



A Vision Transformer-Enhanced ResU-Net50 Model for Accurate MRI Brain Tumor Segmentation Using DBO Optimization

N. M. Ramalingeswararao¹ and G. Nagaraju²

ABSTRACT

For precise evaluation and therapeutic planning, automatic brain cancer segmentation from MRI scans is essential. In this work, a hybrid deep learning framework for image segmentation comprising a Vision transformer and ResU-Net50 is proposed. It combines the strong local feature extraction capabilities of ResNet50 with the global representation learning capabilities of a Vision Transformer embedded in a U-Net framework. To improve mask quality and prediction stability, a preprocessing pipeline powered by Dung Beetle Optimization (DBO) is used to optimize thresholding, morphological enhancement, and test-time augmentation. Comparing DBO-based preprocessing to non-optimized preprocessing, experimental results on a T1-CE MRI dataset with 3064 annotated images show that the former greatly enhances segmentation performance, resulting in increases of 2.3% in precision, 1.9% in Dice Coefficient, and 3.2% in Jaccard Index. With Dice scores of 0.9630, 0.9873, and 0.9862 for pituitary, glioma, and meningioma tumors, respectively, the suggested Vision-ResU-Net50 model further achieves exceptional tumor-wise segmentation results, outperforming cutting-edge techniques like Edge U-Net, Improved U-Net, CNN, and deep learning models by margins of 4–17% in Dice and 5–15% in Jaccard Index, the suggested approach is a highly efficient and dependable for clinical brain tumor analysis.

Article information:

Keywords: ResNet50, Vision Transformer, U-Net framework, Dung Beetle Optimization, Dice Coefficient, Jaccard Index

Article history:

Received: December 23, 2025

Revised: March 22, 2026

Accepted: May 21, 2026

Published: June 13, 2026

(Online)

DOI: 10.37936/ecti-cit.2026203.265404

1. INTRODUCTION

Brain tumors pose a serious risk to human health since they can be either benign or malignant, starting in the brain tissue or spreading to other regions of the body. They develop quickly and show a variety of clinical symptoms, frequently resulting in serious neurological disorders that impair the quality of life of patients and even put their lives in jeopardy. Early and precise detection and therapy are essential to enhance prognostic and therapeutic success because of the heterogeneity, deep placement, and complicated anatomic structures of brain tumors [1-2]. Deep learning techniques are extensively used in the analysis and diagnosis of clinical images, leveraging their strong feature-extraction capabilities to automate brain cancer segmentation and diagnosis. The

large amount of MRI data helps the deep learning techniques to learn the complex features associated with brain cancer, making the analysis and diagnosis process precise and efficient. The techniques have shown significant success in the automated diagnosis and segmentation of brain tumors using the U-Net model, indicating strong feature-extraction capabilities compared to other methods and making them reliable and efficient for the analysis and diagnosis of brain cancer [3, 4].

Recent advancements in image segmentation have been achieved through the application of deep learning techniques, specifically the application of Convolutional Neural Network (CNN) and Transformer techniques. The Transformer model is highly efficient in the analysis and determination of global image features. In contrast, the CNN model is advantageous

¹The author is with the Research Scholar of JNTUK, Kakinada from SRKR Engineering College research center, Andhra Pradesh, India, Email: muralinakkina@gmail.com

²The author is with the SRKR Engineering College, Andhra Pradesh, India, Email: gnagaraju@srkrec.edu.in

¹Corresponding author: muralinakkina@gmail.com

in the determination and analysis of local image features, making the model efficient in the extraction of spatial features. However, the techniques have shown some limitations, as the CNN model might ignore the potential global spatial patterns. In contrast, the application of the Transformer model is restricted by the large computational requirements and the need to develop large datasets [5, 6].

In the domain of computer vision, a Mamba Network—a new deep learning approach—has been reported to have achieved significant success. Its application to various image segmentation, object detection, and image classification tasks has shown significant performance improvements due to efficient feature extraction and modularity. The Mamba Network outperforms existing CNN-based architectures in terms of multi-scale feature fusion and contextual information collection. Moreover, it significantly reduces computational complexity compared to existing Transformer-based architectures. Medical images, especially MRI brain tumor images, have complex textures and structures with high internal heterogeneity and irregularities. These complex structures pose challenges for the Mamba Network in precisely segmenting and delineating brain tumor boundaries due to variations in tumor size, shape, position, and contrast with respect to neighboring tissues [7-8]. Despite having advantages over existing architectures in terms of computational complexity and sequence modeling, its application to medical image segmentation tasks is still in its early stages. In contrast, Vision Transformers have been thoroughly validated for medical image analysis tasks and has shown to have robust global context modeling capabilities using self-attention mechanisms. This work uses Vision Transformer for its stability, seamless integration with CNN-based architectures, and its effectiveness in modeling long-range spatial dependencies in medical images.

ResNet50 is one of the U-Net variants, which is highly effective for feature extraction tasks and enables the construction of deeper networks without facing vanishing gradients issues [9]. This work uses Vision ResNet50 for its feature extraction capabilities and proposes Vision ResU-Net50 as a hybrid model for brain tumor image segmentation tasks. This article presents a detailed analysis of Vision ResU-Net50 for brain tumor image segmentation tasks. It also describes the Architectural details of U-Net, Vision Transformer, and ResNet50 variants, along with their advantages and disadvantages for medical image analysis tasks.

The fundamental hypothesis of this research is that effectively incorporating Vision Transformer and ResNet50 under a U-Net structure, improves segmentation performance compared to existing architectures, including those that utilize the relatively recent Mamba architectures. This study proposes that,

through the use of both convolutional network capabilities and transformer network self-attention, more accurate boundary detection, increased Dice similarity, and better generalization capabilities will be obtained, especially with regard to diverse tumor types. The primary contributions of this work are:

1. The development and presentation of a novel framework, referred to as Vision-ResU-Net50, which has been utilized effectively in brain tumor segmentation, incorporates U-Net, Vision Transformer, and ResNet50 architectures into a unified structure. This study proposes that this framework will allow more accurate and effective segmentation of tumor tissue, leveraging the capabilities of all incorporated architectures.
2. The development and presentation of novel pre-processing and post-processing techniques, including the utilization of the Dung Beetle Optimizer.

2. LITERATURE REVIEW

2.1 Evolution of the U-Net model

The original U-Net framework has also undergone improvements through a few recent updates, such as those proposed in [10], which are specifically designed to address certain problems associated with brain tumor segmentation. Further improvements to the U-Net-based models can also be incorporated through the use of attention mechanisms, as proposed in [11], to improve accuracy and robustness through more efficient use of resources. Similarly, the use of ResNet and DenseNet architectures helps to improve the use and development of more complex networks, thereby allowing for a more efficient use of gradient flow, as proposed in [13]. Further, transfer learning-based approaches, such as those proposed in [14], also show promise in improving segmentation accuracy, especially when applied in scenarios with a lack of medical image resources. However, despite such improvements, a few problems persist, such as those related to dataset imbalance, noisy and anomalous data, and tumor characteristics, which often prove to be a hindrance during training and testing and when applying such models to new patient data. However, several recent approaches have improved segmentation performance, such as Edge U-Net, which uses edge detection to improve segmentation accuracy, and ensemble-based approaches, which utilize multimodal data to improve accuracy. Recent studies have also focused on developing even more complex networks, such as ResU-Net, which uses a combination of residual structures and the original U-Net framework, allowing for even more precise feature extraction and overcoming gradient vanishing problems [17].

Despite such improvements, the original and recent versions of the U-Net framework also suffer from a few limitations, such as those related to its ability

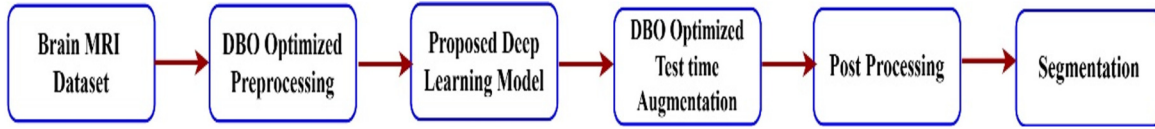


Fig.1: Overview of Dog ID using Contrastive Learning technique with binary cross-entropy loss function.

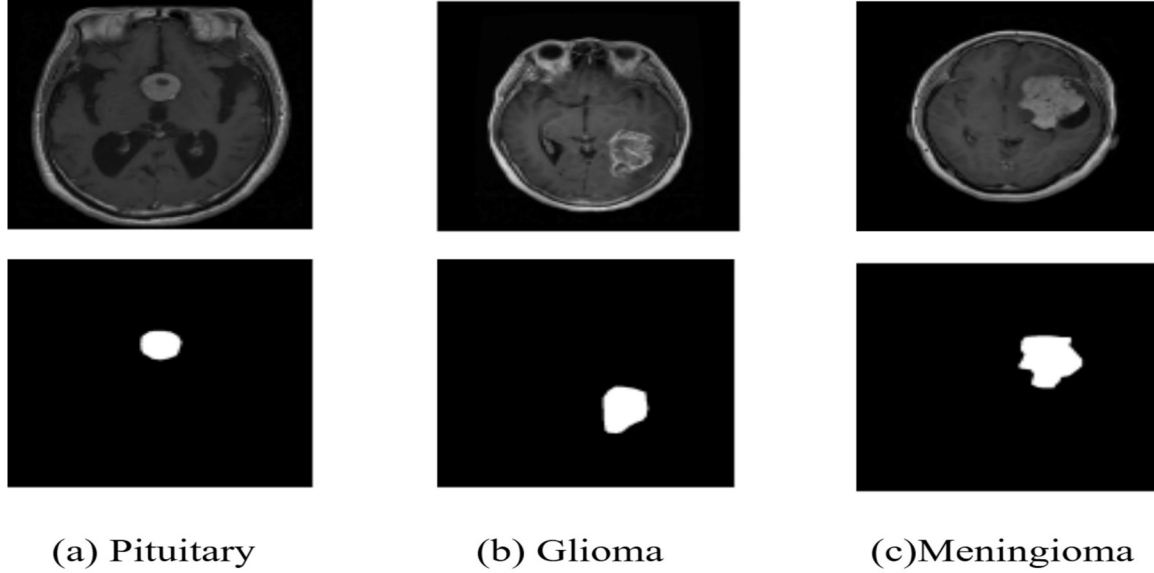


Fig.2: Brain MRI scans of different tumors associated with a mask.

to generalize and segment different types of tumors, insufficient ability to handle smaller tumor regions, and its ability to handle dataset heterogeneity, and its ability to improve and preserve segmentation accuracy, especially when faced with different contrast and noisy conditions, even when using approaches such as those proposed in Edge U-Net and ensemble-based approaches, which often prove to be computationally expensive and inefficient, especially when faced with heterogeneous tumor conditions. A more effective and flexible solution is required to address these issues. In order to improve segmentation efficiency, feature extraction, and resilience across various patient groups, Vision-ResU-Net50 is suggested. This hybrid architecture incorporates Vision transformer and ResNet50 into the U-Net architecture. The vanishing gradient issue that is typical of deeper networks is lessened by ResNet50's residual connections, which enhance the flow of gradients and deep feature learning. The enhanced ability to retain spatial features results from the skip connections, which are optimized to allow the delineation of the boundaries of the tumor formations with greater precision. The vision transformer improves the ability to extract global features. Therefore, the integration of the model ensures greater accuracy and precision in the segmentation of small and asymmetrical tumor formations with minimal over-segmentation inaccuracies.

3. RELATED METHODS AND MATERIALS

This section describes the procedures employed in this study, including the preparation of the Brain MRI database, the proposed U-Net structure with a hybrid Vision ResNet50 encoder, and the evaluation criteria used to assess the segmentation efficiency. Figure 1 depicts the comprehensive workflow of the suggested technique.

3.1 Brain MRI Dataset

This study uses a customized dataset from Kaggle, taken from the T1-CE MRI database, containing 3064 MRI scans of the brain [18], to train the proposed tumor segmentation algorithm. The images are of high quality, with an image size of 512×512 pixels, along with a 24-bit color space to capture significant visual details for tumor investigation. The incorporation of precise real-life tumor masks for every image is a key aspect of the suggested methodology, as it is vital to the proposed supervised training strategy. Three distinct tumor classes are evident in the images, namely pituitary, glioma, and meningioma, with 940, 1430, and 710 instances, respectively. This classification is a step that increases the flexibility and efficacy of the proposed model while also highlighting the diversity that exists among different brain tumors. Figure 2 illustrates the different MRI images and their corresponding tumor masks, highlighting the complexity

and diversity of the dataset, as well as the different challenges that are associated with each type of tumor.

Pituitary tumors, for instance, are usually benign and take the form of adenomas, causing hormonal imbalances that may lead to metabolic syndrome, such as Cushing's syndrome and GH deficiency, and may also cause macroadenoma, which may compress the optic chiasm and cause blindness [19]. Glioma tumors range from benign to highly malignant, including multifocal glioblastoma, characterized by infiltrative behavior. Glioblastoma, or GBM, is a highly malignant brain tumor that is associated with a poor prognosis, with a median survival rate of less than 15 months, and whose boundaries require careful delineation for appropriate treatment [20]. Meningioma is a generally benign tumor that may also be malignant or grade II or III, and may cause visual and neurological disturbances through mass effect [21]. The level of clinical seriousness presented by these tumors calls for advanced image segmentation techniques, such as structural characterization for meningioma, delineation of tumor margins for glioma, and localization for pituitary tumors.

3.2 Preprocessing

The preprocessing pipeline follows a structured series of operations, ensuring consistency and optimal performance in segmentation tasks. The proposed workflow is as follows:

Step 1 (Image Resizing): The MRI images are resized to 256×256 pixels, ensuring consistency in input image size and minimizing computation cost.

Step 2 (Grayscale Conversion): The images are converted from 24-bit color mode to 8-bit grayscale mode, allowing us to focus more on the intensity variations, which are more relevant for tumor detection than color variations.

Step 3 (DBO-Based Threshold Optimization): The optimal threshold is determined using Dung Beetle Optimization (DBO) to ensure optimal accuracy in segmentation results.

Step 4 (Morphological Optimization Using DBO): The morphological operations, including dilation, erosion, opening, and closing, are optimized using DBO.

Step 5 (Test-Time Augmentation (TTA)): The optimal augmentation strategy is determined using the DBO framework, including rotation and flipping.

The initial preprocessing step involves reducing the resolution of the input images to 256×256 pixels. The reduction in resolution normalizes the data, reduces memory consumption, and optimizes computation cost. It strikes an optimal balance between providing enough detail for detailed analysis and keeping the resolution low enough for optimal computation. The second preprocessing step involves converting 24-bit color mode images to 8-bit grayscale mode. The

proposed framework will be able to process images more efficiently, focusing more on variations in intensity, which are more relevant in tumor detection than color.

3.2.1 Dung Beetle Optimization (DBO) Algorithm

Dung Beetle Optimization Algorithm is a nature-inspired metaheuristic algorithm, and its inspiration comes from the behavior and navigation patterns of dung beetles. This algorithm has been able to successfully mimic different behavioral patterns, such as rolling, dancing, foraging, and stealing, to explore and exploit the search space effectively. The dung beetles represent different candidate solutions, and each dung beetle's location represents a potential solution to the optimization problem.

The location of each dung beetle is updated based on environmental cues and different strategies of interaction, and its update is described by a general formula specified in Equation (1).

$$X_i^{t+1} = X_i^t + \alpha (X_{best} - X_i^t) + \beta * R \quad (1)$$

where X_i^t represents the current location, X_{best} represents the best solution obtained so far, and R represents a random vector, while α and β are control parameters. Dung Beetle Optimization Algorithm successfully optimizes different parameters, such as threshold, morphological, and test time augmentation, by maximizing the Dice similarity coefficient between predicted and ground-truth masks.

DBO Optimization Procedure

1. Initialize the population of Dung Beetles with random candidate solutions
2. Fitness evaluation using the Dice Coefficient
3. Best solution identified
4. Position update using exploration and exploitation techniques
5. Boundary constraint implementation
6. End of algorithm: maximum iterations reached or convergence obtained
7. Optimal parameters output: threshold, morphological operation, TTA configuration

Thresholding is an essential stage in the segmentation process, as it directly influences the transformation of the estimated probability maps into binary tumor masks. In the literature, the most commonly adopted approach is the utilization of a constant threshold value. In practice, the utilization of a constant value does not take into consideration the differences in contrast among the MRI images, which directly influences the segmentation performance. Instead of using a constant value, the Dung Beetle Optimization algorithm is used to determine the optimal binary value for the segmentation process. In the optimization algorithm, the Dung Beetle represents the possible range of the binary value, which can vary from 0.1 to 0.9. In contrast, the optimal value is de-

terminated based on the maximization of the Dice Coefficient in relation to the anticipated and ground truth masks, according to Equation (2).

$$C_{Dice} = -\frac{2\Sigma z_{true}z_{predicted} + \delta}{\Sigma z_{true} + \Sigma z_{predicted} + \delta} \quad (2)$$

Here, the z_{true} represents the ground truth mask, and the $z_{predicted}$ represents the binarized predicted mask obtained using the model. The symbol δ represents a small value added to prevent division by zero. The Dung Beetle Optimization (DBO) algorithm finds the optimal threshold (Th) that maximizes the dice similarity value. This ensures the accuracy of the segmentation process while avoiding false positives and false negatives. After the threshold optimization process, the segmentation masks may still contain some minor noise artifacts and may not be entirely accurate with regard to the tumor region and its shapes. To improve the accuracy of the segmentation masks, morphological processing is performed. This process applies various morphological transformations, including dilation, erosion, opening, and closing to improve the accuracy of the segmentation masks. However, the selection of the optimal morphological transformation and the corresponding kernel size has traditionally posed a complex problem and required proper customization of the process. Therefore, the DBO algorithm automatically determines the optimal morphological transformation and the required kernel size. In the DBO algorithm, the dung beetle represents a unique combination of the operation type (o) and the kernel size (k) and can be of various types, such as opening to remove noise artifacts, closing to remove gaps in the tumor region, and dilation to expand the tumor region. The fitness function uses the dice similarity value to determine the most suitable combinations of the dung beetle and select the optimal values. It is important to emphasize here that the entire optimization process performed using the DBO algorithm is carried out only on the training and validation datasets. The optimal parameters obtained using the optimization process are then used to evaluate the model on the test dataset. During the testing process, the model operates independently without the ground truth masks and therefore ensures a realistic scenario with no risk of data leakage.

Test-Time Augmentation considerably improves segmentation reliability. Based on image manipulations like flipping, rotating, and scaling, models may generate conflicting predictions. DBO is used to optimize TTA techniques in order to mitigate this problem and guarantee that the optimal collection of enhancements is applied while testing. Every dung beetle in DBO symbolizes a different set of enhancement parameters, such as rotations (0° , 90° , 180° , and 270°) and vertical and horizontal flips. The ultimate prediction uses the average of the most effective enhanced outputs after the fitness function as-

sesses the precision of segmentation for each enhancement scenario. The segmentation consistency across MRI images has been significantly increased by this method.

3.3 Proposed Hybrid Deep Learning Model

The proposed hybrid model integrates the vision transformer into ResU-Net50 in order to improve the segmentation accuracy. The architectural details of the proposed model are illustrated in Figure 3. In order to solve the degradation issue that arises when training deeper networks, ResNet50, a variation of the Residual Network architecture, incorporates residual learning [23]. ResNet50 mitigates diminishing gradients by using shortcut connections that bypass one or more layers, enabling the training of significantly deeper networks. ResNet50 allows gradients to propagate over several layers when used as an encoder within the U-Net architecture, improving feature extraction without appreciably raising the computational load. The extensive contextual data and minute details included in the deep features that the ResNet50 encoder extracts are essential for distinguishing between malignant and normal cells in medical images.

In the proposed model, a Vision-ResNet50 has been used to improve the encoder path. In contrast to standard U-Net architectures, both the encoding and decoding paths incorporate changes to the convolution blocks. As depicted in Figure 3, each convolutional block in the ResNet50 encoder network follows a “ $2\times$ design”. In this context, the number of convolutional layers in each block doubles. Such a modification ensures that deeper features can be extracted during each stage, thereby allowing the network to effectively identify more complex spatial features and tumor characteristics without increasing the computational complexities associated with the network. The proposed network maintains the conventional U-Net structure during the decoding process; however, it ensures that more up-sampling operations are incorporated, along with optimal integration of relevant features encoded during the process. This ensures that the fundamental features retrieved during the down-sampling process are effectively utilized during the up-sampling process, thereby ensuring more accurate results.

The proposed framework incorporates the Rectified Linear Unit (ReLU) activation function, which has been selected considering its simplicity during computation, along with its effectiveness in resolving vanishing gradient problems, which commonly arise during deep network architectures. The use of the ReLU activation function has been preferred considering its non-linear nature, along with its ability to ensure efficient gradient computation, thereby allowing faster convergence during network training. The proposed framework has selected the ReLU activa-

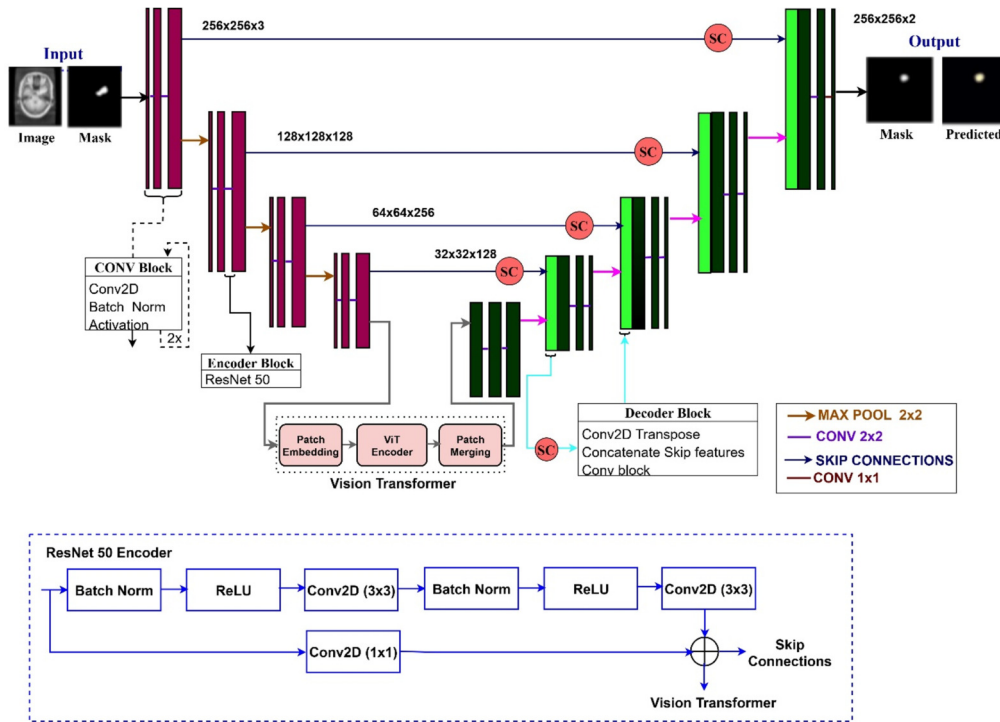


Fig.3: Architectural flow of the proposed Vision-ResU-Net50 model.

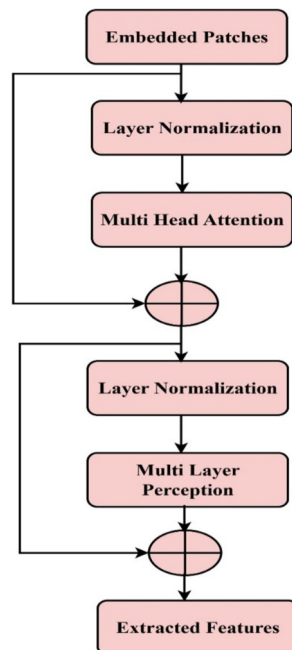


Fig.4: Structure of Vision Transformer (ViT).

tion function considering its better performance and reduced computation, when compared with other activation functions, such as sigmoid and tanh, which have been commonly utilized during medical image segmentation tasks.

The skip connections, which are essential for restoring spatial as well as structural details lost in the encoding process, have also been refined. These connections emphasize the most pertinent features at each level of resolution by applying a dynamic weighing process and immediately concatenating feature maps derived from the encoder onto the decoder. A type of deep learning framework for applications in computer vision is called a Vision Transformer. ViT is added to the suggested hybrid architecture to enhance its capacity for global visualization while improving the abilities of ResNet50 in extracting local features. The ViT structure uses self-attention strategies to describe interactions between various areas of an image, as shown in Figure 4. The ResNet50 encoded image is split up into 16x16 pixel fixed-size patches. To ensure non-redundant spatial distribution, a stride of 16 prevents overlapping. After being flattened to a 1-D vector, each picture patch is subsequently linearly projected using a learned projection onto a smaller-dimensional embedding space. The model distinguishes across patch positions by adding embedded positional information to every token in order to preserve spatial context. Each patch can then obtain contextual data from every other patch

owing to its self-attention process, establishing dependency relationships that are essential for recognizing tumors with different forms and positions. Patch merging then transforms the output from the ViT encoder into the necessary size of the first decoder block of the U-Net structure. This technique utilizes low-level spatial data in conjunction with high-level conceptual characteristics to accurately localize and segment brain tumors.

3.4 Assessment Metrics

The metrics used to assess the efficacy of the proposed model are discussed in this section. These parameters measure the degree of reliability, accuracy, and resemblance between the true and the estimated segmentation masks [24]. Significance of these metrics, along with mathematical formulae, is provided in Table 1.

4. EXPERIMENTAL RESULTS AND ANALYSIS

This section includes experimental results and provides a full analysis. The paper outlines the implementation as well as the experimental setup for evaluating the proposed deep learning model for the segmentation of brain tumors. The paper contains quantitative and qualitative results, an evaluation of the metrics, and a comparison with the existing methods. The problem of class imbalance was addressed by the

Table 1: Mathematical formulae and significance of assessment metrics.

Metric	Formulae	Significance
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$ <i>TP</i>-True Positives, <i>TN</i>- True Negatives <i>FP</i>-False Positives, <i>FN</i> -False Negatives	It measures the percentage of correctly categorized pixels in the segmentation mask when compared with the real mask. A high score of accuracy implies the model accurately identifies tumor as well as non-tumor areas in images.
Precision	$Precision = \frac{TP}{TP + FP}$	It calculates the percentage of accurately predicted malignant pixels among all predicted malignant pixels. Its main goal is to reduce false positives.
Recall	$Recall = \frac{TP}{TP + FN}$	It measures the percentage of real tumor pixels that can be detected by the model accurately. A high recall score means that the model is able to identify tumor locations with little loss of important features.
F1-Score	$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall}$	It provides a single indicator of the model's aggregate segmentation efficiency by balancing both false negatives and false positives.
Pixel Accuracy	$PA = \frac{\sum_{j=1}^M TP_j}{\sum_{j=1}^M (TP_j + FP_j + TN_j + FN_j)}$ Where <i>M</i>-No. of Classes	A comprehensive overview of the model's performance is provided by Pixel Accuracy, which is the percentage of properly categorized pixels throughout the whole dataset.
Dice Coefficient (DC)	$DC = \frac{2 * X \cap Y }{ X + Y }$ Where <i> X </i> and <i> Y </i> are the true and predicted masks	It provides a more precise measure of assessment for datasets that are unbalanced by measuring the spatial overlapping between the anticipated and reality segmentation masks.
Jaccard Index (JI)	$JI = \frac{ X \cap Y }{ X \cup Y }$	The algorithm compares anticipated and true segmentation masks by calculating the intersection-to-union ratio.

application of the filtration method, ensuring the proportionate distribution of the classes. The data was split into the training and test sets using the “patient-wise” method, where 80% and 20% of the patients were allocated to the respective sets. The allocation ensures that the images from the same patients are not included in the test and train sets, avoiding the data leakage problem. All experiments ran on a system with the NVIDIA GPU (e.g., RTX 3080, 3090), an Intel Core i7, and 16 GB RAM. The proposed model uses Python and the libraries Tensor Flow and Torch, which are the most popular libraries for deep learning. The system was sufficient to train the proposed model.

The model uses the Adam optimizer with an initial value of 0.001. The “step decay” method reduces the learning rate. The batch size was 16, and the model trains for 50 epochs. The random seed is fixed to ensure reproducibility.

Segmentation accuracy is evaluated prior to and after using DBO optimization, as shown in Table 2, to assess the efficacy of DBO-inspired thresholding, morphological processing, and Test-Time Augmentation. The findings show that DBO-driven im-

provements produce notable increases in the accuracy of segmentation. According to the findings of Table 2, DBO-threshold optimization successfully eliminated noise artifacts associated with segmentation masks, resulting in a 2.3% increase in precision. Recall dropped by 0.4%, suggesting the post-processing inhibited small tumor areas. Dynamic threshold selection improves segmentation accuracy, as evidenced by the 1.9% increase in the dice coefficient. Likewise, the Jaccard Index rose by 3.2%, indicating that following optimization, tumor borders were more precisely defined.

The ablation study in Table 3 demonstrates that each component of the DBO-based preprocessing pipeline contributes incrementally to performance improvement. The initial gain comes from the optimization of the threshold, and then comes the morphology and time to arrival, which further improve the quality of segmentation.

Figure 5 shows a visual representation of the segmentation results, providing informative insights into how well the model performs from a subjective perspective. For each type of tumor, a visual representation illustrates how well the segmentation has been

Table 2: Analysis of DBO optimized preprocessing.

Metric	Before Applying DBO	After Applying DBO	% Improvement
Precision	0.8589	0.8796	2.3
Recall	0.9827	0.9783	-0.4
Dice Coefficient	0.9203	0.9389	1.9
Jaccard Index	0.8469	0.8756	3.2

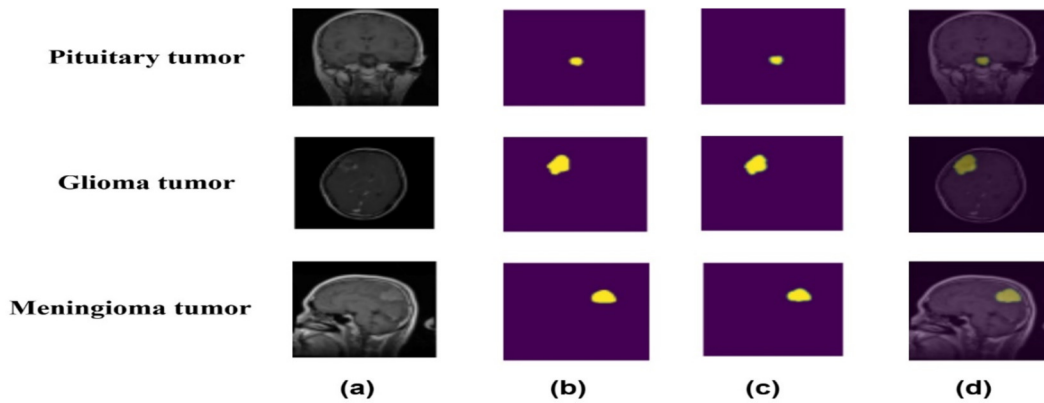


Fig.5: Results of segmentation : (a) MRI images of the brain (b) Real mask (c) Predicted mask (d) Overlapped results.

Table 3: Ablation analysis of proposed work.

Model	Dice Coefficient	Jaccard Index
Base model without DBO	0.9203	0.8469
Base model +Threshold Optimization	0.9287	0.8592
Base model +Threshold Optimization+ Morphology Optimization	0.9345	0.8698
Base model +Threshold Optimization+ Morphology Optimization+ TTA Optimization	0.9389	0.8756

Table 4: Quantitative evaluation of the proposed model for different tumors.

Metric	Pituitary tumor	Glioma tumor	Meningioma tumor
Accuracy	0.9978	0.9987	0.9994
Precision	0.9356	0.9354	0.9364
Recall	0.9455	0.9456	0.9463
F1-Score	0.9416	0.9421	0.9419
Pixel Accuracy	0.9834	0.9839	0.9837
Dice Coefficient	0.9630	0.9873	0.9862
Jaccard Index	0.9113	0.9483	0.9467

performed compared to the original ground-truth segmentation masks. Each type of tumor has different characteristics, and these affect how segmentation has been performed, thus requiring a different approach for evaluation and interpretation. Table 4 presents the quantitative evaluation of the proposed model for different classes. The model achieves a Pixel Accuracy of 98.34%, 98.39%, and 98.37%, respectively for three tumor classes. However, when dealing with imbalanced datasets, such as medical image segmentation, where tumor areas are a very small percentage of the overall image, a high value of Pixel Accuracy does not necessarily indicate a good segmentation result. Therefore, a deeper evaluation and analysis are required, especially when a classifier might perform poorly in segmenting tumor areas from non-tumor areas, even if a high value of Pixel Accuracy has been obtained, such as 90% if all pixels are classified as belonging to the background, even if the image consists of 90% tumor and 10% background, and a classifier fails to segment even a single pixel from the tumor area. In this regard, the Jaccard Index and Dice Coefficient, which are more reliable metrics for measuring segmentation accuracy, especially when dealing with imbalanced datasets, are emphasized. The Jaccard Index values, namely 0.911, 0.948, and 0.946, and Dice Coefficient values, namely 0.983, 0.987, and 0.986, indicate a substantial overlap with the actual masks, implying a high degree of accuracy and efficacy in capturing intricate features of tumors. The robust performance of the model in maintaining distinctive boundaries among tumors also speaks volumes for its reliability and accuracy in volumetric analysis and clinical decision-making processes.

The efficacy and accuracy of the model in precisely locating tumor areas are also evident from the remarkable performance of the Glioma and Meningioma classes, which attained a Dice Coefficient value of 0.9873 and 0.9862, respectively. This also implies a high degree of concurrence and agreement between the actual and estimated tumor segmentations, thereby ensuring a reliable and accurate model for tumor segmentation and detection. Conversely, the performance of the proposed model in relation to the Pituitary class, although relatively lower, is still praiseworthy, with a Dice Coefficient value of 0.9630. The inherent difficulties associated with pituitary tu-

mors, such as distinctive morphological features and proximity to vital brain areas, may account for such differences in performance compared to other tumor classes. Although a relatively lower Dice value is observed for this class compared to other tumor types, the proposed model has also ensured a high accuracy level in tumor segmentation and detection, and to avoid over-segmentation and misclassification of non-tumor areas as cancerous, a high precision level needs to be maintained.

This may be due to over-sectioning, extensive surgery, or improper dimension determination. The proposed model can attain an average precision of 93.5%. It is vital to strike an appropriate balance between precision and recall to ensure correct tumor detection without misclassification of normal brain tissues. The proposed approach strikes an optimal balance between precision and recall, thus providing an effective tumor detection rate. The F1 score of 94.1% further supports the findings, thus providing a balance between precision and recall. One of the main advantages of the proposed approach is its high average recall rate of 94.5%, thus providing significant therapeutic benefits for mitigating under-segmentation risk. Under-segmentation risk is a major hazard in medical treatment, where some of the tumor tissues may be missed. This may cause an under-evaluation of the tumor tissues, thus leading to improper treatment. The proposed approach ensures almost complete detection of tumor tissues, thus providing a high recall rate. This minimizes the risk of under-segmentation, thus providing an effective approach for treatment. The proposed approach segments images more effectively than other approaches currently used for brain tumor image segmentation. The proposed approach's segmentation results are comprehensively compared with the best techniques currently being used for brain tumor image segmentation. The proposed approach's performance, based on the Jaccard index and Dice coefficient for pituitary, glioma, and meningioma for various tumor classes, is presented in Table 5.

The suggested approach performs quite well when assessing glioma segmentation, obtaining a Dice value of 0.9873 and a Jaccard Index value of 0.9483. These outcomes are better than many current techniques. Similarly, the suggested method outperforms com-

Table 5: Comparative analysis of segmentation performance of different methods.

Method	Metric	Pituitary tumor	Glioma tumor	Meningioma tumor
Edge U-Net [16]	Dice coefficient	0.8432	0.8190	0.8181
	Jaccard Index	0.7743	0.8478	0.7985
CNN [25]	Dice coefficient	0.8517	0.8273	0.8263
	Jaccard Index	-	-	-
Deep Learning [26]	Dice coefficient	0.8691	0.8441	0.8432
	Jaccard Index	-	-	-
Improved U-Net [27]	Dice coefficient	0.9148	0.8885	0.8875
	Jaccard Index	-	-	-
Proposed	Dice coefficient	0.9630	0.9873	0.9862
	Jaccard Index	0.9113	0.9483	0.9467

peting methods such as Edge U-Net, enhanced U-Net, CNN, and Deep learning for the segmentation of Meningioma tumor, with a Dice coefficient value of 0.9862 and Jaccard Index value of 0.9467. For segmentation of pituitary tumors, the suggested technique produces noteworthy results, with a Dice coefficient value of 0.9630 and a Jaccard Index value of 0.9113, demonstrating its superiority over current approaches, despite the fact that they perform better in this class when compared to other tumor classes. Several major architectural breakthroughs have contributed to the suggested hybrid architecture's enhanced performance. First, the incorporation of residual learning alleviates the vanishing gradient issue common in deep CNN designs. The integration of the ViT structure efficiently captures inter-patch relations. Existing segmentation models, especially Edge U-Net and Improved U-Net, suffer from the extraction of deep features due to a lack of residual connections, resulting in the loss of important spatial information. On the other hand, ResUNet50 uses residual blocks to enable the transfer of information between layers while preserving structural details, thus enabling deeper networks to learn more feature representations. The tumor boundaries are well defined, and the segmentation is improved.

The improvement in the model's performance results from the following components:

- (i) The ResNet50 encoder is responsible for the improvement in the performance of the model through the application of residual learning in the feature extraction process.
- (ii) The application of the Vision Transformer enhances the performance by capturing contextual dependencies, thus improving the performance in the segmentation of the irregular tumor regions.
- (iii) The DBO-based preprocessing enhances the performance by refining the generated masks through the optimization process.

5. CONCLUSION

To develop an accurate and reliable model for the segmentation of brain tumors in MRI scans, the authors have developed the Vision-ResU-Net50 model, which is a hybrid model that combines the strengths of ResNet50, the Vision Transformer, and the U-Net model. The model is designed to handle the challenges associated with the accurate and reliable segmentation of brain tumors in MRI scans, as well as the need to ensure the model is reliable and accurate. The model is improved by applying the DBO-optimized preprocessing method, which enhances the quality of the generated masks. The model was tested using three types of tumors, and the results show that the model is the best among the other models, including the Edge U-Net, Improved U-Net, CNN, and other deep learning models. The model achieved high Dice and Jaccard values, 0.9630, 0.9873, and 0.9862, in the segmentation of pituitary, glioma, and meningioma, respectively. The improvement in the Dice and Jaccard index values is 17% and 15%, respectively, indicating the model's potential in preserving the boundaries and its generalization capabilities.

AUTHOR CONTRIBUTIONS

Conceptualization, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; methodology, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; software, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; validation, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; formal analysis, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; investigation, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; writing—original draft preparation, Ramalingeswara Rao N. M.; writing—review and editing, Ramalingeswara Rao N. M. and Dr. G. Nagaraju; supervision, Dr. G. Nagaraju.

References

- [1] A. Abo-El-Rejal, S. A. Ayman and F. Aymen, "Advances in breast cancer segmentation: A comprehensive review," *Acadlore Transactions on AI and Machine Learning*, vol. 3, no. 2, pp. 70–83, 2024.
- [2] S. Zahoor, I. U. Lali, M. A. Khan, K. Javed and W. Mehmood, "Breast cancer detection and classification using traditional computer vision techniques: A comprehensive review," *Current Medical Imaging Reviews*, vol. 16, no. 10, pp. 1187–1200, 2020.
- [3] M. Moghbel, C. Y. Ooi, N. Ismail, Y. W. Hau and N. Memari, "A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography," *Artificial Intelligence Review*, vol. 53, pp. 1873–1918, 2020.
- [4] S. Vijh, M. Saraswat and S. Kumar, "Automatic multilevel image thresholding segmentation using hybrid bio-inspired algorithm and artificial neural network for histopathology images," *Multimedia Tools and Applications*, vol. 82, pp. 4979–5010, 2023.
- [5] A. Ashurov *et al.*, "Improved breast cancer classification through combining transfer learning and attention mechanism," *Life*, vol. 13, no. 9, p. 1945, 2023.
- [6] X. Zhang *et al.*, "HiViT: A Simpler and More Efficient Design of Hierarchical Vision Transformer," in *Proceedings International Conference on Learning Representations (ICLR)*, 2023.
- [7] J. Ruan, J. Li and S. Xiang, "VM-UNet: Vision Mamba UNet for medical image segmentation," *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024.
- [8] X. Ma, Y. Du and D. Sui, "A U-shaped architecture based on hybrid CNN and Mamba for medical image segmentation," *Applied Sciences*, vol. 15, no. 14, p. 7821, 2025.
- [9] H. Yang, Y. Liu and T. Xia, "Defect detection scheme of pins for aviation connectors based on image segmentation and improved ResNet-50," *International Journal of Image and Graphics*, vol. 24, no. 1, p. 2450011, 2024.
- [10] J. Shedbalkar, K. Prabhushetty, and R. H. Havaladar, "UNet and transformer-based model for multi-modality brain tumor segmentation," *International Journal of Electronics and Communication Engineering*, vol. 10, no. 8, pp. 22–35, 2023.
- [11] G. Huang *et al.*, "Channel-Attention U-Net: Channel Attention Mechanism for Semantic Segmentation of Esophagus and Esophageal Cancer," in *IEEE Access*, vol. 8, pp. 122798–122810, 2020.
- [12] K. Avazov, S. Mirzakhilov, S. Umirzakova, A. Abdusalomov and Y. I. Cho, "Dynamic focus on tumor boundaries: A lightweight U-Net for MRI brain tumor segmentation," *Bioengineering*, vol. 11, no. 12, p. 1302, 2024.
- [13] A. H. Nizamani, Z. Chen, A. A. Nizamani and U. A. Bhatti, "Advance brain tumor segmentation using feature fusion methods with deep U-Net model with CNN for MRI data," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 9, p. 101793, 2023.
- [14] K. Swaraja, K. Meenakshi, H. B. Valiveti and G. Karuna, "Segmentation and detection of brain tumor through optimal selection of integrated features using transfer learning," *Multimedia Tools and Applications*, vol. 81, pp. 27363–27395, 2022.
- [15] T. A. Soomro *et al.*, "Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review," in *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 70–90, 2023.
- [16] M. G. Allah Ahmed, A. M. Sarhan and N. M. Elshennawy, "Edge U-Net: Brain tumor segmentation using MRI based on deep U-Net model with boundary information," *Expert Systems with Applications: An International Journal*, vol. 213, no. PA, 2023.
- [17] Y. Chang, Z. Zheng, Y. Sun, M. Zhao, Y. Lu and Y. Zhang, "DPAFNet: A residual dual-path attention-fusion convolutional neural network for multimodal brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 79, p. 104037, 2023.
- [18] J. Cheng *et al.*, "Enhanced performance of brain tumor classification via tumor region augmentation and partition," *PLOS ONE*, vol. 10, no. 12, p. e0144479, 2015.
- [19] D. Shah and J. Sen, "Pituitary macroadenoma: A comprehensive case study of surgical intervention and postoperative management," *Cureus*, vol. 16, no. 4, p. e59387, 2024.
- [20] N. Rasool and J. I. Bhat, "A critical review on segmentation of glioma brain tumor and prediction of overall survival," *Archives of Computational Methods in Engineering*, vol. 32, pp. 1525–1569, 2025.
- [21] L. Rogers *et al.*, "Meningiomas: Knowledge base, treatment outcomes, and uncertainties. A RANO review," *Journal of Neurosurgery*, vol. 122, no. 1, pp. 4–23, 2015.
- [22] Z. Li, W. Liu, H. Zhao and W. Pu, "Dung beetle optimization with composite population initialization and multi-strategy learning for multi-level threshold image segmentation," *Signal, Image and Video Processing*, vol. 19, no. 255, 2025.
- [23] W. Fang, Z. Yu, Y. Chen, T. Huang, T. Masquelier and Y. Tian, "Deep residual learning in spiking neural networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21056–21069, 2021.

- [24] S. H. H. Abdalgadir and M. Öztürk, "Image enhancement and U-Net based brain tumor segmentation using MRI," *European Journal of Engineering and Applied Sciences*, vol. 8, no. 1, pp. 31–38, 2025.
- [25] S. Kumar and D. Kumar, "Human brain tumor classification and segmentation using CNN," *Multimedia Tools and Applications*, vol. 82, pp. 7599–7620, 2023.
- [26] K. T. Islam, S. Wijewickrema and S. O'Leary, "A deep learning framework for segmenting brain tumors using MRI and synthetically generated CT images," *Sensors*, vol. 22, no. 2, p. 523, 2022.
- [27] P. Zheng, X. Zhu and W. Guo, "Brain tumour segmentation based on an improved U-Net," *BMC Medical Imaging*, vol. 22, no. 199, 2022.



N. M. Ramalingeswararao is currently pursuing a Ph.D. at Jawaharlal Nehru Technological University Kakinada in the Department of Electronics and Communication Engineering. His area of specialization is Image Processing. He completed his M.Tech degree in Digital Electronics and Communication Systems. His research interests include image fusion, medical image processing, and deep learning applications.



G. Nagaraju is an Associate Professor in the Department of Electronics and Communication Engineering at SRKR Engineering College. He received his Doctorate from KL University. He has presented numerous papers in international journals and conferences. His areas of interest include Image Processing, Signal Processing, and VLSI Design. He has extensive teaching experience at various academic levels.