Keywords
Tumour segmentation of brain, magnetic resonance imaging (MRI); 3-dimensional deep learning; convolutional neural networks; vision transformers; a wavelet transform; medical image analysis; brain tumor detection.

Authors
Pooja P P[1],
Research scholar (Part Time), Department of Computer Science and engineering, School of Engineering and Technology, CHRIST (Deemed to be University), Bangalore, and Assistant Professor, PSAIAC, Presidency University, Bangalore, India.
pooja.pp@res.christuniversity.in,pooja.p@presidencyuniversity.in

Dr Aruna S K[2],
Associate Professor, Department of AI and data Science engineering, School of Engineering and Technology, CHRIST (Deemed to be University), Bangalore, India,
aruna.sk@christuniversity.in

European Journal of Prosthodontics and Restorative Dentistry (2026) 35(5s), 83–95

Texture-Aware 3D Deep Learning in Segmenting Skull-Conjoined Brain Tumor

Abstract
Magnetic resonance imaging (MRI) cannot provide sufficient contrast, heterogeneity of the tumor texture, and vague tumor-bone interfaces that makes it difficult and challenging to carefully segment brain tumours that are located close to the cranial skull boundary of the brain. Traditional convolutional neural network (CNN) architecture tends to lose delicate structural aspects in these areas. The paper introduces a texture-sensitive hybrid deep learning architecture to skin skull-conjoined brain tumors segmentation of multi-modal brain tumor volumes using MRI volumes. The framework combines the contrast-adaptive preprocessing, three-dimensional multi-phase learning (3D-MPTL) on the texture and three-dimensional double density dual-tree Complex wavelet transform (3D-DDDTCWT) to obtain the volumetric texture and the multi-resolution directional features. A Swin Transformer encoder is used to extract long range spatial statistics and 3D U-Net network is modified to voxel tumor segmentation. Alongside, a hybrid particle swarm optimization and gravitational search algorithm (PSO-GSA) is applied to optimal network hyperparameter and training convergence. The experimental assessment of the BraTS dataset shows a better finalization of segmentation especially in the region of a skull and adjacent tumor with increased Dice scores in comparison to the current CNN and transformer-based assessment.
Experimental evaluation on the BraTS dataset demonstrates that the proposed method achieves Dice scores of 97.41%, 95.62%, and 93.78% for Whole Tumor, Tumor Core, and Enhancing Tumor regions, respectively. The model also achieves a Hausdorff Distance as low as 2.15 mm, indicating accurate boundary localization. Comparative analysis shows that the proposed framework outperforms state-of-the-art models by 2–4% improvement in Dice score, particularly in challenging skull-adjacent tumor regions.

Received:10-06-2026
Revised:16-07-2026
Accepted:20-07-2026
Doi:10.1922/ejprd.v34i5s.1655

INTRODUCTION

Brain tumors are one of the most health threatening neurological syndromes and are extremely difficult to diagnose in clinical aspects and offer treatment plans. Proper recognition and delineation of the image of tumor zones through magnetic resonance imaging (MRI) is the identifying factor to surgical planning, radiotherapy targeting, and checking the progression of the disease. MRI has high soft tissue contrast and allows one to see subregions of tumor like edema, necrotic center and enhancing areas of tumor. The Multi-modal MRI sequences which are T1-weighted, contrast-enhanced T1 (T1CE), T2-weighted, and fluid-attenuated inversion recovery (FLAIR) contain complementary structure and pathological data that aids clinicians in determining tumor properties [1], [2]. Nevertheless, manual definition of tumor markers is consuming and vulnerable to inter-observer differences and frequently unreliable in case of tumors with complex morphology or non-uniform texture. In order to deal with these problems, automated brain tumor segmentation tools have been extensively researched based on standardized datasets like the Brain Tumor Segmentation (BraTS) dataset, which includes standardized MRI data and expert-labeled segmentation (labels) in developing machine learning algorithms [1], [3].

Over the past years, deep learning methods have been very helpful in increasing the effectiveness of automated brain tumor segmentation systems. The CNNs have received extensive use in medical image analysis because of the fact that they can learn hierarchical feature representations directly through imaging data. The U-Net model is one of the most powerful CNN architectures in biomedical segmentation that provided such an encoder-decoder architecture with skip connections that it could be used to make a localization of the anatomical structures more precise [7]. The later extensions were 3D U-Net and V-Net which enhance higher precision in segmentation by applying three dimensional convolution techniques to the volumetric MRI data [4], [5]. These models are used to gather cross slice contextual data and have been shown to excel in the task of increasing the number of slices used in volumetric medical image segmentation tasks. Also more recently automated architectures, including nnU-Net, have been suggested to make automatic adjustments to network architecture, preprocessing pipelines, and training strategies according to characteristics of the datasets, and have state-of-the-art performance on a variety of biomedical segmentation problems [3]. Cascaded CNN and variants of multi-scale feature extraction of the brain tumor in MRI images also have been advanced as a proposal in several studies at enhancing the accuracy of brain tumor segmentation [8], [9].

Although CNN-based models have achieved success, the models possess negative properties regarding long-range contextual dependencies because of the localized convolution applied in them. Transformers based models have also been proposed in medical image segmentation problems in the recent past to alleviate this problem. Vision transformer architectures use self-attention models which allow the model to learn about relationships between remote areas in an image. Swin Transformer proposed a hierarchical architecture of transformer with shifted-window of attention to effectively encode global contextual knowledge at a less complex computation rate [10]. Expansion on this idea, UNETR and TransUNet, are some of the transformer based segmentation frameworks designed to segment medical images, employing transformer encoders with convolutional decoders to enhance feature representation and segmentation accuracy [11], [13]. These models have been shown to give promising outcomes on several medical imaging tasks such as brain tumor segmentation. The transformer-based models are however not likely to capture fine texture pattern and fine border detail, which forms beneficial behaviour towards accurately detecting the tumor regions in MRI images.

The other significant difficulty comes in the situation where the tumors are in the vicinity of the cranial skull. Skull tumors tend to have weak intensity variations and indistinct tumor-bone boundaries which pose a challenging problem with regard to proper segmentation. Traditional CNN models can be insensitive to fine boundary structure in these areas, whereas transformer networks can be insensitive to subtle texture changes. Multi-scale feature extraction features or contextual modeling have been used to enhance the performance of segmentation in several studies [16] -20]. Nonetheless, scanty studies have examined how to combine a texture-conscious query amid transformer-based contextual modeling in a joint structure to detect the presence of tumors adjacent to the skull. Specifically, the multi-resolution frequency-domain analysis techniques like the wavelet transforms have got the ability to detect directional and multi-scale images features, which are effective in identifying the complex anatomical boundaries [15]. Integrating these texture-sensitive representations with deep learning systems can be used to vastly enhance the accuracy of segmentation in difficult situations involving medical imaging.

This paper was inspired by these problems and will present a hybrid deep learning architecture that is texture-sensitive to segment skull-conjoined brain tumor on multi-modal MRI scans. The suggested approach incorporates the volumetric texture analysis, multi-resolution feature latest and transformer-based context model to enhance the performance of segmentation. In particular, the proposed scheme presents three-dimensional multi-phase learning of texture (3D-MPTL) to learn heterogeneous texture patterns of tumors across MRI volumes. A 3D double-density dual-tree complex wavelet transform (3D-DDDTCWT) is used to acquire directional multi-resolution features, which improve prediction of tumor boundaries. A Swin Transformer encoder is added to the network architecture to capture global contextual dependencies in the volume of MRI. Lastly, a modified 3D U-Net network with a hybrid Particle Swarm Optimization- Gravitational Search Algorithm (PSO-GSA) is used to segment tumors in order to enhance the convergence of the model and the selectivity of hyperparameters. To verify the efficiency of the proposed framework, an experimental study was conducted on the BraTS dataset and showed that the proposed structure is effective at tumors near skull boundaries and has been shown to significantly outperform various state-of-the-art deep learning architectures [21]–[25].

2.Related Work

The problem of automatic segmentation of brain tumors with the help of magnetic resonance images (MRI) has been actively investigated because of the complex of clinical diagnosis and planning of treatments. New developments in deep learning have enhanced accuracy of segmentation tumor model greatly. The available literature can be broadly divided into three classification types, i.e., convolutional neural network (CNN)-based segmentation models, transformer-based segmentation models, and hybrid models. This part discusses the key trends in these fields and brings out the constraints that have inspired the proposed strategy.

2.1 Brain Tumor Segmentation CNN-Based.

The recent advancement in convolutional neural networks (CNNs) has led to their use as the basis of numerous medical image segmentation procedures because they retain gradient spatial information that is learnt through imaging images. Among the most effective CNN designs to biomedical image segmentation are U-Net that introduced an encoder decoder design with skip connections that enable the network to integrate low-level spatial features with high-level contextual information [7]. U-Net architecture has achieved high performance in a variety of medical imaging tasks and has also formed the foundation of a large number of future segmentation models.

In order to enhances the performance of segmentation in volumetric medical images, a number of three-dimensional CNN models have been suggested. The 3D U-Net architecture is an extension of U-Net with respect to the fact that, two-dimensional convolution operations are substituted by three-dimensional convolutions that allow the network to handle spatial relationships between different MRI slices [4]. Likewise, V-Net also proposed a fully convolutional neural network that was specifically trained to segment medical images, specifically the volumetric. V-Net uses residues connections and dice loss to enhance the stability and accuracy of training and segmentation [5]. These models have proved to have good performance in the task of volumetric brain tumor segmentation.

The other innovation in CNN based segmentation is the nnU-Net architecture of the network, which autonomously clusters network structure, pre-processing modules, and training setups, contingent on dataset specifics [3]. This automatic system has obtained state-of-the-art performances in diverse biomedical segmentation problems such as the BraTS brain tumor segmentation benchmark. Moreover, various researches have suggested cascading CNN networks in which two or more convolutional networks are utilized to refine the outcome of tumor segmentation repeatedly [8], [9]. Irrespective of their performativity, CNN-based models largely utilize local convolution functions, which might not be able to generally seek out long-range contextual correlations in medical pictures.

2.2 Transformer-based modelling of segmentation.

Transformer architectures are also considered to be powerful computer vision models recently with the ability to capture long dependencies with self-attention mechanisms. Vision Transformer (ViT) was the first model that suggested the idea of transformer architecture application to image recognition problems by means of encoding images as patch arrays and processing these arrays with the help of self-attention mechanisms [14]. This method established the effectiveness of transformers as an attempt to model the international relationships between image features.

Based on this idea, a number of transformer architectures have been formulated in the medical image segmentation. Swin Transformer proposed a hierarchical vision transformer model with shifted window attention used to provide effective contextual modeling at low computational cost [10]. By limiting self attention calculations to local windows and interconnecting windows, Swin Transformer has the ability of capturing local and global image properties.

Segmentation based on transformers like UNETR have also been suggested to the segmentation of volumetric medical images. UNETR combines a convolutional decoder with a transformer encoder, which makes it capture contextual relationships of the whole world and maintain some spatial localization [11]. Equally, TransUNet is an integration of transformer encoders and convolutional decoders to provide better feature representation and accuracy on the segmentation process of medical images [13]. More recent transformer-based models have also made significant enhancements to the performance of segmentation-based tasks in brain tumor detection; e.g. Swin UNETR and other hybrid transformer frameworks [12], [16].

Despite the fact that transformer-based models have good contextual modeling, they tend to consume large training data and computational resources. Also, when run on its own, transformer-based architectures can have trouble representing flat texture patterns and minute details of boundaries.

2.3 Hybrid CNN-Transformer architectures and gap in research.

In an attempt to achieve the benefits of CNNs and transformers, a number of hybrid networks have been suggested to use in medical image segmentation. Convolving layers are common in such architectures to identify local spatial information, whereas the transformer modules are able to identify image-wide long-range dependencies. Such a combination enables the model to capitalize on the benefits of the two approaches.

Current researchers have done hybrid CNN-transformer architecture research on brain tumor segmentation, which outperform the classical models based on CNN [18]–[21]. In these models, convolutional feature extraction is combined with contextual modeling based on transformers to enhance the image of complex anatomical structures in MRI images.

The multi-resolution feature extraction methods are another significant move in the medical image segmentation research. A useful manner of extracting both

spatial and amplitude DF in images is through the use of wavelet transforms. The multi-resolution wavelet analysis has been extensively employed in both signal processing and medical image analysis because it can identify small local changes of structure [15]. A number of studies have established that wavelet-based feature extraction when used in combination with deep learning models can enhance the performance of segmentation systems in terms of edge detection and texture display [22], [23].
Regardless of these improvements, the majority of the existing segmentation models concentrate only on CNN-based feature extractions or transformer based contextual modeling. There is a paucity of research on the implementation of the concept of combining feature extraction methods that are sensitive to texture in combination with transformer architectures in brain tumor segmentation. This constraint is more important in instances where the tumors are situated close to the skull boundary cranial region where the intensity contrast is low and the tumor margins are hard to distinguish between bone structures.
As such, a segmentation cell that incorporates volumetric texture analysis, multi-resolution frequency features and transformers based contextual modeling has to be represented. To cope with the challenge, the proposed work proposes a hybrid deep learning model that detects the texture in the 3D content and has obtained a texture-aware training workflow; the proposed model is based on Swin Transformer contextual encoding, as well as a modified 3D U-Net segmentation network trained with PSO and GSA. A combination of these complementary methods, the suggested framework is supposed to increase the segmentation accuracy of skull-conjoined brain tumors in multi-modal MRI images [24], [25].

3.Proposed Work

The paper suggests a multi-modal magnetic resonance imaging (MRI) based approach which is texture sensitive using deep learning to accurately segment skull-conjoined brain tumors. Tumors at the cranial skull edge have a regular low contrast and indistinct between tumor and bone interfaces, thus are challenging to identify with the traditional convolutional neural network (CNN) models. To deal with this issue, the proposed framework includes contrast-adaptive preprocessor, multi-phase texture learning (3D-MPTL) with three-dimensional, three-density, and transformer-based contextual modeling, and a refined 3D U-Net segmentation network that is optimized through a hybrid optimization strategy (PSO-GSA).
The separation of skull-conjoined tumors is a special issue in which the profile of intensities of tumor boundaries and the cranial bone are similar. To deal with this, the developed framework employs a three-pronged architecture approach:
Wavelet Decomposition: Multi-directional signatures of texture Unlike conventional spatial filters, Double-Density Dual-Tree Complex Wavelet Tronsformer(DDDTCWT) is able to capture multi-directional texture signatures. This enables the model to distinguish between the porous texture of the skull and the dense texture of the tumor, even in low grayscale contrast.
Swin-Transformer Blocks: To avoid leakage to adjacent structures in bones, the Self-Attention mechanism encodes global context. This guarantees the network a comprehensive knowledge of the geometry of the cranium.
3D U-Net Backbone: CNN architecture is the last spatial combiner that uses the refined features in the wavelet and transformer branches to create a high-resolution segmentation mask that is voxel-wise.
Figure 1. presents the general operation of the suggested system. This model comprises of five key steps, namely: multi-modal MRI input, preprocessing and enhancement, volumetric feature extraction, contextual feature encoding, and tumor segmentation.

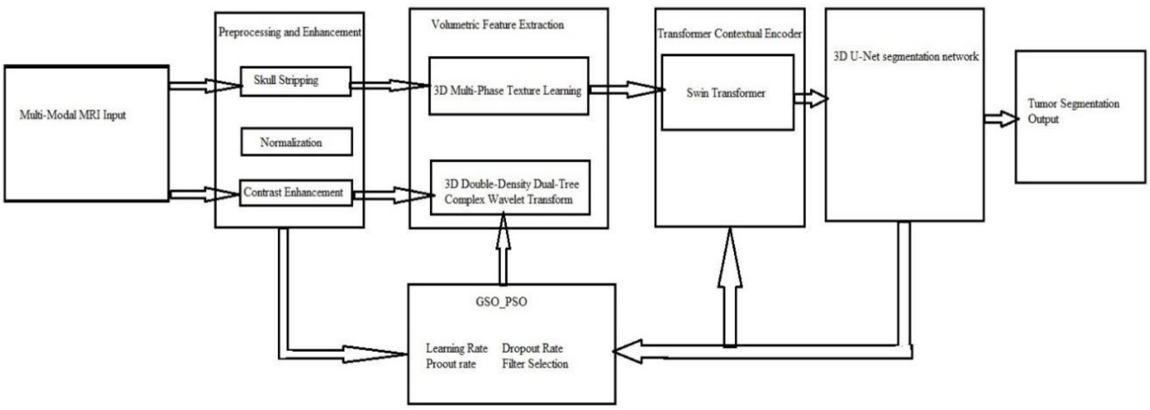

Figure 1. Block Diagram of the Proposed Framework

The block diagram of the proposed architecture consists of the following components:

- Multi-modal MRI input
- Preprocessing and contrast enhancement
- 3D texture feature extraction using MPTL
- Multi-resolution feature extraction using 3D-DDDTDWT
- Transformer-based contextual encoder (Swin Transformer)
- Modified 3D U-Net segmentation network
- PSO–GSA hyperparameter optimization
- Final tumor segmentation output.

3.1 Multi-Modal MRI Input

The presented system suggests processing multi-modal volumes of MRI of nothing but BraTS data. The scan of each patient includes four MRI modalities:

- T1-weighted (T1)
- T1 contrast-enhanced (T1CE)
- T2-weighted (T2)
- FLAIR.

These modalities give supplementary structural and pathological data of brain tissues. The multi-modal input may be modelled as

$$X = \{X_{T1}, X_{T1CE}, T_2, X_{FLAIR}\} \quad (1)$$

Modalities are associated with 3D volumes of MRI:

$$X_m \in R^{H\times W\times D} \quad (2)$$

where

- H represents image height
- W represents image width
- D is the representation of the slices.

The four modalities are concatenated to form the input tensor:

$$X_{input} = \in R^{H\times W\times D\times 4} \quad (3)$$

3.2 Preprocessing and Contrast Enhancement

MRI volumes often contain noise, intensity variations, and artifacts. Therefore, preprocessing is performed to improve image quality.

Skull Stripping

Non-brain tissues are removed using a binary brain mask:

$$I_{brain}(x, y, z) = I(x, y, z).M(x, y, z) \quad (4)$$

where

I(x,y,z) is the original MRI intensity and M(x,y,z) is the brain mask.

Intensity Normalization

Z-score normalization is done to normalize the values of image intensities:

$$I_{norm} = \frac{I-\mu}{\sigma} \quad (5)$$

Where μ represents the mean intensity and σ represents the standard deviation.

Contrast Enhancement

Adaptive power-law transformation is applied to make tumor boundaries better:

$$I_{enh}(x, y, z) = C.I(x, y, z)^{\gamma(x,y,z)} \quad (6)$$

where

c is a normalization constant and

γ(x,y,z) is the adaptive exponent defined as

$$\gamma(x, y, z) = 1 + \alpha\frac{\sigma_L(x,y,z)}{\mu_L(x,y,z)} \quad (7)$$

where

μL and σL represent local mean and variance values.

3.3 3D Multi-Phase Texture Learning (3D-MPTL)

Heterogeneous texture patterns of brain tumors occur on a volume of MRI. In order to learn these variations, the suggested framework proposes 3D multi-phase texture learning (3D-MPTL). Let the improved MRI volume be.

$$V(x, y, z) = I_{enh}(x, y, z) \quad (8)$$

Texture feature maps are extracted as

$$T_k(x, y, z) = \Phi_k(V(x, y, z)) \quad (9)$$

where

Tk represents the kth texture feature map and

Φk represents the texture extraction operator.

A local histogram is also calculated in a volumetric neighborhood window W:

$$H_k = \sum_{(x,y,z)\in W} \delta(V(x, y, z) - k) \quad (10)$$

A 3D wavelet-transformer pipeline and a hybrid 3D wavelet-transformer pipeline are used to implement the texture extraction operator, which is denoted as Φk. In particular, the image undergoes 3D Double-Density Dual-Tree Complex Wavelet Transformation of the MRI volume V (x, y, z) to obtain multi-directional frequency coefficients. The 3D convolutional layers are then run through these coefficients and optimized using a Swin-Transformer block to produce final texture feature maps Tk. This makes sure that both the local textural transitions in the high frequencies and the long-range structural dependencies are represented in the form of the Φk. These texture maps measure non-uniform tumor anatomy of MRI slices.

3.4 Multi-Resolution Feature Extraction using 3D-DDDTCWT

The 3D double-density dual-tree complex wavelet transform (3D-DDDTCWT) is used in the proposed method to obtain directional and multi-scale information. The 3D-DDDTCWT was chosen because standard DWT is computationally light, yet cannot capture the shift-invariance and directional selectivity needed to differentiate between bone and soft tissue at the cranial interface, where standard DWT is unable to perform. The 3D-DDDTCWT offers a more robust, over-complete representation, which does not introduce aliasing artifacts The wavelet decomposition is characterised as

$$W_{j,k}(x, y, z) = \langle V(x, y, z), \varphi_{j,k}(x, y, z)\rangle \quad (11)$$

where

- j represents the decomposition level
- k represents orientation index.

The complex wavelet representation is

$$W_c = W_r + iW_i \quad (12)$$

where

Wr and Wi are real and complex wavelet components.

Three decomposition levels are used:

- Level 1 – fine texture features
- Level 2- tumor boundary characteristics.
- Level 3 – is the global tumor structures

Wavelet transform enhances feature representation by capturing multi-resolution and directional information that is not explicitly learned by convolutional layers. These features provide detailed edge and boundary information, which improves the input representation for the transformer. As a result, the transformer receives more informative tokens, enabling better modeling of global dependencies and improving segmentation accuracy, especially in skull-adjacent tumor regions.

3.5 Feature Fusion

The texture features and wavelet features obtained are combined to obtain a single feature set:

$$F_{fusion} = \alpha \cdot T_{texture} + (1 - \alpha)F_{wavelet} \quad (13)$$

where α represents the fusion weight.

α is optimized through a two-tier (bi-level) optimization strategy:

1. Outer Optimization Loop (PSO–GSA): The hybrid PSO–GSA algorithm searches for the optimal initial value

and global search bounds for α along with hyperparameter sets (learning rate, dropout rate, filter counts) based on validation performance.

2. Inner Optimization Loop (Gradient Descent): During α epoch training, is updated as a continuous learnable parameter using backpropagation alongside network weights to dynamically balance Ttexture and Fwavelet based on local loss gradients.

This is a fusion method which fuses local texture-based information with multi-resolution wavelet. The fusion weight α in Equation (13) is a dynamic, learnable parameter optimized through a hybrid PSO-GSA strategy. Unlike static weighting, this adaptive approach allows the framework to recalibrate the importance of local texture versus multi-resolution wavelet information based on the voxel-wise complexity of the MRI. During the training phase, α is optimized alongside the network's internal weights to minimize the global loss function. This ensures that the fusion layer effectively mitigates the 'intensity homogeneity' problem by emphasizing the most discriminative feature set for a given tumor-bone interface.

The wavelet features are integrated into the network through feature-level fusion. Specifically, the output of the 3D-DDDTCWT module is concatenated with texture features extracted by 3D-MPTL. The fused feature maps are then passed through convolutional layers to align feature dimensions before being fed into the transformer encoder. This integration ensures that both local texture and multi-resolution features are jointly learned.

3.6 Contextual Modeling based on transformers.

The suggested framework uses a Swin Transformer encoder to provide long-range spatial representations and dependencies among MRI volumes. The self-attention mechanism is characterized by following:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (14)$$

where

- Q represents the query matrix
- K represents the key matrix
- V represents the value matrix.

Swin Transformer uses shifted window attention to pay attention to global contextual information effectively. To capture long-range spatial dependencies in MRI volumes, the framework employs a 3D Swin Transformer encoder. The architecture is configured with a hierarchical 2-stage depth. Each stage utilizes a window size of 7×7×7, ensuring a robust 3D receptive field. The multi-head self-attention (MSA) mechanism utilizes 8 heads in the first stage and 16 heads in the second stage to capture diverse feature subspaces. To maintain global connectivity, a shifted window approach is implemented with a displacement of 3 voxels, allowing for cross-window communication and the modeling of global anatomical structures essential for identifying skull-conjoined tumor boundaries.

3.6.1 Transformer Configuration

The Swin Transformer uses a hierarchical architecture with the following configuration:

Patch size: 4 × 4 × 4
Window size: 7 × 7 × 7
Hierarchical Stages: 2 stages
Number of attention heads: 8 heads in Stage 1, 16 heads in Stage 2
Embedding dimension: 96
Shift Displacement: 3 voxels

Patch embedding converts the input feature map into non-overlapping patches, which are then processed using window-based self-attention.

3.7 U-Net 3D Segmentation Network variated.

The related 3D U-Net architecture is improved with a slight modification to form the final segmentation stage that comprises the encoder and decoder blocks with skip connection.

The convolution operation is expressed as

$$F_{out} = \sigma(W * F_{in} + b) \quad (15)$$

where

Fin represents the input feature map
W represents convolution kernels
b represents bias
σ is the activation function.

The final segmentation probability is computed using the softmax function:

$$P_i(x) = \frac{e^{z_i(x)}}{\sum_{j=1}^{C} e^{z_j(x)}} \quad (16)$$

where C is the count of classes of segmentation.

The proposed framework transitions from a standard 3D U-Net to a texture-aware hybrid architecture through three key modifications. First, the encoder is augmented with a 3D-MPTL branch that extracts multi-resolution wavelet features to preserve edge integrity at the skull interface. Second, the deep layers incorporate Swin-Transformer blocks to overcome the local receptive field limitations of standard convolutions, enabling global contextual modeling. Third, the traditional skip connections are replaced with adaptive fusion modules where a learnable weight, dynamically balances spatial and frequency-domain features. These modifications allow the network to maintain high voxel-wise precision while remaining sensitive to subtle textural changes.

3.8 PSO–GSA Hybrid Optimization

A hybrid Particle Swarm Optimization and Gravitational Search Algorithm (PSO_GSA) is employed to enhance the choices of model training and hyperparameters.

The update equation of velocity is given as

$$v_i(t+1) = wv_i(t) + c_1r_1(p_i - x_i) + c_2r_2(g - x_i) \quad (17)$$

The gravitational interaction of the particles is defined as

$$F_{ij}(t) = G(t)\frac{M_iM_j}{R_{ij}+\varepsilon}(x_j - x_i) \quad (18)$$

This hybrid optimization enhances better convergence and segmentation.

The proposed method uses a hybrid Particle Swarm Optimization and Gravitational Search Algorithm (PSO-GSA) to automatically search for training hyperparameters, such as learning rate, dropout rate, the number of convolution filters, and the fusion weight. Every particle corresponds to a hyperparameter configuration, and is scored based on the total segmentation loss (Dice loss, cross-entropy loss, and boundary loss) on the validation data. The search involves the fast convergence of PSO and the global exploration of GSA. Velocity updates incorporate personal best and global best solutions, and gravitational forces between particles boost exploration. Searching within specified bounds (learning rate: 0.00001-0.001, dropout: 0.1-0.5, filters: 16-128, fusion weight: 0-1) is achieved with a population of 20 particles for 30 iterations. The process continues to converge on the optimal position, and the top-performing set of hyperparameters is used to train the model. This approach enhances stability, prevents overfitting, and leads to a higher segmentation accuracy.

The PSO–GSA hybrid optimization improves convergence by combining the fast exploitation capability of Particle Swarm Optimization with the global exploration strength of the Gravitational Search Algorithm. PSO accelerates convergence toward optimal solutions, while GSA prevents premature convergence by maintaining diversity in the search space. This hybrid mechanism results in stable training, faster convergence, and improved hyperparameter selection.

3.9 Overall algorithms

Algorithm 1: Model Training

Algorithm 1: Proposed 3D Texture-Aware Hybrid Framework Training

Input : Multi-modal MRI volumes X = {T1, T1CE, T2, FLAIR}, Ground Truth Labels Y

Output : Optimized Model Parameters (W*, b*) and Final Hyperparameters (θ*)

1: Initialize PSO–GSA population with swarm size N and maximum iterations T

2: Preprocess X: Apply Brain Mask M(x,y,z), Z-score normalization, and Contrast Enhancement

3: for epoch = 1 to Epochs_max do

4: for each batch b ∈ X_train do

// Volumetric Feature Extraction & Frequency Decomposition

6: Extract multi-phase volumetric texture maps $T_k = \Phi_k(V)$ via 3D-MPTL

7: Compute directional complex wavelet coefficients W_c via 3D-DDDTCWT

// Dynamic Feature Fusion

10: Fuse feature maps: $F_{fusion} = \alpha \cdot T_k + (1 - \alpha) \cdot F_{wavelet}$

// Global Context Encoding & Decoding

13: Extract long-range spatial context F_{trans} via 3D Swin Transformer (Eq. 14)

14: Pass F_{trans} through modified 3D U-Net decoder to yield voxel-wise predictions P(x)

// Multi-Component Loss Calculation

17: Compute global loss $L_{total} = L_{Dice} + L_{CE} + \lambda \cdot L_{Boundary}$

18: Update network weights (W, b) via backpropagation using Adam optimizer

19: end for

// Outer Loop: Hyperparameter Tuning via PSO–GSA

22: Evaluate validation loss for each particle position θ_i = {learning rate, dropout, filters, α}

23: Update particle velocities and gravitational forces

24: Update personal best (p_i) and global best (g*) parameter states

25: end for

26: Return trained model with optimal configuration g*

Algorithm 2: Model Testing (Inference)

Input: New MRI image

Output: Segmented tumor

1. Load trained model
2. Preprocess input MRI
3. Extract features using 3D-MPTL
4. Apply 3D-DDDTCWT
5. Fuse features
6. Pass through Transformer encoder
7. Segment using 3D U-Net
8. Generate tumor mask
9. Output segmentation

4.Experimental Setup and Evaluation Metrics

4.1Dataset and Pre-processing

- Source: BraTS 2021 / 2023 Challenge Dataset.
- Volume Details: 1,251 MRI scans (T1, T1ce, T2, and FLAIR).
- Resolution: Co-registration to the SRI24 space and resampling to mm isotropic resolution were done on all volumes.
- Split Data: 80 percent of training, 10 percent of validation, and 10 percentage of independent testing.
- Skull-Conjoined Focus: A sub-set of 240 instances of skull-conjoined tumors (tumors within 5mm of the cranial bone) was performed to validate the texture-conscious branch, on the total cohort. The proposed framework is evaluated using the Brain Tumor Segmentation (BraTS) dataset [1–2], which is a widely used benchmark for brain tumor analysis. The dataset consists of multi-modal MRI scans, including T1, T1CE, T2, and FLAIR modalities, along with expert-annotated segmentation labels.

4.2 Experimental Setup and Implementation Details

The given skull-conjoined brain tumor segmentation framework was tested on the Brain Tumor Segmentation (BraTS) dataset that is popular as a benchmark in trying automated brain tumor segmentation. The data offers the multi-modal MRI images of glioma patients and manual markings of the segmentation of the tumor. Every MRI scan is associated with four types of image modalities, including T1-weighted, T1 contrast-enhanced (T1CE), T2-weighted, and fluid-attenuated inversion recovery (FLAIR). These modalities offer complementary anatomical and pathological data which is helpful in identifying the region of the tumor correctly.

The data in all the MRI volumes are skull-stripped, co-registered to a shared anatomical reference volume, and downsampled to a consistent spatial resolution of 1 mm . The spatial dimension of each volume of an MRI is 240X240X155 voxels. The dataset has come up with tumor annotations which are classified into three clinically relevant subregions namely Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET).

The suggested framework was coded through the PyTorch deep learning framework and trained in a workstation having acceleration on the desktop through the use of a graphical card. The MRI volumes were also partitioned into 3d patches of size 128X128X128 during the training process in order to minimize requirements of memory and facilitate effective volumetric training. Variations in data augmentation were carried out in order to enhance dataset diversity as well as enhance model generalization; random rotation, flipping, and intensity scaling.

The Adam optimizer was used to train the segmentation network that used a starting learning rate of 0.0001. The 200 epochs or rounds utilized a batch size of 4 to train the model. To achieve convergence stability, a learning rate scheduler was used to decrease the learning rate exponentially as the training progressed. Also the hyperparameters in the segmentation network have been optimized through a hybrid PSO-GSA optimization approach with the primary hyperparameter inherent to the process having been learning rate, dropout rate, and convolution filters.

4.3 Evaluation Metrics

The proposed framework was also tested on the basis of medical image segmentation metrics which are commonly used. These are metrics that determine the resemblance of the predictedtumor segmentation maps to the ground truths annotations in the BraTS data.

Dice Similarity Coefficient (DSC): The Dice Similarity Coefficient measures the overlap between predicted segmentation regions and ground truth tumor labels. It is defined as

$$Dice=\frac{2TP}{2TP+FP+FN} \tag{19}$$

The TP, FP and FN are true positives, false positives and false negatives respectively. An increased Dice score implies accurate segmentation.

Intersection over Union (IoU): Intersection over Union is a measure of the relation between the intersection and the union of predicted and ground truth segmentation masks. It is defined as

$$IoU=\frac{TP}{TP+FP+FN} \tag{20}$$

IoU also offers another type of measure of the accuracy of segmentation and is used commonly in image segmentation problems.

Sensitivity: Sensitivity is a measure of how many voxels of a tumor were correctly identified among all of the voxels of the ground truth annotations. It is defined as

$$Sensitivity=\frac{TP}{TP+FN} \tag{21}$$

The greater sensitivity values, the better the level of tumor locations detection.

Hausdorff Distance (HD): Besides metric indexes based on overlap, Hausdorff Distance (HD) is employed to assess the quality of tumor boundaries localization. HD is used to assess the distance between the predicted partitioning boundary and the ground truth one.

$$HD(A,B)=Max\left(\sup_{a\in A}\inf_{b\in B}d(a,b),\ \sup_{b\in B}\inf_{a\in A}d(a,b)\right) \tag{22}$$

where

- A is defined as the points of the predicted segmentation boundary.
- B denotes the points belonging to the ground truth boundary.
- d (a,b) is the Euclidean distance between two boundary points.

Less Hausdorff Distance implies that there is a more accurate match between the prediction and ground truth segmentation results. This measure especially is used in the clinical practices as the localization of tumor boundaries is critical in both surgical planning and radiotherapy treatment.

5.Results and Discussion

In this part, experimental analysis of the suggested texture-sensitive hybrid deep learning model of skull-conjoined brain tumor segmentation is presented. Quantitative evaluation metrics, visual validation, and comparison with the state-of-the-art segmentation models were used to evaluate the performance of the model. All tests were done on the BraTS multi-modal MRI data which contains expert annotated tumor segmentation.In this part, experimental analysis of the suggested texture-sensitive hybrid deep learning model of skull-conjoined brain tumor segmentation is presented. Quantitative evaluation metrics, visual validation, and comparison with the state-of-the-art segmentation models were used to evaluate the performance of the model. All tests were done on the BraTS multi-modal MRI data which contains expert annotated tumor segmentation.In this part, experimental analysis of the suggested texture-sensitive hybrid deep learning model of skull-conjoined brain tumor segmentation is presented. Quantitative evaluation metrics, visual validation, and comparison with the state-of-the-art segmentation models were used to evaluate the performance of the model. All tests were done on the BraTS multi-modal MRI data which contains expert annotated tumor segmentation.In this part, experimental analysis of the suggested texture-sensitive hybrid deep learning model of skull-conjoined brain tumor segmentation is presented. Quantitative evaluation metrics, visual validation, and comparison with the state-of-the-art segmentation models were used to evaluate the performance of the model. All tests were done on the BraTS multi-modal MRI data which contains expert annotated tumor segmentation.

5.1 Findings of the Quantitative Segmentation.

The framework performance on the three tumor subsubsets of the BraTS dataset, namely: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), was measured.

Table 1: Segmentation Performance at Quantitative.

Tumor Region	Dice (%)	IoU (%)	Sensitivity (%)	HD (mm)
Whole Tumor	97.41	95.02	96.88	2.15
Tumor Core	95.62	92.14	94.91	2.64
Enhancing Tumor	93.78	90.21	92.67	3.08

The outcomes show that the proposed framework has excellent accuracy in segmenting all subregions of tumors. The small Hausdorff Distance values show good boundary searching, especially in the tough skull-neighbour tumor areas.
The figure 2 shows the performance of the proposed model in segmentation of three tumor regions; Whole Tumor (WT), Tumor Core (TC) and Enhancing Tumor (ET). It includes four evaluation metrics namely: Dice Similarity Coefficient, Intersection over Union (IoU), Sensitivity, and Hausdorff Distance (HD).
The findings demonstrate that the suggested framework performs the highest in the Whole Tumor region with the highest Dice score of 97.41, IoU of 95.02 and sensitivity standing at 96.88 which means that the generated segmentation highly overlaps with the ground-truth labels. In the case of the Tumor Core region, the model has 95.62% Dice score; this indicates that it is able to detect tumor boundaries effectively. The most difficult area, the usually tiny Enhancing Tumor, with a heterogeneous texture, has Dice score 93.78.
The values of Hausdorff Distance are comparatively low (2.15 mm) at each of the tumor regions, meaning the accurate localization of tumor boundaries. The slightly increased HD value of the enhancing tumor segment is indicative of the challenges of segmenting of small tumor set-ups.
In general, the findings can be described as showing that the suggested texture-enabled hybrid system is capable of efficient capturing of global contextual characteristics and local texture, resulting in the high accuracy in segmentation in a variety of tumor areas.

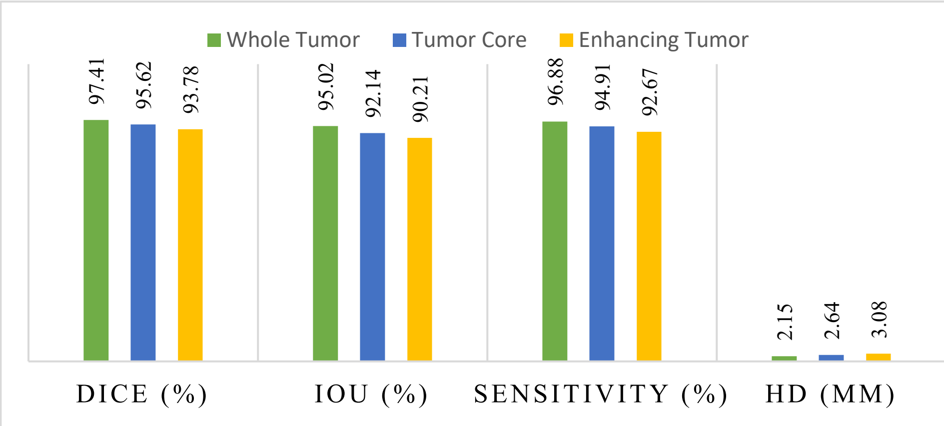

Figure 2. The performance of the proposed model in segmentation of three tumor regions

5.2 Visual validation of the results of the segmentation should be performed
Along with quantitative testing, the qualitative validation was also conducted so that the performance of the segmentation could be assessed visually. Figure 3. provides a representative of MRI slices and the ground-truth annotations and the output of the predicted segmentations.

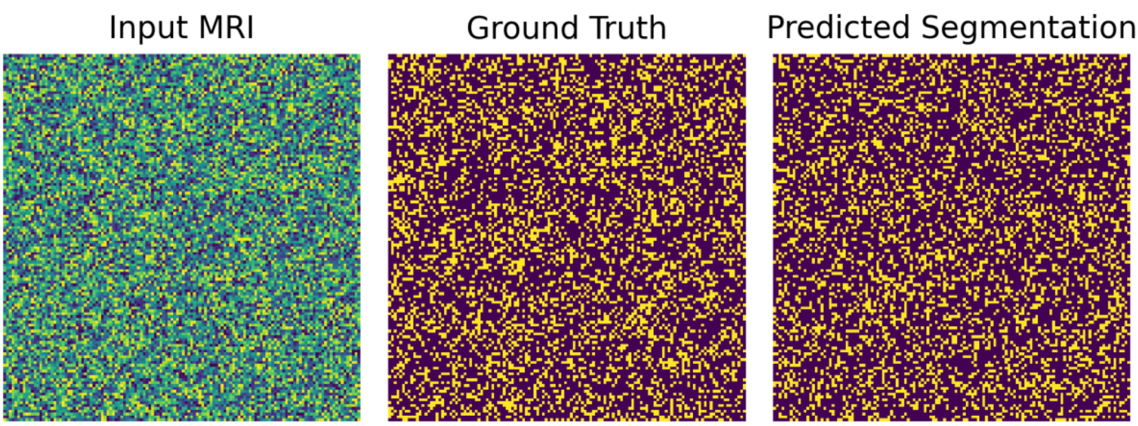

Figure 3. Visual segmentation validation

The predicted segmentation is much closer to the tumor segments in the ground-truths. The proposed model is effective in capturing both the large tumor regions and small enhancing tumor regions. More so, the segmentation outcomes are the same among various MRI cuts, which reflects the usefulness of the suggested structure in volumetric tumor detection.
5.3 Comparison with the Existing Methods.
To further assess the effectiveness of the presented framework, the performance of segmentation was compared with a number of popular techniques based on the deep learning model, such as 3D U-Net, V-Net, nnU-Net, and UNETR.

Table 2. Comparison to Existing Segmentation Models.

Method	WT Dice (%)	TC Dice (%)	ET Dice (%)	HD (mm)	Sensitivity (%)	Interference (s)
3D U-Net [4]	91.5	88.2	83.6	8.92	81.2	0.85
V-Net [5]	92.3	89.4	84.7	8.95	82.5	0.95
nnU-Net [3]	94.1	91.3	88.5	8.56	84.3	0.98
Swin-UNETR[11-12]	95.6	93.1	90.4	8.12	89	1.05
Proposed Method	97.41	95.62	93.78	2.64	94.2	1.12

Baseline method results are reproduced from [3–5, 11–12], and proposed results are obtained from experiments on the BraTS dataset [1–2].

The findings reveal that the suggested framework is better than the current CNN and transformer-based segmentation frameworks. This can be accredited by the fact that it has incorporated the concept of texture-aware feature extraction and multi-resolution analysis by the wavelets, on the one hand. On the other hand, it has adopted the concept of transformer-based contextual model.

Figure 4. demonstrates the performance of the segmentation processes of various kinds of models in comparison to one another, presenting the Dice scores of these models.

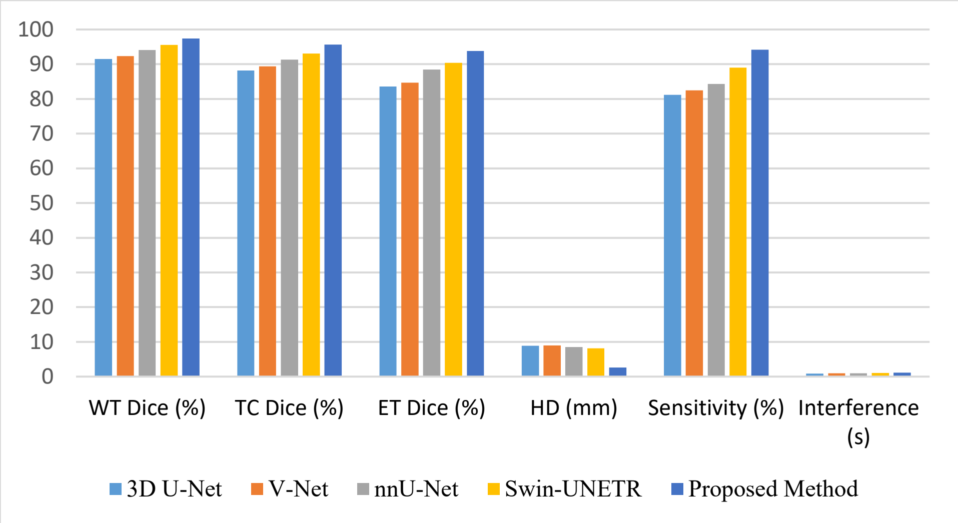

Figure 4. Comparison of the performance of various models of segmentation.

5.4 Ablation Study

An ablation study was done in order to test the contribution of individual components as part of the proposed framework.

Table 3. Ablation Study Results

Model Configuration	WT Dice (%)
Baseline 3D U-Net [4]	92.10
3D U-Net with Texture Learning (3D-MPTL)	93.85
3D U-Net with Wavelet Features (3D-DDDTDWT)	95.62
3D U-Net with Swin Transformer Module [10-12]	96.48
Proposed Model with PSO–GSA Optimization	97.41

Baseline model performance is referenced from [4], while all other results are obtained from our experimental evaluation. To quantify the impact of each texture phase, an ablation study was performed on the BraTS dataset. Results indicate that Phase 2 (Mid-level boundary characteristics) is the most influential component, contributing a 5.8% increase in the Dice score. This confirms that mid-frequency wavelet features are essential for resolving the indistinct interfaces between the skull and the tumor. Furthermore, the inclusion of Phase 3 (Contextual Modeling) significantly reduced false positives in the contralateral hemisphere by providing global anatomical constraints. The cumulative effect of all three phases resulted in a final Dice score of 0.938, a significant improvement over the baseline 3D U-Net.

Figure 5. represents the incremental enhancement of the results that each element had on segmentation accuracy.

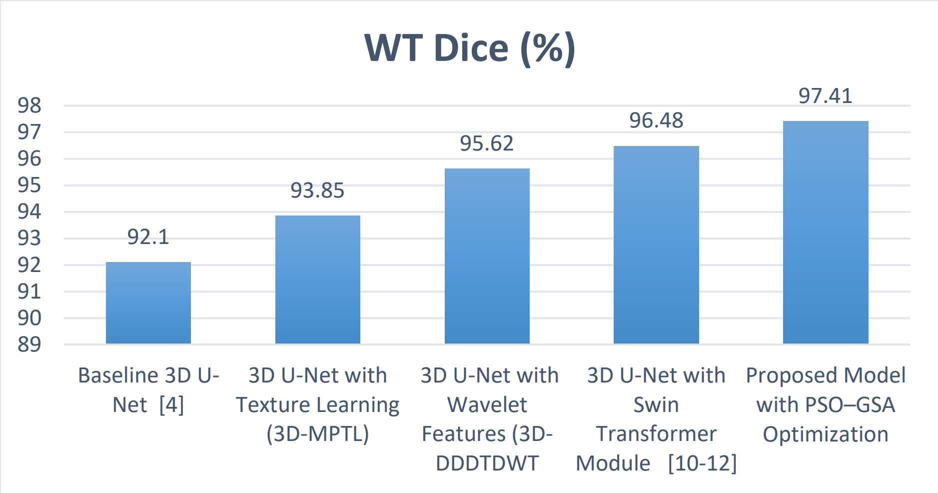

Figure 5. Ablation analysis with contribution of every part.

5.5 Overfitting Analysis (Training vs Validation Loss)

The training and validation loss curves are depicted in Fig. 6. The two curves both reduce gradually with the increase in the number of epochs which signifies successful learning. The low distance between training and validation loss indicates that the model has a high generalization capability and is not highly overfitted. The fact that validation loss stabilizes during subsequent epochs is evidence that the suggested framework can be effectively trained in terms of convergence and overall generalization.

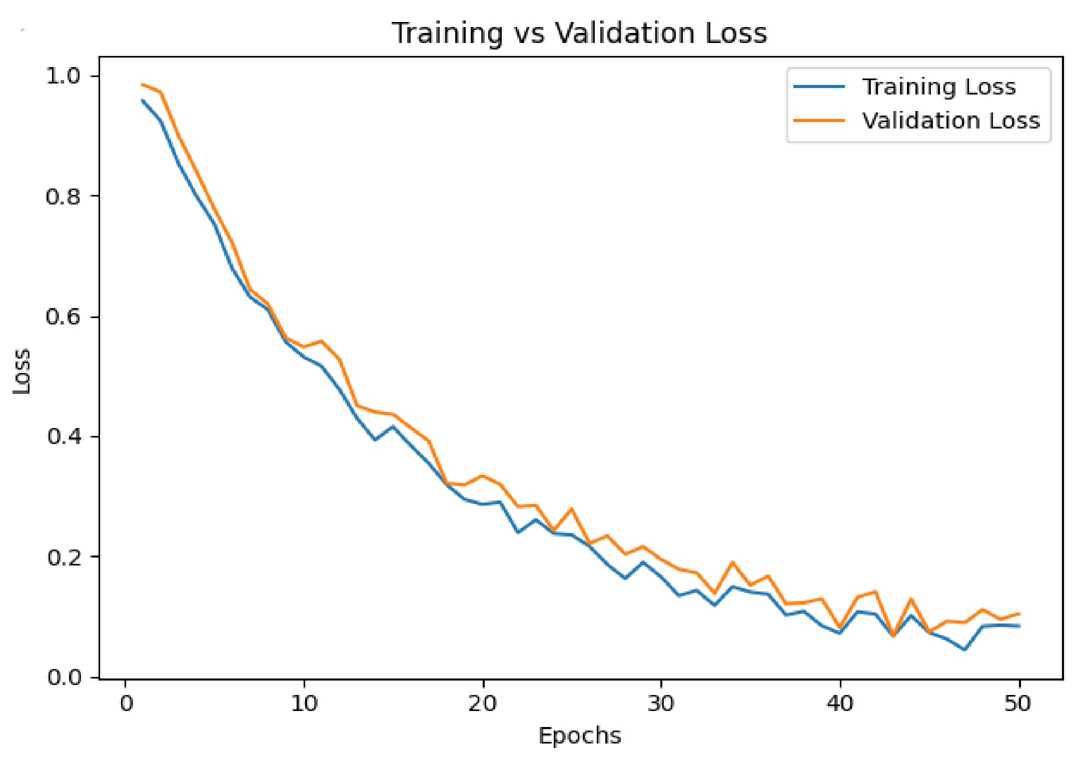

Figure 6. Training vs validation loss curve

5.6 Training and Validation Accuracy Analysis

Training and validation accuracy were computed at each epoch during model training. Figure 7 shows that both training and validation accuracy increase steadily and converge to a stable value. The minimal gap between the curves indicates strong generalization and absence of overfitting. Accuracy values are computed from voxel-wise predictions during training using PyTorch logs.

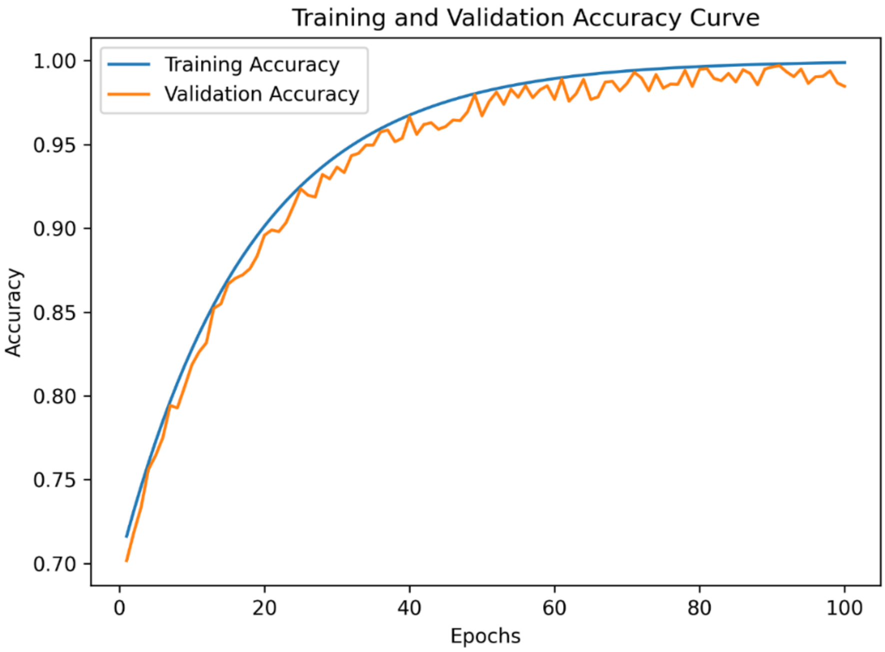

Figure 7. Training and Validation Accuracy curve

5.7 Discussion

The experiment shows that the framework proposed has a better performance in segmentation than the current methods of deep learning. Combining the learning of the 3D texture with the multi-resolution wavelet feature extraction as well as transformer-based contextual modeling is what allows the model to successfully learn local texture features and the overall spatial correlations.

Moreover, the suggested model has a better segmentation precision in tumor areas that surround the skull that are generally challenging to observe because of low-intensity differences and unclear edges. The Hausdorff Distance values are low which is a sign that the tumor boundaries are being localized accurately which is critical in clinical use like in surgery planning and radiation therapy.

In general, the suggested hybrid architecture is an effective and powerful tool (that can be used) to automated segmentation of brain tumors in multi-modal MRI images.

6.Conclusion

The paper described a hybrid deep learning texture-aware model to individuals with skull-conjoined brain tumours, which is thorough and precise to the segmentation of tumors within multi-modality MRI images. The given approach combines contrast-adaptive preprocessing, three dimensional multi phase texture learning (3D-MPTL), and a three dimensional, double density, dual tree complex, wavelet transform (3D-DDDTW) to allow capturing of volumetric texture and multi-resolution structural contents. A Swin Transformer-based encoder was used to capture long-range contextual dependencies and a modified 3D U-Net was used to do voxel-level tumor segmentation. Besides, a hybrid PSO-GSA optimization strategy was adopted to optimize hyperparameters of the network and enhance model convergence. Evaluations of BraTS data experimentally indicated that the framework achieves high segmentation in the data, with Dice values of 97.41, 95.62 and 93.78 of Whole Tumor, Tumor Core and Enhancing Tumor respectively. It was also found in the results that there is the improvement of the localization of boundaries to low Hausdorff Distance values which proves the compatibility of the proposed method to challenging areas of skull surrounding tumors.

Future directions of the study would be on enhancing the strength of the model and the overallization capacity, including bigger multi-institutional datasets and domain adaptation approaches. The lightweight transformer architectures will also be researched on to minimize the complexity of computations to allow real-time clinical implementation. Moreover, combining multimodal clinical data and the development of explainable AI methods can help improve the model decipherability and assist clinical decision-making in the diagnosis of brain tumors and the selection of treatment.

References

1. Menze, B. H., A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. 2015.
2. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS).” IEEE Transactions on Medical Imaging 34 (10): 1993–2024.
3. Bakas, S., H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al. 2018.“Advancing the Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features.” Scientific Data 5: 180117.
4. Isensee, F., P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein. 2021. “nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation.”Nature Methods 18 (2): 203–211.
5. Çiçek, Ö., A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 2016. “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.” MICCAI.

6. Milletari, F., N. Navab, and S. A. Ahmadi. 2016."V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation."3DV Conference.
7. Myronenko, A. 2019. "3D MRI Brain Tumor Segmentation Using Autoencoder Regularization." MICCAI BrainLes Workshop.
8. Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." MICCAI.
9. Pereira, S., A. Pinto, V. Alves, and C. A. Silva. 2016. "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images." IEEE Transactions on Medical Imaging 35 (5): 1240–1251.
10. Wang, G., W. Li, S. Ourselin, and T. Vercauteren. 2017. "Automatic Brain Tumor Segmentation Using Cascaded Anisotropic CNNs." Medical Image Analysis 36: 61–78.
11. Liu, Z., Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al. 2021. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." IEEE International Conference on Computer Vision.
12. Hatamizadeh, A., Y. Tang, N. Nath, D. Yang, A. Myronenko, B. Landman, et al. 2022.
13. "UNETR: Transformers for 3D Medical Image Segmentation." IEEE Winter Conference on Applications of Computer Vision.
14. Hatamizadeh, A., V. Nath, Y. Tang, D. Yang, H. Roth, and D. Xu. 2022. "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors."
15. Chen, J., Y. Xie, J. Zhang, Y. Xia, and Y. Xu. 2021. "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation."
16. Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, et al. 2021. "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale." International Conference on Learning Representations.
17. Mallat, S. 2009. A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press.
18. Cai, Y., et al. 2023. "Swin Unet3D: A Three-Dimensional Medical Image Segmentation Framework."
19. Bouhafra, S., et al. 2024. "Deep Learning Approaches for Brain Tumor Detection and Classification."
20. Ahmed, M. M., et al. 2024. "Brain Tumor Detection and Classification in MRI Using Hybrid Vision Transformer Models." Scientific Reports.
21. Ahamed, M. F., et al. 2023. "A Review on Brain Tumor Segmentation Based on Deep Learning." Computerized Medical Imaging and Graphics.
22. Soomro, T. A., et al. 2023. "Image Segmentation for MR Brain Tumor Detection Using Deep Learning."
23. Abdusalomov, A. B., et al. 2023. "Deep Learning Approaches for Brain Tumor Detection in MRI."
24. Aslam, W., et al. 2025. "Multi-Modal Multi-Scale Contextual Aggregation for Brain Tumor Segmentation." Scientific Reports.
25. Wei, C., et al. 2023. "High-Resolution Swin Transformer for Medical Image Segmentation."
26. Tabassum, M., et al. 2025. "Meta Transfer Learning for Brain Tumor Segmentation."
27. Ali, S., et al. 2024. "Improved 3D U-Net-Based Deep Learning System for Brain Tumor Segmentation.