medical imaging modalities may be used for automated brain tumor segmentation. As a result of its excellent resolution and capacity to produce improved visualization of connective tissues, magnetic resonance imaging (MRI) is often used to diagnose brain tumors. [10]. Deep learning for medical image analysis is also becoming more common, particularly for identifying and localizing brain tumors. Segmentation tasks often relied on conventional machine learning methods. CNNs are a relatively advanced kind of deep learning method, resulting in significant progress in the discipline and enhanced automatic segmentation performance. In various medical image-processing applications, including brain tumor segmentation, CNNs exhibited outstanding results [11, 12].

CNN-based enhanced framework for brain tumor analysis has been presented by [13]. This work uses an object identification model called YOLOv2 to locate and recognize objects in an image. The suggested method has been tested and confirmed on

different issues of BraTS datasets, including BraTS (2018, 2019, and 2020), to detect tumors from MRI images. The proposed technique attained accuracy scores exceeding 0.90 in localizing, segmenting, and classifying brain tumors.

Many researchers have applied a hybrid method that integrates the characteristics of different classifiers with CNN to overcome the BT segmentation challenges [12]. Maqsood et al. [14] suggested a deep learning model and multiclass support vector machine SVM to diagnose BT, which had an accuracy of 97.47%. Meanwhile, Khairandish et al. [12] investigated the BT segmentation performance by combining the CNN feature and different classical classifiers. CNN-SVM obtained the better performance with an accuracy of around 98.4. Younis et al. [15] presented visual geometry group (VGG16) as a better version of CNN and Ensemble learning for detecting brain tumors. The models have been evaluated by computing the model prediction accuracy and obtained 98.5 and 98.1 for the VGG16 and ensemble models. A significant drawback of the methods mentioned above is their reliance on a limited number of measures to evaluate the performance of their suggested models. This neglect caused an omission of crucial details that impacted the robustness of the assessment.

CNN-based UNet architecture has become more frequently employed in medical image processing, including classification and segmentation [16, 17], in the encoder-decoder CNN design known as UNet, which is based on ignoring the encoder and decoder layers interconnection. These skip connections aim to learn abstract representations of the input data without losing the high-quality features of the input image [18]. Several studies have employed UNet and UNet-based algorithms for BT segmentation, with the BraTS dataset as their image source. A three-layer encoder-decoder fully convolutional neural network (FCNN) model with a 3D structure was proposed by [19]. The researchers adopted four specific evaluation methods to demonstrate the effectiveness of their proposed method in segmenting brain tumors relying on the BraTS2020 dataset. Their findings correspond to 0.75, 0.87, and 0.76 as dice score values for the necrotic (core tumor), Edema (total tumor), and augmenting tumors. M. Lin et al. [20] produced a 3D Context U-Net model employing a deep supervision learning model to automate brain tumor segmentation. This segmentation was performed using Multiparametric MRI data from BraTS 2019. Their approach obtained 0.8693, 0.8013, and 0.7782 dice scores for Edema, necrotic, and enhancing tumors.

The authors in [21] incorporated a three-dimensional attention module into the decoder. The study utilized the straightforward U-Net design and successfully obtained a dice score 0.704 with the BraTS 2019. Theophraste et al. [22] presented a brain tumor segmentation model applying deeply-supervised 3D UNet-based self-ensembled. This model was developed as a solution for the BraTS 2020 challenge. They introduced two models derived from separate training processes, generating feature maps for BT segmentation. Each individual's two label maps were combined, considering each ensemble's performance for specific tumor subregions. Their proposed models attained 0.81, 0.91, and 0.85 dice values for the enhancing tumor, Edema, and necrotic. Kumar et al. [23] employed 3-D CNN to accurately distinguish gliomas and their constituents in MRI scans using dense connectivity patterns to minimize weight and incorporate residual connections. During training, hard mining was employed to prepare for challenging segmentation tasks. Hard mining was achieved by progressively raising the dice coefficient threshold to select the more complicated examples as the epoch count grew. The architecture attained 0.744, 0.876, and 0.714 dice values for necrosis, Edema, and enhancing tumors. The previously mentioned works have some limitations in their performance, which obstruct their models' abilities in segmenting and predicting brain tumors. 3D model's high processing complexity limits these studies.

To sidestep the intricate framework and the high processing requirements for 3D models, many researchers preferred 2D structured models to solve the BT detection and segmentation challenges. To automatically separate tumor areas from healthy tissue in the BraTS 2020 dataset, Sidratul et al. [24] suggested a 2D U-net design. The model attended 93% for dice. Unfortunately, the model lacked the hyper-parameters needed to evaluate its performance consistency, which was this work's most significant limitation. Using many versions of the BraTS datasets for training and improvement, Al Nasim et al. [25] introduced an enhanced structure of a 2D U-Net network. This study shows that the 2019 BraTS dataset yielded necrotic, Edema, and enhancing dice scores of 0.85, 0.94, and 0.88, respectively, which are statistically indistinguishable from the 2017–2018–2020 BraTS datasets. The fact that these models have only two dimensions is their biggest weakness; two-dimensional approaches tend to be inaccurate because they can't use deeper features.

Various CNN architectures have implemented diverse attention mechanisms to enhance their segmentation skills and overcome the two-dimensional methods related to their weak,

inaccurate data representation. The network's segmentation accuracy can be improved by training it to prioritize prominent parts of the input image through attention processes [21, 26]. Recent works indicate that the UNet-attention model integrates attention mechanisms into the existing UNet architecture. The UNet-attention model has demonstrated strong performance in multiple domains of medical image segmentation, including the separation of brain tumors. This achievement suggests that the model may be used in future potential applications within the medical imaging area [27]. A modified UNet architecture proposed by Myronenko et al. [28] attained a dice coefficient of 0.82 to segment brain tumors. Havaei et al. [29] gained 0.81 as a dice score by utilizing a wholly connected CNN built on the UNet framework. In this work, the researchers propose a hierarchical strategy that employs multiple UNet models on the BraTS dataset to boost the efficacy of segmentation. Few researchers have used the BraTS dataset to segment BTs by integrating attention mechanisms into UNet configurations. Ali et al. [30] proposed the self-attention module to capture and include long-range connections between regions within the input image. A dice coefficient of 0.87 was found to be produced by this model. Zhang and colleagues introduced a novel attention-based UNet architecture for predicting different tumor types, including Enhancing tumors. The suggested model exhibited differences in predicting tumors with low performance in predicting Enhancing tumors with a dice score of 0.51 when applied UNet with SoftMax loss. The researchers devised the channel attention technique to achieve feature equalization across the various channels of the input image [31]. The main limitation of all pre-reviewed methods was that their low performance led to insignificant tumor prediction areas due to their weak UNet structure and powerful hyperparameters.

On the other hand, feature extraction is vital to achieving better segmentation and prediction performance. Many works have investigated the importance of applying the different forms of wavelet transformation in enhancing brain tumor segmentation thoroughness. In their work, Hajiabadi et al. compared the methods that used variant forms of wavelet transform cooperating with deep learning to improve the brain tumor segmentation quality. They presented UNet CNN with multiwavelet transform as a supplementary component within deep networks. Their work was tested with the BRAT2015 dataset, and they achieved a 91.8 dice score in segmenting different MRI images from their dataset [32]. The authors in [33] suggested a texture feature extraction method such as GLCM (gray-level cooccurrence matrix) and DWT with SVM. The model attained 97% accuracy in detecting brain tumors from T1-weighted MRI images. Their work was limited to applying a single brain MRI modality.

In our work, multilevel texture features obtained from applying 2D discrete wavelet transformation and the elements of attention map obtained from using spatial and channel attention mechanisms of UNet have yielded encouraging outcomes in predicting different types of tumors involved in the BraTS dataset as medical imaging datasets.

3. Deep learning and UNet attention

Models incorporating deep learning principles in medical image analysis have recently become more prominent. Among these models, the UNet architecture has demonstrated excellent performance in medical picture segmentation tasks, such as brain tumor segmentation. The UNet design uses a symmetrical encoder-decoder network with skip links to segregate objects in an image properly. Although UNet architecture usually works, there are cases where it fails to extract all necessary information from an image, leading to incorrect segmentation. The UNet architecture incorporates attention approaches to address this issue and enhance the efficiency of the segmentation process. Through attention processes, by focusing on what's most crucial, the network enhances images. The accuracy of tumor segmentation in medical images is improved by the attention-based UNet model, which selectively supplements the essential components needed for an effective segmentation procedure.

This study classifies brain tumors using the BraTS dataset and the UNet attention model. Our analysis finds that the UNet attention model outperforms other cutting-edge approaches, including the classic UNet architecture, with an impressive 99% accuracy rate. These findings suggest that medical image segmentation and related tasks could benefit from attention-based UNet models. Fig. 1 explains the recommended design of the deep UNet attention approach for identifying brain tumors in brain segments. The brain image is retrieved from the database in the first step. The next phase is feature extraction, followed by data augmentation and pre-processing. The technique is divided into two parts: testing and training. Deep CNN is used in all stages to predict the segment during the model training process and to categorize the results with the testing data to determine whether a tumor is present.

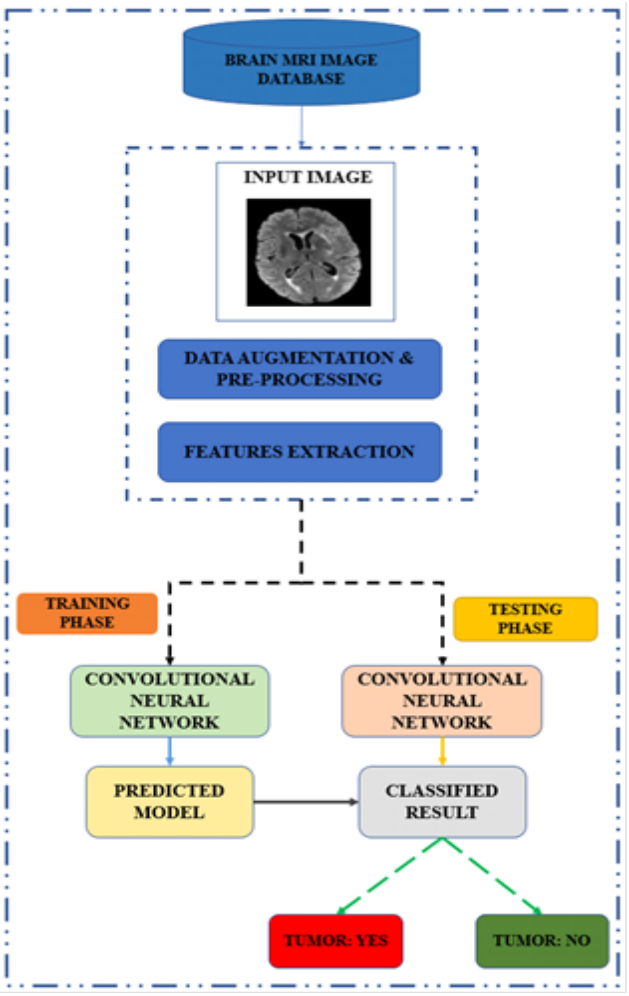

Figure. 1 Deep UNet-attention proposed framework for bain tomur detection

This study uses a UNet-based deep learning algorithm with spatial and channel attention techniques to extract features from tumor-containing brain MRI data. Fig. 1 depicts the suggested deep-learning scheme for brain tumor determination. First, retrieve the brain image from the database. Feature extraction, data augmentation, and pre-processing follow. The method includes testing and training. Deep CNN is utilized throughout, with DL first. Deep convolutional neural networks anticipate the model during training and categorize outcomes during testing. The model also considers tumors. Deep learning produced better categorization results on datasets, including medical images. Multiple convolutional layers collect information from input images to create feature maps. A model with a 99% overall accuracy was achieved using the BraTS dataset.

4. Dataset collection

For the BraTS dataset collection, MRI scans of brain tumors were obtained from various institutions, including hospitals, colleges, and other organizations. The dataset was analyzed for diagnosing and segmenting brain tumors. The dataset is modified annually, and the most recent release (BraTS 2020) includes annotated MRI scans from 335 individuals [8]. The dataset's modalities consist of T1-weighted MRI and T2 MRI. (T2-weighted MRI), FLAIR or fluid-attenuated inversion recovery and T1CE MRI. Different forms of information regarding the brain tissue are provided by each modality, which can aid in identifying and segmenting the tumor. While T2-weighted MRI is sensitive to Edema and alterations in the white matter, T1-weighted MRI offers high contrast between grey and white matter. When looking for Edema and necrosis surrounding a tumor, FLAIR MRI is beneficial, and T1-weighted MRI with contrast enhancement can be used to show specific regions of active tumor development. The T1CE MRI sequence provides detailed images of the body's interior anatomy and highlights aberrant structures such as tumors. However, a more accurate and thorough image of the brain and the tumor may be made by integrating data from several modalities. For this reason, each patient in the BRATS dataset has data from various modalities, and many algorithms for brain tumor segmentation and diagnosis combine these modalities.

5. Methodology

The UNet attention model is applied to complex operations, including convolution, attention mechanisms, and loss computation. These processes successfully enable the model to separate brain tumors from MRI images, improving medical diagnosis and treatment. Fig. 2 explains our suggested model methodology for BT segmentation using CNN with deep UNet-attention architecture. The following subsections will explain our applied methodology with its theoretical and mathematical representation.

5.1 Pre-processing

The most essential step in any segmentation and classification system is the pre-processing to obtain the necessary information for the segmentation procedure. The input images must be precisely aligned and resized to the designated intensity level. In our methodology, intensity normalization has been applied to enhance the segmenting procedure and create a more stable and effective process. The intensity normalization technique aims to homogenize the intensity levels of MRI images across different scans, patients, and imaging modalities. Minimizing intensity differences and

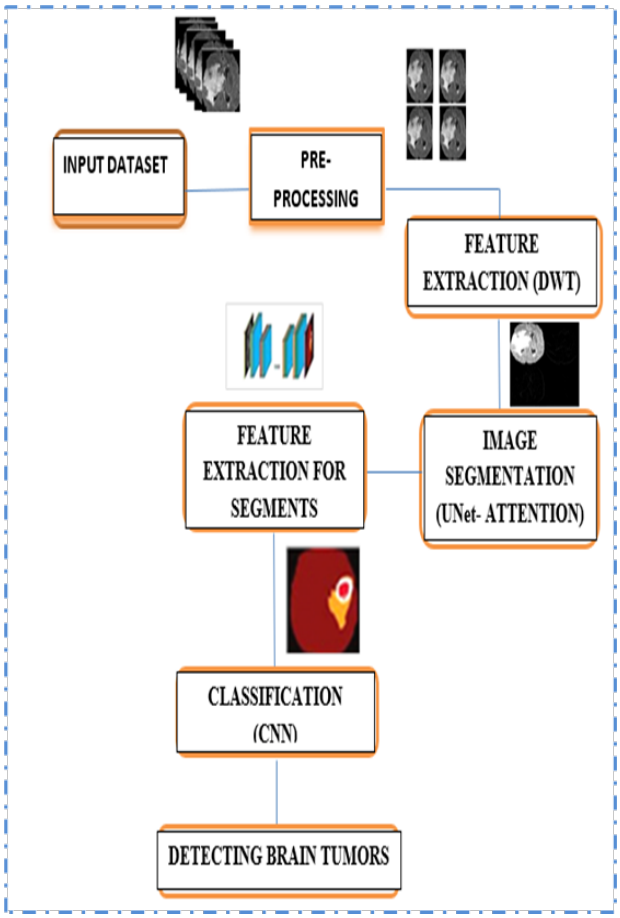

Figure. 2 Methodology of deep UNet-based multi-attention model and DWT featured data for BT segmenting and predicting

focusing on tumor-specific features, standardized MRI intensity distributions have created more reliable and accurate descriptions of the core anatomical structures.

Applying intensity normalization reduces the impact of MRI image brightness and contrast mismatches, which improves the model's generalizability, makes it easier for the segmentation model to adapt to different datasets, and ensures that intensity differences do not introduce biases. It also improves the learning algorithm by removing intensity-dependent noise, leading to a more focused and refined model, and makes it better at identifying tumors and non-tumor areas.

$$\acute{v} = \frac{v-min}{max-min}(new_{max} - new_{min}) + new_{min} \quad (1)$$

Eq. (1) describes the min-max normalization employed in our model to convert the input data from its original units into a new interval for the input feature. The symbols "*min*" and "*max*" represent the data's least and maximum absolute values. $\acute{v}$ means the updated value of each entry inside the dataset.

The variable v, represents the previous value of each entry in the dataset. The functions new_{max} and new_{min} represent the range's maximum and minimum values, respectively, corresponding to the required boundary values [34]. In addition to the raw data normalization, data augmentation procedures like rotation, flipping, and scaling are used to improve its diversity and resiliency.

5.2 Feature extraction

One of the most vital parts of any classification system is feature extraction. The wavelet transform is a significant feature extraction and discrimination method, a signal-processing technique representing images in both the time and frequency domains. The discrete wavelet transform (DWT) considers the temporal aspect and the concurrent resolution of frequency. DWT is extensively employed in various fields to solve many challenging issues in image processing. DWT was developed utilizing multi-resolution analysis and uses two fundamental functions. The scale function represents the low-pass filter, and the wavelet function represents the high-pass filter. *x[n]* performs half-band decomposition in the DWT using these two filters. The mathematical representation of the first level of decomposition of the discrete DWT is represented in the following:

$$low = \sum_n x[n].L[2k-n] \quad (2)$$

$$high = \sum_n x[n].H[2k-n] \quad (3)$$

The k values are between one and two, representing the wavelet function's scaling factor. Symbol *L[.]* refers to low-pass fillers, and *H[.]* refers to high-pass filters. The variable *x[n]* marks the input image. The variables "*low*" and "*high*" reflect the outputs of the two filters [35]. In our proposed brain tumor segmentation and prediction model, one-level decomposition of 2D DWT is applied to the pre-processed images to reduce dimensionality and discriminate the selected features. Fig. 3 shows the four sub-bands of 2D DWT results from images. low-low frequency (LL), low-high frequency (LH), high-low frequency (HL), and high-high frequency (HH) are the abbreviations for the respective bands. The LL sub-band contains all the required information, as seen in Fig. 3. A reduction in input image size necessitated the elimination of all sub-bands except the LL sub-band, which was therefore preserved. The resulting images with a length of 28*128 are fed to convolutional neural networks of deep UNet attention structure.

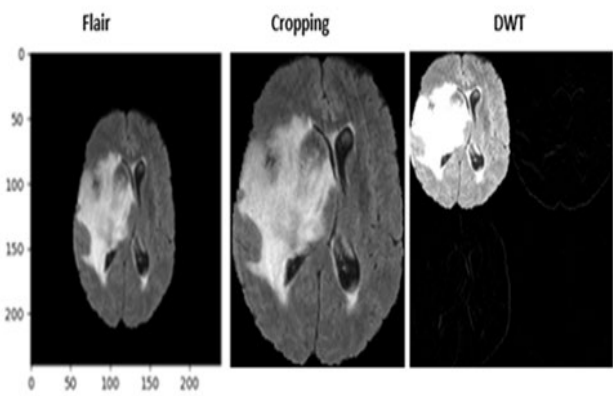

Figure. 3 2D DWT applied on flair image

5.3 Data preparation and gathering

This section uses the BRATS dataset to describe deep learning with the UNet-attention model for the brain tumor detection training process. A new data generator class implemented as a subclass of the Keras sequence class has been created as a part of the procedure. The generator class loads pre-process and prepares the data for the model in batches. This step enables the model to be trained incrementally on smaller pieces of the data and prevents memory problems caused by loading the entire dataset. The generator class provides a variety of methods, including the retrieving process, which produces batches of data using the data generation function. The T1CE, FLAIR, and segmentation images are loaded and resized utilizing this technique for each instance in the set to the proper sizes. The process modifies the indexes after each epoch to take the shifting of the data into account, which changes the order of the data. The batch size, input data dimensions, list of image IDs, number of input channels, and flag for data shuffle are all used to initialize the generator class. For training, validation, and testing purposes and as inputs to the UNet-attention model, three instances of the generator class are created at the end of the procedure. These generators produce their respective data sets using the training, validation, and testing generators. These newly created datasets have been employed for model training, hyperparameter tuning, and evaluation. MRI image conversion into a 3D tensor format for data entry into the UNet-attention model. The ground truth labels for the UNet-attention model are created using the pertinent annotations of the various tumor regions. The BraTS 2020 dataset has around 500 images in its original dataset. In this study, 149 images have been used. These images have been partitioned into two sets for model training and validation phases in a four-to-one ratio. In contrast, fifteen percent of the training set has been utilized as testing samples. So, the dataset has 149 3D Brats images, 30 for validation, and 119 for training our unit attention model.

5.4 Model selection and training

Using the BRATS dataset, our UNet model for segmenting brain tumors with an attention mechanism has been constructed and trained. UNet Attention architecture, which combines the channel and spatial attention systems, has been created. The number of convolutional layers, filters, and other hyperparameters has been decided depending on the experimental phase. A deep learning framework, such as TensorFlow, has been employed to implement the UNet-attention model, including the layers of the UNet architecture's spatial and channel attention mechanisms.

Encoder and decoder components make up the UNet architecture. The encoder section of the model utilizes convolutional and max pooling layers to perform downsampling on the input images, hence facilitating the extraction of relevant information. Upsampling and convolutional layers in the decoder portion rebuild the image to its original size while preserving the essential details. The UNet design implements the attention mechanism to concentrate on the crucial areas of the image while ignoring unimportant information. UNet-attention model architecture for BT Segmentation constructed from the following:

Convolutional layer: One way to calculate a convolutional layer's output is as follows:

$$y = \sigma(Wx + b) \tag{4}$$

The weight tensor, denoted as W, represents the weights of a neural network. The bias tensor, denoted as b, represents the network's biases, and the activation function, denoted as σ. The tensor y represents the network's output. Tensor x represents the network's input. Convolution layers iteratively extract increasingly complex image representations by sequentially processing localized information at each layer. Over time, this process effectively segregates pixels across a space of high dimensions based on their semantic attributes. The model's predictions depend on the information obtained from a wide receptive field, acquired sequentially. So, a linear transformation and a non-linear activation function are applied one after the other to the layer l output to produce the feature map. The most common option for activation functions is the ReLU activation function.

$$\sigma\left(x_{i,c}^{l}\right) = \max\left(0, x_{i,c}^{l}\right) \quad (5)$$

In the context where "i" represents spatial dimensions and "c" represents channel dimensions, the formulation of feature activations can be expressed as the following equation where the convolution operation is denoted by *:

$$x_c^l = \sigma\left(\sum_{c' \in F_l} x_{c'}^{l-1} * k_{c',c}\right) \quad (6)$$

x_c^l feature activation with c as feature dimension in layer l. c' is the feature dimension of the previous layer l-1, σ refers to the ReLU activation function. $k_{c',c}$ is the convolution kernel with dimension c' and depth c. The spatial subscript (i) is eliminated to keep the notation as simple as possible.

Encoder and decoder pathways: The UNet-attention model consists of a path for the encoder, which downsamples the input image and extracts features, and a pathway for the decoder, which upsamples the features and creates the segmentation map. The following equation can be used to illustrate the encoder pathway:

$$E_{i+1} = (Conv(E_i)) \quad (7)$$

Conv is a convolutional layer with a stride of 2 for downsampling the input, E_i is the output of the i^{th} encoder block, and E_{i+1} is the output of the $(i+1)^{th}$ encoder block. Concatenation of the up-sampled feature map U_i and the equivalent feature map C_i represent the decoder pathway are represented by the formulas:

$$D_i = Concatenate(U_i, C_i) \quad (8)$$

$$D_{i+1} = (Conv(D_i)) \quad (9)$$

Conv is a convolutional layer with a stride of 1 for upsampling the features, D_i is the output of the i^{th} decoder block, and D_{i+1} is the output of the $(i+1)^{th}$ decoder block.

Attention gates: The primary role of these gates in the UNet attention model is to highlight informative portions of the input images selectively. The following equations can be used to represent the attention gates:

First, calculate the feature maps for the encoder pathway using the following:

$$g_i = W_i \times X_i + b_i \quad (10)$$

Where g_i refers to the i^{th} encoder layer, the i^{th} encoder layer's learnable weight is designated as W_i. The input to the i^{th} encoder layer is X_i. The bias designation for the i^{th} encoder layer is b_i.

Second, calculate the feature maps for the decoder pathway using:

$$h_j = V_j \times Y_j + c_j \quad (11)$$

Where the computed feature map for the j^{th} decoder layer is represented by h_j, the j^{th} decoder layer's learnable weight is designated as V_j. The input to the j^{th} decoder layer is Y_j. The bias designation for the j^{th} decoder layer is c_j.

Create attention maps: The attention maps. a_i can be computed by multiplying the feature maps and applying the SoftMax function. Apply the attention maps to the decoder route feature maps using the following equation to choose which informative sections to emphasize.

$$a_i = softmax(g_i \,.\, h_i) \quad (12)$$

g_i is the feature map derived from the i^{th} encoder layer, and the feature map for the i^{th} decoder layer is represented by h_j. The element-wise multiplication's normalized exponential values are calculated using the SoftMax function. The softMax function is defined for a vector x.

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \quad (13)$$

x_i refers to the output vector where n is the sum of all the classes that make up this vector. The SoftMax activation function transforms the outputs of a neural network into a vector of probabilities, representing the distribution of possibilities for each input class in the n-class of the multiclass classification task. The SoftMax activation function generates a vector of n entries as output, where each item represents the probability of the input belonging to class i at index i. The SoftMax function scales the input values, ensuring that the probabilities' total equals 1. Categorization and attention mechanisms frequently employ this technique to determine the relative relevance of various components. Finally, apply the attention maps to the decoder feature maps by using the following equation:

$$U_i = a_i \odot h_i \quad (14)$$

Where U_i refers to the i^{th} decoder layer's weighted feature map. The attention map for the i^{th} layer is a_i.The feature map for the i^{th} decoder layer is represented by h_i.

Spatial attention mechanism: This method is used to compute the spatial attention map as explained in the following:

$$A(l) = \sigma(W_s(l) \times A(l-1) + B_s(l)) \quad (15)$$

A(l) refers to the calculated attention map features at layer *l*. σ shows the sigmoid functions, which are added as activation functions for the model. *Ws* refers to the learnable parameters (weights) for the spatial attention mechanism for layer *l*. *A(l−1)* refers to the activation output of the prior layer of the model. *Bs(l)* refers to the bias term for the spatial attention mechanism.

Channel attention mechanism: The channel attention technique uses global pooling and element-wise operations to refine feature maps. A feature map's spatial data is consolidated using average and maximum pooling. This method generates separate spatial context identifiers for the intermediate and maximum-pooled features. The two identifiers are then transmitted to a joint connection to create the channel attention map.

$$M_c(A) = \sigma\big(AvgPool(A) + MaxPool(A)\big) \quad (16)$$

Where *Mc* refers to the generated channel attention map, A refers to a spatial attention map. σ is the sigmoid activation function. The deep UNet initializes the model architecture with the attention function, which accepts the input layer. Max pooling is applied after two 3x3 convolutional layers with the ReLU activation function and 32 filters. Two more convolutional layers with 64 and 128 filters are used, followed by max pooling for each set. The spatial attention function employs a 3x3 convolutional layer structure and a sigmoid as an activation function with a multiplication operation within each pair of convolutional layers as an integral component of the attention mechanism. The model employs two more convolutional layers, each with 256 or 512 filters and a maximum pooling operation. The dropout function is applied at 0.2 to avoid the model overfitting. The model's decoder part begins with an up-sampling layer and then concatenates the relevant feature map from the encoder part. Two sets of 256 and 128 filter convolutional layers are then applied, followed by upsampling and concatenation with each set's matching feature map from the encoder component. Two more convolutional layers with 64 and 32 filters each are used to repeat the process. The successive two convolution 2D layers are identical to the first two convolution 2D layers. However, with 64 rather than 32 filters, get the max pooling 2D layer pool output as input. After several convolution 2D and max pooling, the network up-samples the feature maps using up-sampling 2D layers. Finally, it concatenates the equivalent feature maps from the down-sampled part using attached layers. This unique layout yields a detailed feature map displaying both high and low features. Upsampling is carried out repeatedly until the input's original resolution is attained. The last convolution layer generates the map for the fourth class of segmentation. Each pixel is assigned a probability distribution across all four classes using the SoftMax activation function. Like the regularisation approach, the dropout regularisation strategy uses an inputted dropout rate to avoid overfitting. Eq. (17) calculates the difference between the predicted and actual segmentation and is often employed as the loss function in the UNet-attention model.

$$L = (1-y) \times \log(1-y') + y \times \log(y') \quad (17)$$

If the l*og* is the natural logarithm, y' is the projected segmentation, and y is the ground truth segmentation. The size output (128, 128, 4) reflects the four classes of brain tumor segmentation (background, necrotic and non-enhancing tumor, Edema, enhancing tumor). Table 1 explains the deep attention configuration for our segmentation task. Figs. 3 and 4 describe the segmentation model performance, displaying the loss of performance and accuracy during the training and validation stage.

5.5 Hyperparameter fine-tuning

The model's hyperparameters are tuned with the aid of the validation set. The model's hyperparameters are adjusted to improve its efficiency.

Hyperparameters aim to investigate various settings related to the attention mechanism, batch sizes, dropout rates, and learning rates. It is essential to acknowledge that a more significant batch rate has not been established to mitigate overfitting, given the limited size of our dataset. The optimal results can be achieved with reduced processing cost and time by systematically analyzing each dataset through trial and error. Hyperparameters for the proposed model are made in advance rather than being learned during training, as explained in Table 2.

6. Model results and evaluation

Numerous measures are frequently employed for model evaluation on the BRATS dataset, including sensitivity, dice similarity coefficient (DSC), jaccard index, also known as intersection over union (IoU),

Table 1. UNet-attention deep model configuration

Layer	Features Size	No. of filters	Layer	Features Size	No. of filters
Input	128×128	2	Concatenate_1	16×16	512
Conv2D_block1	128×128	32	Conv2D_block12	16×16	256
Conv2d_block2	128×128	32	Conv2d_block13	16×16	256
Maxpoolin2D_1	64×64	32	Up_Sampling2D_2	32×32	256
Conv2D_block3	64×64	64	Conv2d_block14	32×32	128
Conv2d_block4	64×64	64	Concatenate_2	32×32	256
Maxpoolin2D_2	32×32	64	Conv2D_block15	32×32	128
Conv2D_block5	32×32	128	Conv2d_block16	32×32	128
Conv2d_block6	32×32	128	Up_Sampling2D_3	64×64	128
Maxpoolin2D_3	16×16	128	Conv2d_block17	64×64	64
Conv2D_block7	16×16	256	Concatenate_3	64×64	128
Conv2d_block8	16×16	256	Conv2D_block18	64×64	64
Maxpoolin2D_4	8×8	512	Conv2d_block19	64×64	64
Conv2D_block9	8×8	512	Up_Sampling2D_4	128×128	64
Conv2d_block10	8×8	512	Conv2d_block20	128×128	32
Dropout	8×8	512	Concatenate_4	128×128	64
Up_Sampling2D_1	16×16	512	Conv2d_block21	128×128	32
Conv2d_block11	16×16	256	Output	128×128	4

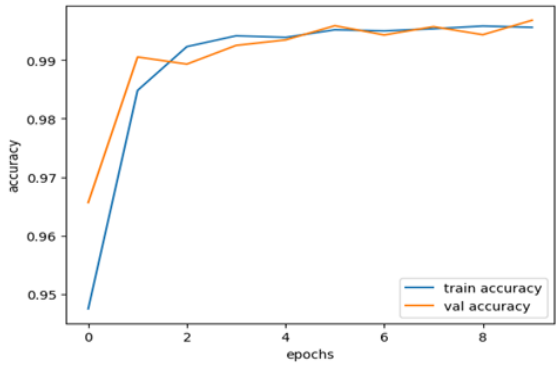

Figure. 4 The accuracy of our segmentation model on both the training and validation data sets

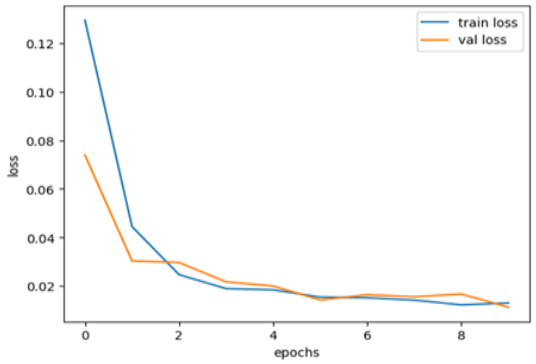

Figure. 5 The loss of our segmentation model on both the training and validation data sets

Table 2. Our proposed model hyperparameters

Parameters	**Values**
learning rate	0.001
Epochs	10
dropout rate	0.1
batch sizes	32

specificity, and hausdorff distance (HD). The model's specificity is measured by its ability to recognize every non-tumor negative pixel. The model's sensitivity describes its capacity to identify all positive instances (tumor pixels).

The dice coefficient calculates the amount of overlap between the predicted segmentation and the actual segmentation. It's outlined as:

$$\text{Dice} = (P + G) / (2 \times |P\ G|) \quad (18)$$

Where P and G are the sizes of the corresponding segmentations, |P G| is the size of the intersection of the two segmentations, with G representing the ground truth segmentation and P defining the projected segmentation. The dice coefficient is a numerical measure that runs between 0 and 1, where 1 indicates total overlap.

Sensitivity and specificity are binary classification measures that quantify the percentage of accurate positive and negative predictions. They are characterized as:

$$\text{Sensitivity} = (TP + FN) / (TP) \quad (19)$$

$$\text{Specificity} = TN / (TN + FP) \quad (20)$$

Table 3. Deep UNet model performance evaluation using different metrics

Type of Metrics	Model Performance with t2	Model Performance with tice	Model Performance with seg	Model Performance with t1	Model Performance with flair
Accuracy	0.9934	0.9914	0.9987	0.9945	0.9986
Mean IoU	0.9023	0.9147	0.9257	0.9375	0.9142
Dice Coeff	0.9147	0.9025	0.8978	0.9145	0.9971
Precision	0.9994	0.9997	0.9991	0.9887	0.9973
Sensitivity	0.9784	0.9571	0.9542	0.9677	0.9988
Specificity	0.9974	0.9847	0.9972	0.9924	0.9915
Loss	0.0134	0.0132	0.0135	0.0154	0.0187

Table 4. Hausdorff distance scores for all types of input images

Evaluation Metrics	Hausdorff distance (mm)
Flair	3.48
t1	3.61
t2	4.234
tice	2.96

FP (false positive) is the segmentation, and G is the ground truth segmentation. Where FP (false positive) is the number of non-tumor voxels that are incorrectly identified as a tumor, FN (false negative) is the number of tumor voxels that are incorrectly identified as non-tumor, and TP (true positive) is the number of correctly identified tumor voxels, TN (true negative) is the number of correctly identified non-tumor voxels.

Hausdorff distance quantifies the spatial dissimilarity between the anticipated and observed segmentations. It's outlined as:

$$H(A,B) = max(h(A,B), h(B,A)) \quad (21)$$

h(A, B) represents Hausdorff distance from set A to set B. The calculation of H (A, B) involves the symmetric Hausdorff distance, which considers dissimilarity in both directions. The directed hausdorff distance between B and A represents the distances between a point in the predicted segmentation and the closest point in the ground truth. Symmetrical Hausdorff distance is calculated by considering both directions to take the dissimilarity in both directions. The hausdorff distance may measure the discrepancy between projected and actual tumor borders, while a lower space suggests a tighter match between the boundaries. Moreover, combining hausdorff distance with other assessment measures, including mean IoU, dice coefficient, accuracy, and recall, is advised for better model performance evaluation.

IoU computes the number of segmentations overlap between predicted and actual segmentation as explained in the following:

$$IoU = [P \cap G] / [P \cup G] \quad (22)$$

$P \cap G$ refers to the intersection in the middle of two segmentations, $P \cup G$ is the size of the union between the two segmentations, P represents the anticipated segmentation, and G is the ground truth segmentation. The suggested model's evaluation

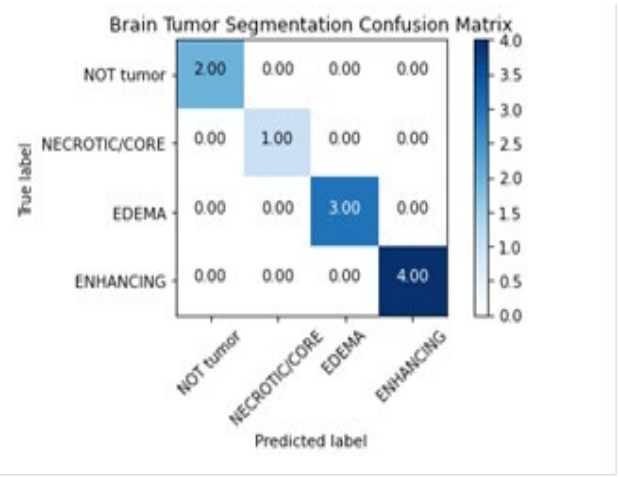

Figure. 6 Deep UNet-attention model confusion matrix

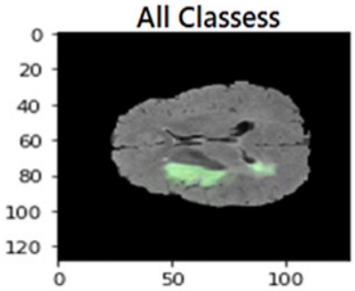

Figure. 7 Deep UNet model predicts all classes in flair image

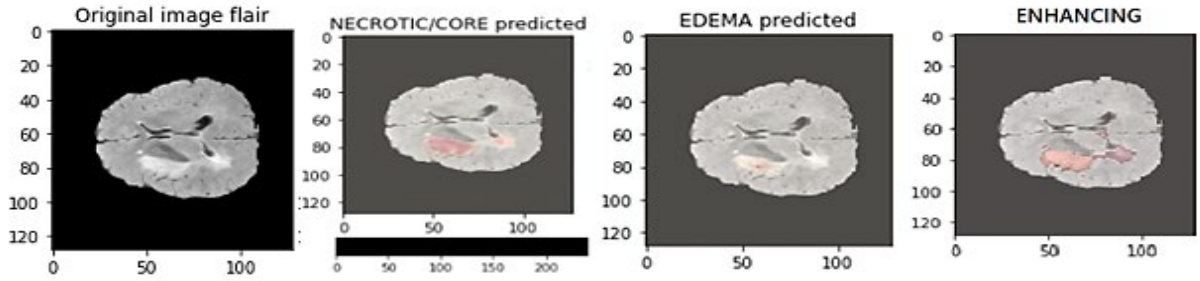

Figure. 8 The deep UNet model performance in predicting the four classes: Not tumor, NECROTIC/CORE, EDEMA, and ENHANCING from flair image

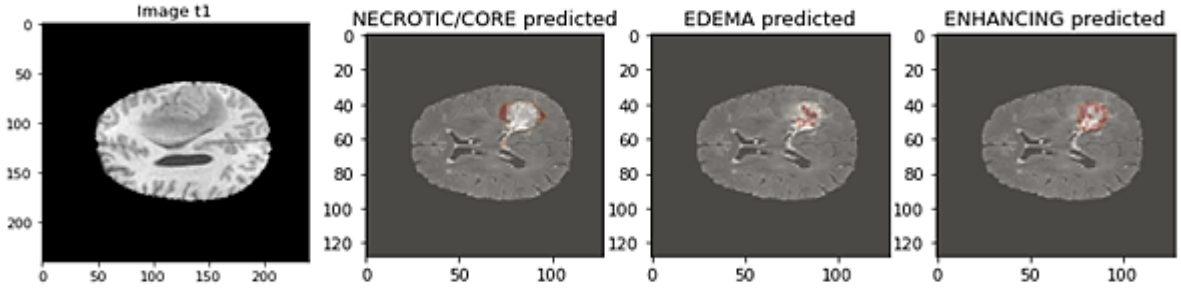

Figure. 9 The deep UNet model performance in predicting the four classes: Not tumor, NECROTIC/CORE, EDEMA, and ENHANCING from the t1 image

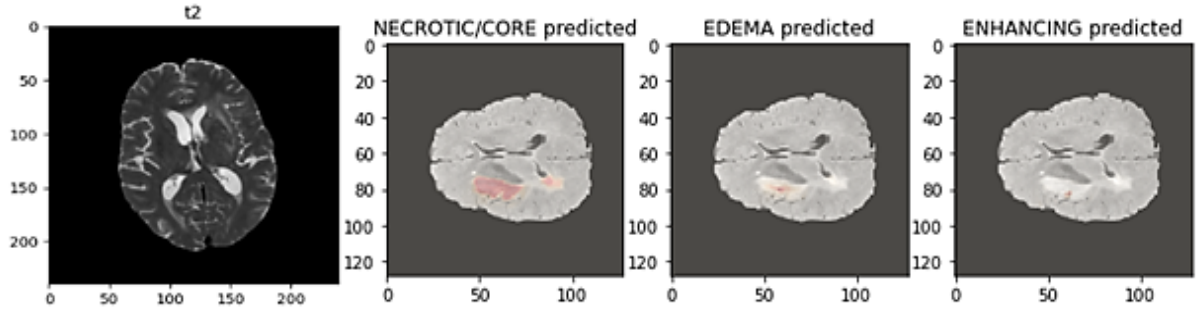

Figure. 10 The deep UNet model performance in predicting the four classes: Not tumor, NECROTIC/CORE, EDEMA, and ENHANCING from the t2 image

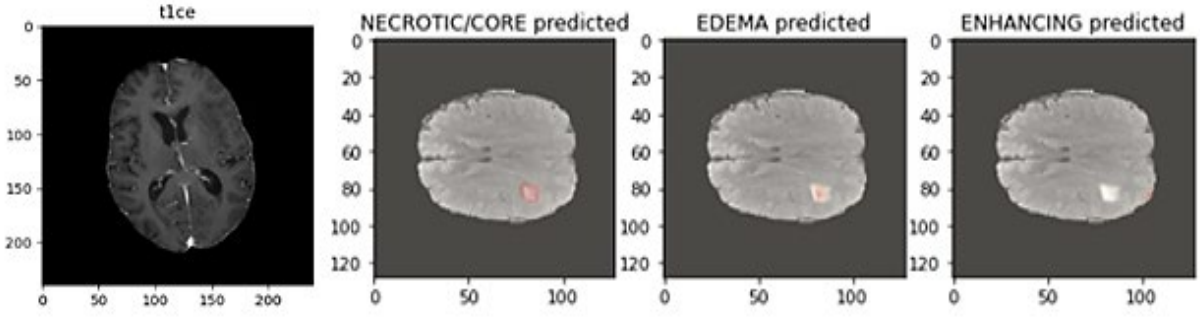

Figure. 11 The deep UNet model performance in predicting the four classes: Not tumor, NECROTIC/CORE, EDEMA, and ENHANCING from t1ce image

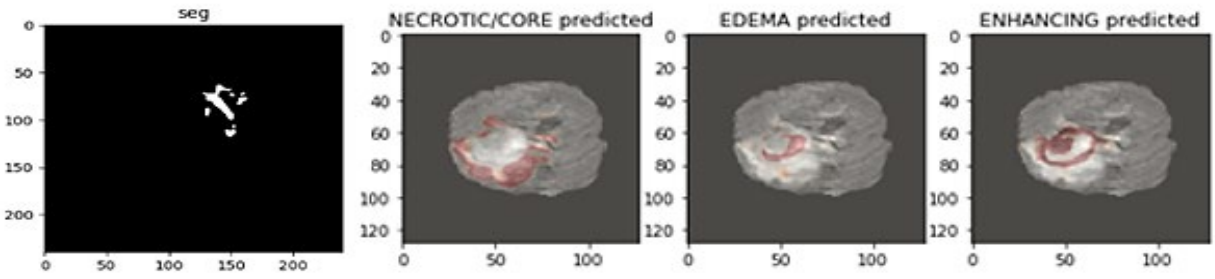

Figure. 12 The deep UNet model performance in predicting the four classes: Not tumor, NECROTIC/CORE, ED EMA, and ENHANCING in seg image

measures are laid out in Table 3; the table explains the evaluation scores for all dataset modalities used in this work. Moreover, our deep UNet segmentation model's Hausdorff distance scores for various input images are shown in Table 4.

7. Brain tumor segments prediction using the trained model

The following are the categories that our model can identify and predict. In input images, regions labeled "Not tumor" (label 0) are free of tumors. In contrast, areas labeled "NECROTIC/CORE" (label 1) contain necrotic or non-enhancing core tumors, and regions labeled "EDEMA" (label 2) include brain fluid accumulation. Areas marked "ENHANCING" (label 3) contain tumors often associated with high microvasculature and malignancies.

Correctly segmenting each image region and labeling each pixel in the segmentation map is required before inference can be performed using the trained model. Our model predicted every location in the original image from the test data, as seen in Figs. 8-12. The model accurately predicts the appearance of each element of the image using a combination of known and unknown information. A confusion matrix frequently assesses a classifier's performance by contrasting the projected class with the actual type. When segmenting brain tumors, the confusion matrix determines each pixel's anticipated labels with the ground truth labels supplied in the dataset. The suggested model's confusion matrix is shown in Fig. 6, explaining how the valid label stacks up against the model's prediction. Fig. 7 displays the results of a deep UNet model's prediction of all classes in a flair image.

8. Comparative analysis

In this work, the BraTS dataset and the integrated model that integrated the texture features extracted from the input dataset (BraTS2020) obtained from applying 2D discrete wavelet transform with a sophisticated UNet framework that incorporates spatial and channel attention mechanisms have been proposed to segregate brain tumors with a remarkable accuracy of 99%. For comparable purposes, the other methods explained in Tables 5 and 6 have been examined and contrasted with our strategy to comprehend our model overperformance. A YOLOv2 and CNN-based enhanced framework for brain tumor analysis has been presented by [13]. On the BraTS dataset, this technique has a 90% accuracy rate. An object identification model called YOLOv2 locates and recognizes objects in an image. However, object detection models might not perform as well as segmentation models like UNet since they are not created expressly for segmentation tasks. The approach's low accuracy may also be caused by CNN's lack of depth or complexity compared to our UNet-attention model.

On the BraTS dataset, deep neural networks and multiclass SVM have been employed for brain tumor diagnosis, which had an accuracy of 97.47% [14]. SVMs are helpful for some classification tasks. However, they are rarely employed for segmentation tasks like brain tumor segmentation. In contrast, our method used a UNet-attention model, created especially for image segmentation tasks, and has been demonstrated to work effectively in medical imaging applications.

The authors in [17] introduced an automated approach utilizing MRI data for brain tumor segmentation. The authors suggested a wholly automated method using a 2D UNet architecture on the BraTS 2020 dataset to isolate tumor areas from healthy tissue. The model had a dice score of 91% with Flair MRI. The model was limited since it only detected and segmented the tumor area without classifying the tumor types. In contrast, our model applied robust feature maps extracted by a multi-attention mechanism to segment and predict the three kinds of brain tumors involved in the BraTS dataset. Table 5 explains the comparative performance that involved different CNN structures based on various evaluation metrics.

Table 5. Comparative related works of BT segmentation

Evaluation Metrics	Our model	Ref. [13]	Ref. [14]	Ref. [17]
Accuracy	0.99	0.90	0.97	98.95
Mean IoU	0.91	0.88	0.95	-
Dice Coeff	0.99	0.87	0.93	91.23
Precision	0.99	0.91	0.92	-
Sensitivity	0.99	0.94	0.91	98.53
Specificity	0.99	0.95	0.93	99.14
Loss	0.02	-	-	0.06

Table 6. Comparative related works with different versions of BraTS in predicting each class

Model	Data set	Dice Score		
		NECROTIC	EDEMA	ENHANCING
self-ensembled 3D U-net [22]	BraTS 2020	0.85	0.91	0.81
3D encoder-decoder [28]	BraTS 2019	0.83	0.89	0.80
Supervised deep UNet [20]	BraTS 2019	0.80	0.86	0.77
3D Fully CNN [23]	BraTS 2020	0.74	0.87	0.71
Three layers encoder-decoder [19]	BraTS 2020	0.75	0.87	0.76
Enhanced UNet [25]	BraTS 2019	0.87	0.95	0.94
Proposed model	BraTS 2020-2019	0.994	0.997	0.991

On the other hand, Table 6 explains the comparative performance that involved different UNet structures. The performance evaluation relies on enhancing the dice coefficient value for predicting each class. It compares our proposed design and other work that applied different CNN-based Unet structures using BraTS 2019 and 2020 to indicate various types of brain tumors involving NECROTIC (core tumor), Edema (whole tumor), and enhancing tumors. In [25], a 2D UNet network has been introduced to be enhanced and trained using several versions of the BraTS datasets. This study shows that the BraTS datasets from 2017, 2018, 2019, and 2020 have similar dice scores for Necrotic, Edema, and Enhancing regions compared to the 2019 BraTS dataset. The dice scores achieved with the 2020 BraTS dataset are 0.88, 0.94, and 0.85 for Necrotic, Edema, and Enhancing tumors, respectively. For justification, our model has been tested with both BraTS 2019 and BraTS 2020 datasets, and the results were almost identical with all evaluation metrics that have been applied in our work. The previously mentioned approaches in [17, 25] relied on a two-dimensional depiction, which limited the amount of information that could be effectively employed as it could not successfully handle aspects that convey depth. In our proposed deep Unet multi-attention model, on the other hand, the attention mechanism effectively gets around the problem of two-dimensional representation by making the data representation better by extracting feature maps and adding robust layers to stop data quality leakage. Agravat et al. [19] introduced a three-layer encoder-decoder structure as a 3D model. They utilized diverse evaluation measures to showcase the efficacy of their proposed technique in segmenting brain tumors using the BraTS2020 dataset. The dice scores for the Necrotic, Edema, and Enhancing tumors were 0.75, 0.87, and 0.76, respectively. In addition, Lin et al. [20] proposed a 3D Context U-Net model that employs deep supervision learning to automate the segmenting of brain tumors. The segmentation was conducted via MRI data obtained from BraTS 2019. Their methodology yielded Dice coefficients of 0.8693, 0.8013, and 0.7782 for Edema (total tumor), Necrotic (core tumor), and Enhancing tumors, respectively. These two previously mentioned models possessed a highly intricate architecture of 600 and 500 epochs for the training phase, resulting in significant implications for the time and memory resources required to achieve their outcomes. In comparison, our suggested model got better results with only ten epochs, making the time and memory consumed less than previously reviewed models due to the multi-attention features that cooperate with texture features from applied DCT.

Theophraste et al. [22] introduced a brain tumor segmentation model that utilizes a deeply-supervised 3D UNet-based self-ensembled approach. This model was created as a resolution for the BraTS 2020 challenge. They presented two models created using distinct training procedures, producing a map for segmenting brain tumors. The two label maps per patient were merged, considering each ensemble's performance for specific tumor subregions. Their proposed models achieved dice coefficients of 0.81, 0.91, and 0.85 for the Enhancing tumor, Edema (total tumor), and Necrotic (core tumor), respectively. The model was constructed from a considerable number of layers to enhance the model's accuracy and enlarge the activation maps of the model, resulting in a substantial elongation of the training duration, hence diminishing the scope of the searchable area within the constrained timeframe of the finals. Kumar et al.

[23] utilized 3D CNN to differentiate gliomas and their components from MRI data. The architecture achieved die scores of 0.744, 0.876, and 0.714 for the Necrotic, Edema, and enhancing tumors, respectively. The model's network architecture consists of 77 layers, which has increased the model's complexity and the requirements for its implementation.

In contrast, our model employed deep UNet based on multi-attention mechanisms with 2D structure. Our model achieved high segmentation and prediction performance that overperformed all previously mentioned works. However, the attention method effectively addresses the limitation of 2D representation by providing enhanced representation for visual features. Our UNet architecture successfully extracts features and adds robust layers to reduce data quality leakage, which can result in information loss due to the limited exploitation of in-depth features caused by the 2D structure.

9. Conclusion

This study aims to enhance brain tumor segmentation's accuracy and therapeutic effectiveness by applying a novel UNet Attention model. The motivation behind our endeavors stemmed from the need to address the limitations of traditional segmentation techniques and provide healthcare professionals with an advanced instrument for accurate diagnosis and treatment strategizing. The fundamental contribution of our study involves integrating attention processes into the architecture of the UNet. A model was developed to highlight significant regions while considering the overall context by integrating spatial and channel attention mechanisms particularly crucial in segmenting brain tumors. A highly effective model was developed by integrating intensity normalization of MRI data pre-processing with attention-driven features and a comprehensive understanding of contextual cues, enabling accurate discrimination between tumor and non-tumor regions. The comprehensive examinations on a diverse dataset confirm our UNet -attention model's superior performance. The model also reduces computational complexity and memory requirements by adopting the discrete cosine transform (DCT) to extract features from medical images. That helps minimize the input array's dimensionality for the UNet model. The assessment criteria demonstrate exceptional performance across various evaluation criteria, including dice score, hausdorff distance, IoU, precision, and specificity. One significant aspect of our model's uniqueness is its ability to effectively identify non-tumor regions, which is paramount in clinical diagnosis. Our proposed model outperformed recent research that employed a variety of deep learning algorithms, including CNN, DNN, and SVM, with an overall accuracy of 99%. Adding attention gates improve feature maps and assists the model in concentrating on crucial areas for better segmentation performance.

Moreover, our algorithm effectively identified the tumor borders and captured the minor distinctions between various tumor types, producing exact segmentation. Implementing intensity normalization and data augmentation strategies was crucial in mitigating overfitting and enhancing the model's ability to generalize to unseen data. These pre-processing steps contributed significantly to the observed high accuracy of our improved model. Also, DWT has been applied to the normalized image as a feature extraction method to decrease the input data dimensionality, which aids in reducing our model computation complexity.

Our method provides a robust framework for medical professionals to exercise informed decision-making, as it is congruent with established medical expertise. This technique can attain improved precision in the identification of tumor boundaries. Incorporating artificial intelligence (AI) with medical expertise has promise in enhancing patient care and optimizing clinical outcomes. Our study outcomes demonstrated the impact of integrating the UNet-Attention model of multiple attention mechanisms that considerably influence segmenting brain tumor tasks. The integration of context awareness, deep learning, and attention processes has exhibited considerable potential in augmenting the precision and value of medical imaging.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

This study involved several vital components: conceptualization, formal analysis, software utilization, validation, inquiry, data curation, writing the initial draught, reviewing and editing, and visualization by Hanaa Al Abboodi, Amera Al-funjan, and Ahmed Aldhahab. Wafaa Salih Abedi, Alaa Abdullah, and Ameer Hamza Siraj were responsible for supervising project management allocation of authorship should be restricted to individuals who have made significant contributions to the reported work.

References

[1] S. Alqazzaz, X. Sun, X. Yang, and L. Nokes,

"Automated brain tumor segmentation on multi-modal MR image using SegNet", *Comp. Visual Media 5*, pp. 209–219, 2019.

[2] G. S. Tandel, A. Balestrieri, T. Jujaray, N. N. Khanna, L. Saba, and J. S. Suri, "Multiclass magnetic resonance imaging BT classification using artificial intelligence paradigm", *Computers in Biology and Medicine,* Vol. 122, p. 103804, 2020.

[3] N. Kesav and M. Jibukumar, "Efficient and low complex architecture for detection and classification of BT using RCNN with Two Channel CNN", *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, pp. 6229–6242, 2022.

[4] H. M. A. Abboodi, A. W. Alfunjan, N. A. Hamza, A. H. Abdullah, and B. H. Shami, "Supervised Transfer Learning for Multi Organs 3D Segmentation with Registration Tools for Metal Artifact Reduction in CT Images", *TEM Journal*. Vol. 12, No. 3, pp. 1342-1353, 2023.

[5] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling", *Current Oncology,* Vol. 29, pp. 7498-7511, 2022.

[6] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies", *Physics in Medicine & Biology*, Vol. 58, No. 13, p. R97, 2013.

[7] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images", *IEEE Transactions on Medical Imaging,* Vol. 35, No. 5, pp. 1240–1251, 2016.

[8] B. H. Menze, A. Jakab, S. Bauer, J. K. Cramer, K. Farahani, J. Kirby, and K. V. Leemput, "The multi-modal Brain Tumor Image Segmentation Benchmark (BRATS)", *IEEE Transactions on Medical Imaging*, Vol. 34, pp. 1993-2024, 2014.

[9] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning", *Computers in Biology and Medicine,* Vol. 111, p. 103345. 2019.

[10] M. Pareek, C. K. Jha, and S. Mukherjee, "Brain tumor Classification from MRI Images andCalculation of Tumor Area", *Advances in Intelligent Systems and Computing*, pp. 73–83, 2020.

[11] D. Konar, S. Bhattacharyya, B. K. Panigrahi, and E. C. Behrman, "Qutrit-inspired fully self-supervised shallow quantum learning network for brain tumor segmentation", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 33, pp. 6331-6345, 2021.

[12] M. O. Khairandish, M. Sharma, V. Jain, J. M. Chatterjee, and N. Z. Jhanjhi, "A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images", *IRBM*, Vol. 43, pp. 290-299, 2022.

[13] M. I. Sharif, J. P. Li, J. Amin, and A. Sharif, "An improved framework for brain tumor analysis using MRI based on YOLOv2 and convolutional neural network", *Complex & Intelligent Systems*, Vol. 7, pp. 2023-2036, 2021.

[14] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-modal brain tumor detection using deep neural network and multiclass SVM", *Medicine*, Vol. 58, p. 1090, 2022.

[15] A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, "Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches", *Applied Sciences*, Vol. 12, p. 7282, 2022.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *Medical Image Computing and Computer-Assisted Intervention*, Part III, Vol. 18, pp. 234-241, 2015.

[17] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasan, and A. Karim, "Brain Tumor Segmentation from 3D MRI Scans Using U-Net", *SN Computer Science*, Vol. 4, p. 386, 2023.

[18] Z. Lai, H. Sun, R. Tian, N. Ding, Z. Wu, and Y. Wang, "Rethinking Skip Connections in Encoder-decoder Networks for Monocular Depth Estimation", *ArXiv Preprint ArXiv*: pp. 2208-13441, 2022.

[19] R. Agravat and M. Raval, "3d semantic segmentation of brain tumor for overall survival prediction", In: *Proc. of International MICCAI Brain Lesion Workshop*, pp. 215–227, 2020.

[20] M. Lin, S. Momin, B. Zhou, K. Tang, Y. Lei, W. J. Curran, T. Liu, and X. Yang. "Fully automated brain tumor segmentation from multiparametric MRI using 3D context U-Net with deep supervision", *Medical Imaging 2021: Computer-Aided Diagnosis*, Vol. 11597, pp. 357-362, 2021.

[21] M. Islam, V. Vibashan, V. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, "Brain tumor segmentation and survival prediction using 3d attention unet", In: *Proc. of International MICCAI Brainlesion Workshop*, pp. 262–272, 2019.

[22] T. Henry, A. Carré, M. Lerousseau, T. Estienne, C. Robert, N. Paragios, and E. Deutsch, "Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks",

In: *Proc. of 6th International Workshop, BrainLes 2020, Lima, Part I6,* 2021.

[23] V. K. Anand, S. Grampurohit, P. Aurangabadkar, A. Kori, M. Khened, R. S. Bhat, and G. Krishnamurthi, "Brain tumor segmentation and survival prediction using automatic hard mining in 3d cnn architecture," In: *Proc. of International MICCAI Brainlesion Workshop*, pp. 310–319, 2020.

[24] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasan, and A. Karim, "Brain Tumor Segmentation from 3D MRI Scans Using U-Net", *SN Computer Science*, Vol. 4, No. 4, p. 386, 2023.

[25] M. A. A. Nasim, A. A. Munem, M. Islam, M. A. H. Palash, M. M. A. Haque, and F. M. Shah, Brain, "tumor segmentation using enhanced u-net model with empirical analysis", In: *Proc. of 2022 25th International Conference on Computer and Information Technology (ICCIT)*, pp. 1027-1032, 2022.

[26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, and B. Glocker, "Attention u-net: Learning where to look for the pancreas", *ArXiv Preprint ArXiv:* pp. 1804-03999, 2018.

[27] M. Noori, A. Bahri, and K. Mohammadi, "Attention-guided version of 2D UNet for automatic brain tumor segmentation", In: *Proc. of 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 269-275, 2019.

[28] A. Myronenko and A. Hatamizadeh, "Robust semantic segmentation of brain tumor regions from 3D MRIs", *Brainlesion: Glioma, Multiple Sclerosis, Stroke, and Traumatic Brain Injuries, Part II*, Vol. 5, pp. 82-89, 2020.

[29] M. Havaei, A. Davy, D. W. Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks", *Medical Image Analysis*, Vol. 35, pp. 18-31, 2017.

[30] T. M. Ali, A. Nawaz, A. U. Rehman, R. Z. Ahmad, A. R. Javed, T. R. Gadekallu, C. L. Chen, and C. M. Wu, "A sequential machine learning-cum-attention mechanism for effective segmentation of brain tumor", *Frontiers in Oncology*, Vol. 12, p. 873268, 2022.

[31] J. Zhang, X. Shen, T. Zhuo, and H. Zhou, "Brain tumor segmentation based on refined fully convolutional neural networks with a hierarchical dice loss", *ArXiv Preprint ArXiv*, pp. 1712-09093, 2017.

[32] M. Hajiabadi, B. A. Savareh, H. Emami, and A. Bashiri, "Comparison of wavelet transformations to enhance convolutional neural network performance in brain tumor segmentation", *BMC Medical Informatics and Decision Making*; Vol. 21, No. 1, pp. 1-11, 2021.

[33] M. Pareek, C. K. Jha, and S. Mukherjee, "Brain tumor classification from MRI images and calculation of tumor area", *Soft Computing: Theories and Applications: Proceedings of SoCTA 2018. Springer,* pp. 73-83, 2020.

[34] G. J. Kamani, R. S. Parmar, and Y. R. Ghodasara, "Data normalization in data mining using graphical user interface: A pre-processing stage", *Gujarat Journal of Extension Education,* Vol. 30, pp. 106-109, 2019.

[35] A. Aldhahab, G. Atia, and W. B. Mikhael, "Supervised facial recognition based on multi-resolution analysis with radon transform", In: *Proc. of 48th Asilomar Conference on Signals, Systems and Computers*, pp. 928-932, 2014.