Biomedical Signal Processing and Control 113 (2026) 108917

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

BrainDx: a dual-transformer framework using PVT and SegFormer for tumor diagnosis

Arshleen Kaur [a], Vinay Kukreja [a], Modafar Ati [b,*], Ankit Bansal [c], Shanmugasundaram Hariharan [d]

[a] Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[b] Computer Science and Information Technology, College of Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates
[c] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[d] Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India

ARTICLE INFO

Keywords:
Brain Tumor Classification
Tumor Segmentation
Pyramid Vision Transformer (PVT)
SegFormer
Deep Learning

ABSTRACT

Context: Brain tumor diagnosis is challenging due to their complex morphology, indistinct boundaries, and subtle variations in Magnetic Resonance Imaging (MRI) scans. Manual diagnosis is time-consuming and error-prone, making the need for automated systems crucial. Recent advancements in deep learning, particularly in transformer models, have led to improved accuracy and speed in medical image analysis.
Objective: This research aims to develop an Artificial Intelligencee (AI) based framework that integrates the Pyramid Vision Transformer (PVT) for tumor classification and the SegFormer for tumor segmentation, thereby enhancing diagnostic accuracy, speed, and reducing human error in brain tumor detection.
Methodology: The proposed framework, BrainDX, utilizes PVT to classify MRI images into tumor types (Gliomas, Meningiomas, Pituitary Tumors, and Healthy Brain), and SegFormer to segment tumor regions in real-time. The dataset consists of annotated MRI images that undergo preprocessing (normalization, resizing, and augmentation). The models are trained and evaluated based on performance metrics, including accuracy, Dice score, Intersection over Union (IoU), and segmentation time.
Results: The framework was evaluated across three benchmark MRI datasets, achieving a classification accuracy of 94.0% and a Dice score of 0.87 for tumor segmentation. SegFormer demonstrated real-time segmentation, processing MRI images in under 50 ms. Both models maintained high efficiency while delivering robust performance, even in cases of irregular tumor boundaries.
Future Scope: Future work will focus on further optimizing the model for real-time clinical use, improving generalization across diverse tumor types and MRI modalities. This AI-powered system has the potential to enhance diagnostic processes and improve patient outcomes significantly.

1. Introduction

Medical counselors face difficulties in diagnosis because brain tumors display various forms and create indistinct boundaries and subtle variations in magnetic resonance imaging scans. The identification process should begin as early as possible, in conjunction with precise diagnostic methods, to achieve the best recovery outcomes and informed therapeutic planning [1]. A diagnostic procedure conducted manually using MRI requires extensive time and delivers uneven results, as it depends on human analysts who produce different interpretive outcomes [2]. The application of deep learning artificial intelligence has become a promising approach to automating brain tumor diagnosis, providing both accelerated testing and standardized analysis results [3]. The latest developments in medical image spatial complexity management are driven by transformer models that operate under the deep learning paradigm. Transformers exhibit better performance than traditional Convolutional Neural Networks (CNNs) [4] because they can identify long-range relationships and multiple feature sizes, which are particularly effective in analyzing tumor areas [5]. The proposed dual-model system combines PVT for identifying different types of tumors

* Corresponding author.
E-mail addresses: arshleen.kaur@chitkara.edu.in (A. Kaur), onlyvinaykukreja@gmail.com (V. Kukreja), modafar.ati@adu.ac.ae (M. Ati), ankit.bansal@chitkara.edu.in (A. Bansal), mailtos.hariharan@gmail.com (S. Hariharan).

https://doi.org/10.1016/j.bspc.2025.108917
Received 29 May 2025; Received in revised form 5 August 2025; Accepted 6 October 2025
Available online 14 October 2025

with SegFormer for performing detailed MRI image segmentation of tumor areas [6]. This research introduces the separate deployment of advanced PVT and SegFormer networks, leveraging their complementary architecture strengths to achieve robust tumor identification through hierarchical attention and real-time tumor segmentation without CRF processing [7]. The framework testing involved a strong application on documented brain tumor data for complete analysis and accuracy measurements, along with Dice score and Intersection over Union (IoU) results, as well as segmentation time assessment [8]. The extensive investigation aims to establish a functional secondary diagnostic system based on artificial intelligence that can analyze brain tumors at an industrial level. The adoption of AI technology offers a significant opportunity for improved medical brain tumor identification and treatment methods [9]. Manual diagnosis of brain tumors takes excessive time, with the added disadvantage of unreliable results, specifically when tumors display similar features or show minor variations [10]. Deep learning, alongwith transformer models, leverages AI processing to verify large observational datasets at both high speed and beyond human quantifiable accuracy [11]. The AI automation of brain tumor classification, along with segmentation functions, effectively reduces medical errors and streamlines clinical processing methods [12]. This is the first study to introduce a dual-transformer framework that combines Pyramid Vision Transformer (PVT) for brain tumor classification with SegFormer for tumor segmentation. This novel pairing leverages their complementary strengths to achieve high accuracy and real-time performance, specifically developed for brain tumor MRI analysis. The clinical usefulness of transformer models extends to their ability to adapt to multiple tumor types, as well as different image resolutions in medical environments [13]. The proposed framework demonstrates the capability to enhance diagnostic operations while helping doctors make better decisions, which results in improved patient outcomes accompanied by optimized care strategies [14].

1.1. Motivation

Manual diagnosis of brain tumors from MRI scans is a time-consuming process that demands significant radiologist expertise and is susceptible to inter-observer variability [12]. With the increasing volume of radiological data, there is a critical need for automated solutions that can facilitate accurate, fast, and consistent tumor detection and delineation [15]. Conventional CNN-based models, although widely adopted, often face challenges in learning global context and require additional post-processing steps (e.g., Conditional Random Fields) to refine segmentation outputs [16]. Moreover, models such as U-Net and DeepLabV3 typically lack the efficiency required for real-time clinical deployment [17].

Recent advancements in vision transformers have demonstrated superior performance in capturing both local and global dependencies in image data [18]. However, most studies employ transformers either for classification or segmentation alone. To the best of our knowledge, this is the first work to propose a dual-transformer framework that combines the PVT for classification and SegFormer for segmentation, specifically designed for brain tumor MRI analysis. This combination aims to address the limitations of existing methods by improving diagnostic accuracy, reducing inference time, and eliminating the need for complex post-processing, thereby making the framework suitable for integration into real-world clinical workflows.

1.2. Contribution of the study

- Introduced a dual-model approach named BrainDX, combining Pyramid Vision Transformer (PVT) for classification and SegFormer for real-time tumor segmentation, enhancing both diagnostic accuracy and efficiency.
- Achieved high classification accuracy of 94.0 % using PVT, effectively distinguishing between multiple brain tumor types and healthy brain tissue.
- SegFormer demonstrated strong segmentation performance, achieving a Dice score of 0.87, which accurately describes tumor boundaries, even in complex cases with irregular shapes.
- Developed a framework capable of segmenting MRI images in under 50 ms, ensuring fast and efficient processing.
- The PVT and SegFormer models maintained high performance while being computationally efficient, with a relatively compact architecture and fast inference time.

2. Methodology

Fig. 1 illustrates the methodology employed for tumor classification and segmentation using deep learning models. **Phase 1: Data Collection** describes the acquisition of MRI images, categorized into tumor types like Gliomas, Meningiomas, Pituitary Tumors, and Healthy Brain tissue. These images are annotated with tumor masks for subsequent analysis. In **Phase 2: Data Preprocessing**, the dataset undergoes cleaning, normalization, and augmentation to enhance data quality and improve model generalization. **Phase 3: Model Selection** highlights the implementation of two models: SegFormer for pixel-wise segmentation of tumor regions and PVT (Pyramid Vision Transformer) for tumor classification. These models are designed to extract relevant features from the preprocessed data and provide accurate tumor diagnosis.

The goal is to effectively execute brain tumor classification and segmentation using the proposed framework. The methodology employs state-of-the-art deep learning architectures, including the PVT for classification, as well as SegFormer for segmentation. Additionally, it incorporates training and evaluation steps, along with data preprocessing and model selection. The entire process is structured using code and mathematical formulas, as shown in the following pseudo-code, which is depicted in the algorithm.

Algorithm: Tumor Classification and Segmentation using Deep Learning Models

Input: Annotated MRI images dataset (Gliomas, Meningiomas, Pituitary Tumors, HealthyBrain)Output: Tumor classification and segmentation results

Step 1: **Data Collection Phase:**

- Load Dataset: Load MRI images with tumor annotations from the dataset (Gliomas, Meningiomas, Pituitary Tumors, Healthy Brain).
- Annotation: Images are annotated with tumor masks for segmentation.

Step 2: **Data Preprocessing Phase:** *Data Cleaning:* Remove corrupted or incomplete images.

- Check for corrupted files using checksum validation.
- Exclude images with missing tumor labels.

Normalization: Normalize the image pixel values to the range [0, 1].
Resizing: Resize all images to 224x224 pixels to standardize the input size.
Data Augmentation: Apply augmentation techniques:

- Rotation: Random rotation between 0° and 45°.
- Flipping: Horizontal and vertical flips.
- Translation: Random shifts along the x and y axes.
- Zooming: Random zooming within a specified range.

Data Split: Divide the dataset into three subsets:

- Training set (70 %)
- Validation set (15 %)
- Test set (15 %)

Step 3: **Model Selection Phase:** *Classification & Feature Extraction Model: PVT Segmentation Model:* SegFormer

Step 4: **PVT Model Implementation:** *Input:* Resized and normalized MRI images. *Feature Extraction (Multi-Scale):* The feature extraction process involves multiple transformer encoder layers to extract relevant features at varying levels of detail.

- $OutputFeaturemap = TransformerEncoder(InputFeatureMap)$

(continued on next page)

(*continued*)

Algorithm: Tumor Classification and Segmentation using Deep Learning Models

Attention Mechanism: The self-attention mechanism computes relationships between image patches to focus on relevant parts of the image.

- $Attention(Q, K, V) = Softmax((Q*K \wedge T)/sqrt(d\_K)*V)$

Output Layer: The final layer classifies the image into one of several categories: Gliomas, Meningiomas, Pituitary Tumors, or Healthy Brain.

- $OutputProbabilities = Softmax(W*FeatureMap + b)$

Loss Function: Use Cross-Entropy Loss to train the model.

- $Cross - EntropyLoss = -\sum (Y*log(p))$ *Where Y is the actual label and pIs the predicted probability.*

Step 5: **SegFormer (Segmentation Transformer) Model Implementation:** *Input:* Resized and normalized MRI images.*Pixel-wise Classification:* Classify each pixel as either tumor or healthy tissue.

- *Segmentation Strategy:* Use Dice Loss combined with Cross-Entropy Loss for optimization.
- $DiceLoss = 1 - (2*|A \cap B|)/(|A| + |B|)$
- $TotalLoss = DiceLoss + CrossEntropyLoss$

Multi-Scale Feature Fusion: Combine features from multiple scales to handle tumors of varying sizes and improve accuracy.

- $OutputFeatureMap(CurrentStage) = FeatureMap(PreviousStage) + Multi - ScaleFeatures$

Output: Pixel-wise segmentation mask indicating tumor regions.

Step 6: **Training Phase:** *Train Models:* Train both the PVT and SegFormer models using the training set (70 % of the data).*Validation:* Use the validation set (15 % of the data) to fine-tune hyperparameters and prevent overfitting. *Optimization:* Utilize data augmentation and other techniques to enhance model robustness and generalization.

Step 7: **Evaluation Phase:** *Classification Evaluation:* Evaluate the classification performance using the test set (15 % of the data).

- *Metrics:* Accuracy, Precision, Recall, F1-Score, Specificity, Area Under Curve (AUC).

Segmentation Evaluation: Evaluate the segmentation performance using the test set.

- *Metrics:* Dice Score, Intersection over Union (IoU), Pixel Accuracy, Boundary F1-Score, Hausdorff Distance, Segmentation Time.

2.1. Data collection (phase 1)

Acquiring data remains essential in all machine learning-based medical image analysis processes. The research utilized annotated MRI images of brain tumors as the dataset. The selected dataset contains sufficient samples with diverse tumor types, allowing for the simultaneous execution of both classification and segmentation operations. The BraTS 2020, Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji) provide publicly available ground truth annotations that were curated and validated by expert neuroradiologists as part of the challenge. The following section details the entire dataset structure, alongside annotation specifications, and outlines all preprocessing methods that optimized data readiness for training purposes.

2.1.1. Dataset description

The research utilizes brain tumor images in a dataset to study medical image analysis classification and segmentation systems. Fig. 2 provides a visual representation of a brain tumor dataset, explaining its format, tumor types, and annotations. The central MRI image serves as the focal point, with data format icons Neuroimaging Informatics Technology Initiative (NIfTI) and Digital Imaging and Communcations in Medicine (DICOM) placed on the sides, showing the formats used for the images. Surrounding the MRI image are icons representing different tumor types (Gliomas, Meningiomas, Pituitary Tumors, and Healthy Brain), connected by arrows to indicate their presence in the dataset. Below the MRI image, annotations are shown with highlighted tumor regions and labeled with "Annotated with Tumor Masks," demonstrating the importance of these labels for segmentation tasks. This clear layout helps viewers understand how the dataset is structured and annotated.

The database contains various brain tumor scans that retain multiple tumor types alongside multiple image modalities. The following provides an in-depth description:

- **Source of the Dataset**: The dataset is sourced from [19–21]. This dataset is well-known in medical imaging for its applications in tumor classification and segmentation tasks.
- **Dataset Size:** The dataset comprises 3,000 images of brain scans, specifically MRI scans. These images are organized into subcategories of gliomas, meningiomas, pituitary tumors, and healthy brain tissue, as shown in Table 1:

2.1.2. Data preprocessing (phase 2)

Preprocessing of the dataset is crucial for standardizing the images, handling data issues, and augmenting the data to improve the generalization of models.

- **Data Cleaning**

The dataset is first cleaned by removing corrupted or incomplete images. Corrupted files are identified through checksum validation or by manually inspecting the images for any inconsistencies (e.g., missing image data, corrupted metadata). Any missing data (e.g., images without corresponding tumor labels) is excluded from the dataset. Corruptions included blank scans, extreme brightness/contrast anomalies, and missing DICOM metadata. Domain experts performed visual inspection to ensure anatomical consistency before annotation.

- **Normalization**

To ensure uniformity in image inputs, normalization is applied to each image. Specifically, pixel values are rescaled to the range [0, 1]. All MRI images were resized to 224x224 pixels to standardize the input dimensions for CNNs and transformer-based models, ensuring uniformity across the dataset. Min-max normalization was preferred over Z-score normalization due to its compatibility with activation functions like ReLU and improved training stability in deep convolutional layers, as illustrated in Fig. 3.

All images were grayscale, and normalization was applied consistently across the single intensity channel.

The distribution of pixel intensities in MRI images is shown in Fig. 4. This helps in understanding the range of intensities in the dataset and confirms the need for normalization.

- **Augmentation**

To prevent overfitting and increase dataset variability, several augmentation techniques are applied: rotation (random rotation between 0° and 45°), flipping (horizontal and vertical), translation (random shifts along the x and y axes), and zooming (random zoom within a specified range). These techniques help improve the model's generalization by exposing it to diverse variations of images. The following Fig. 5 illustrates the augmentation process. No oversampling methods, such as Synthetic Minority Over-sampling Technique (SMOTE) or Adaptive Synthetic Sampling (ADASYN), were applied, as the dataset classes were already balanced. Additionally, spatial augmentation was considered more suitable for preserving anatomical consistency in medical imaging tasks.

The dataset is divided into three subsets: training, validation, and test sets, following a 70–15-15 ratio as shown in Fig. 6. This means that 70 % of the total images are allocated to the training set, which consists of 2100 images. During model training, the validation set comprises 450 images to both adjust hyperparameters and prevent overfitting. The evaluation of a model's final performance is conducted using a test set

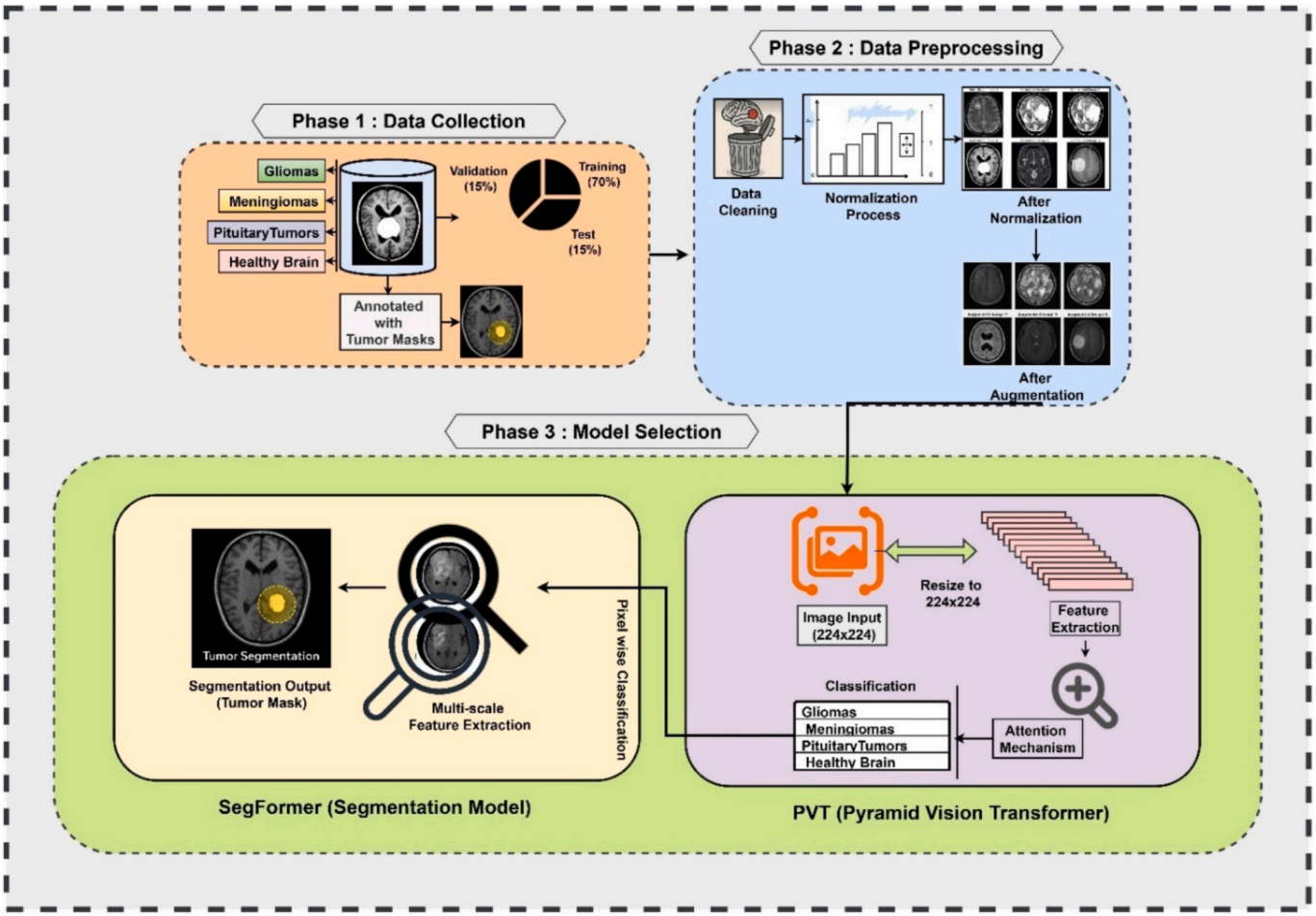

Fig. 1. BrainDX: Proposed Framework.

that contains 450 images after training and validating the model. The data split enables proper training, validation, and testing of models to determine their performance evaluation measures accurately. The distribution of tumor types in Fig. 6 verifies the appropriate representation of all classes in the training set.

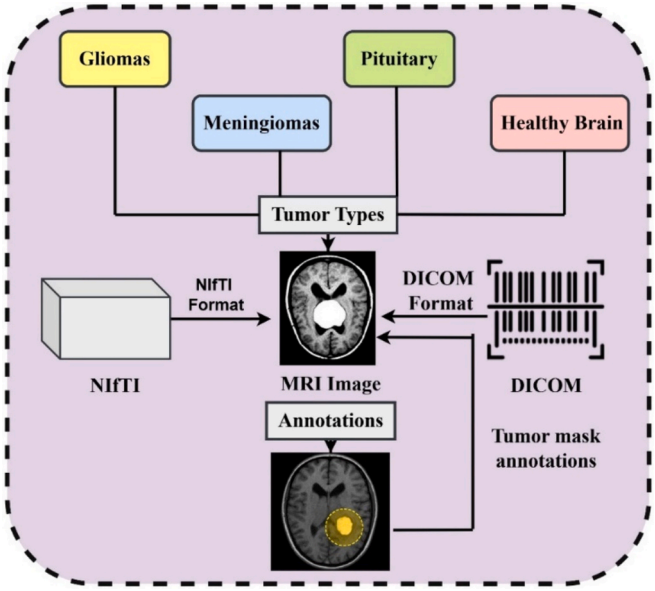

Fig. 2. Overview of Dataset Structure.

2.2. Model selection (phase 3)

The following part selects PVT for classification and feature extraction alongside SegFormer for segmentation applications. Justifications for choosing these models are provided, along with details on their implementation, including any customizations made for improved performance on the brain tumor dataset.

2.2.1. PVT (pyramid vision transformer)

Image classification benefits from the PVT, which obtains global and local visual features during its extraction process [22]. PVT's self-attention mechanism outperforms CNNs since it gathers essential image components from complete image regions without dependence on local processing, thus making it appropriate for brain tumor diagnostics that spread their diagnostic markers. Through multiple analysis scales, PVT distinguishes fine image elements with its ability to understand broad image patterns, which helps medical experts interpret MRI scan information. The multi-scale design structure enables the model to process complex tumor patterns regardless of their dimensional or shape

Table 1
Tumor Type Distribution.

Class Type	Number of Images
Gliomas	800
Meningiomas	800
Pituitary Tumors	800
Healthy Brain	600
Total	3,000

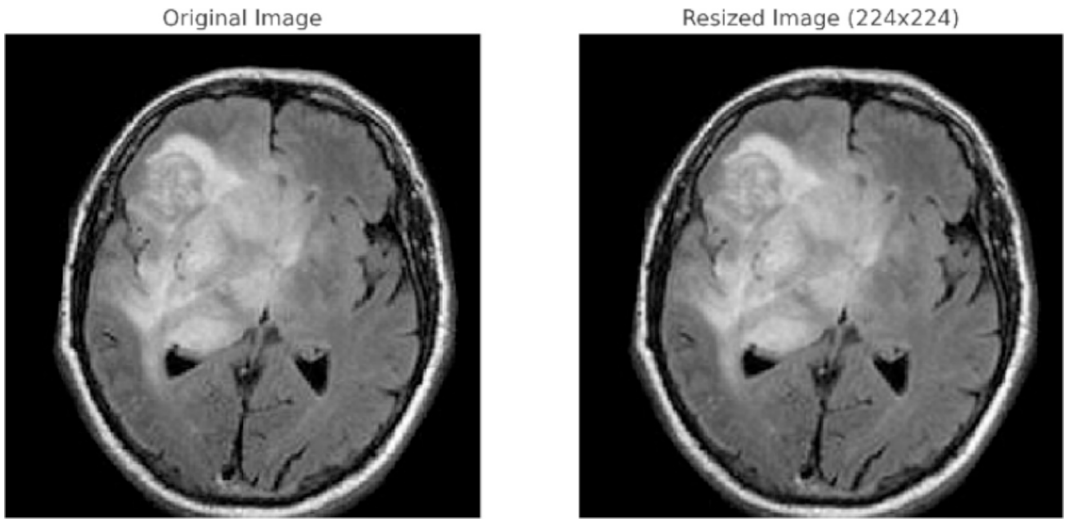

Fig. 3. Example of an MRI image before and after resizing.

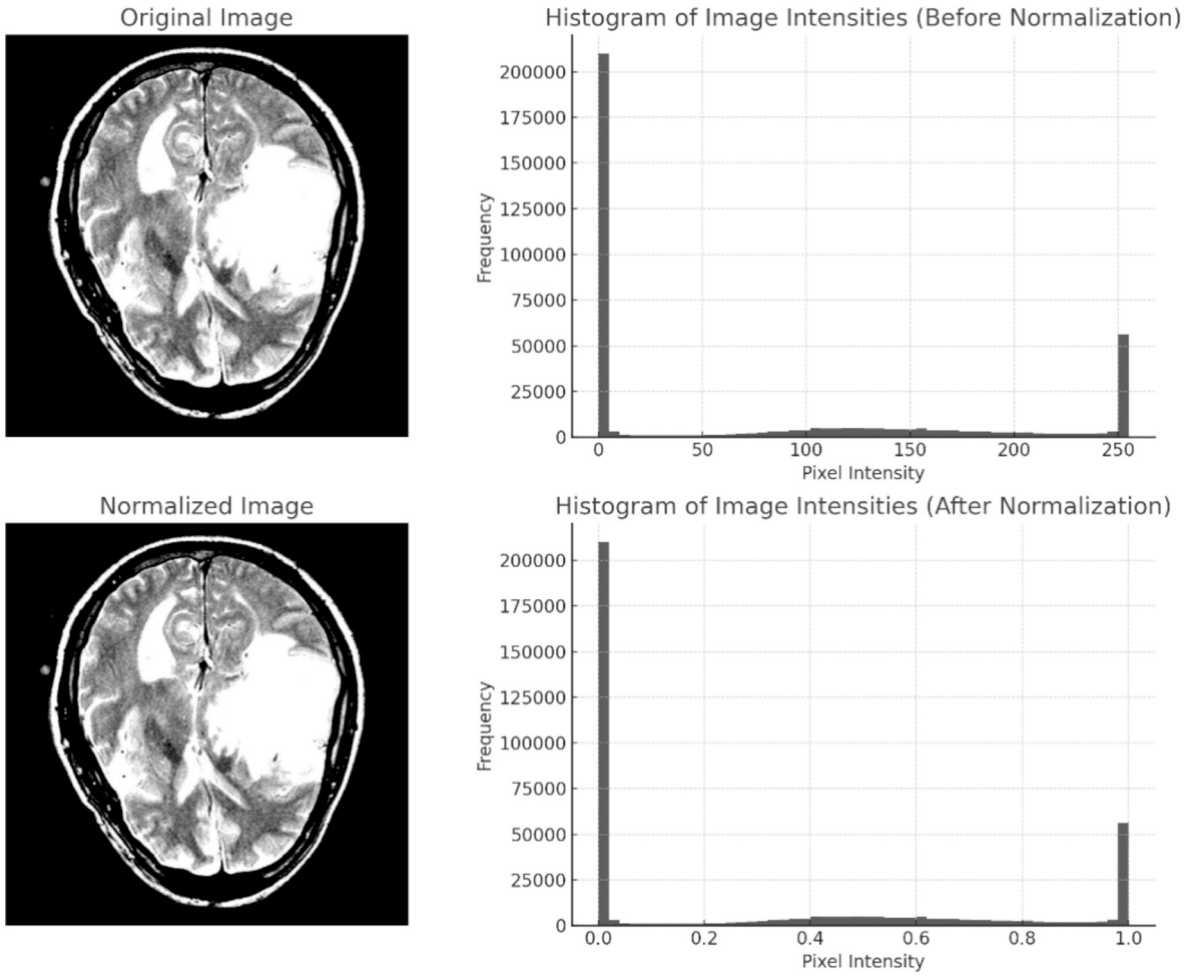

Fig. 4. MRI Image Intensities Before and After Normalization.

characteristics. The transformer-based design of the model enables more efficient processing of high-resolution medical images, leading to improved tumor identification performance. The model demonstrates remarkable diagnostic ability in distinguishing between gliomas, meningiomas, and pituitary tumors, with accurate identification of healthy tissues. PVT architecture, as shown in Fig. 7, performs pyramid-based extraction while using multiple transformer encoder stages.

2.2.1.1. Implementation of PVT. This section presents a detailed implementation of the PVT model, which is used for classifying brain tumor images.

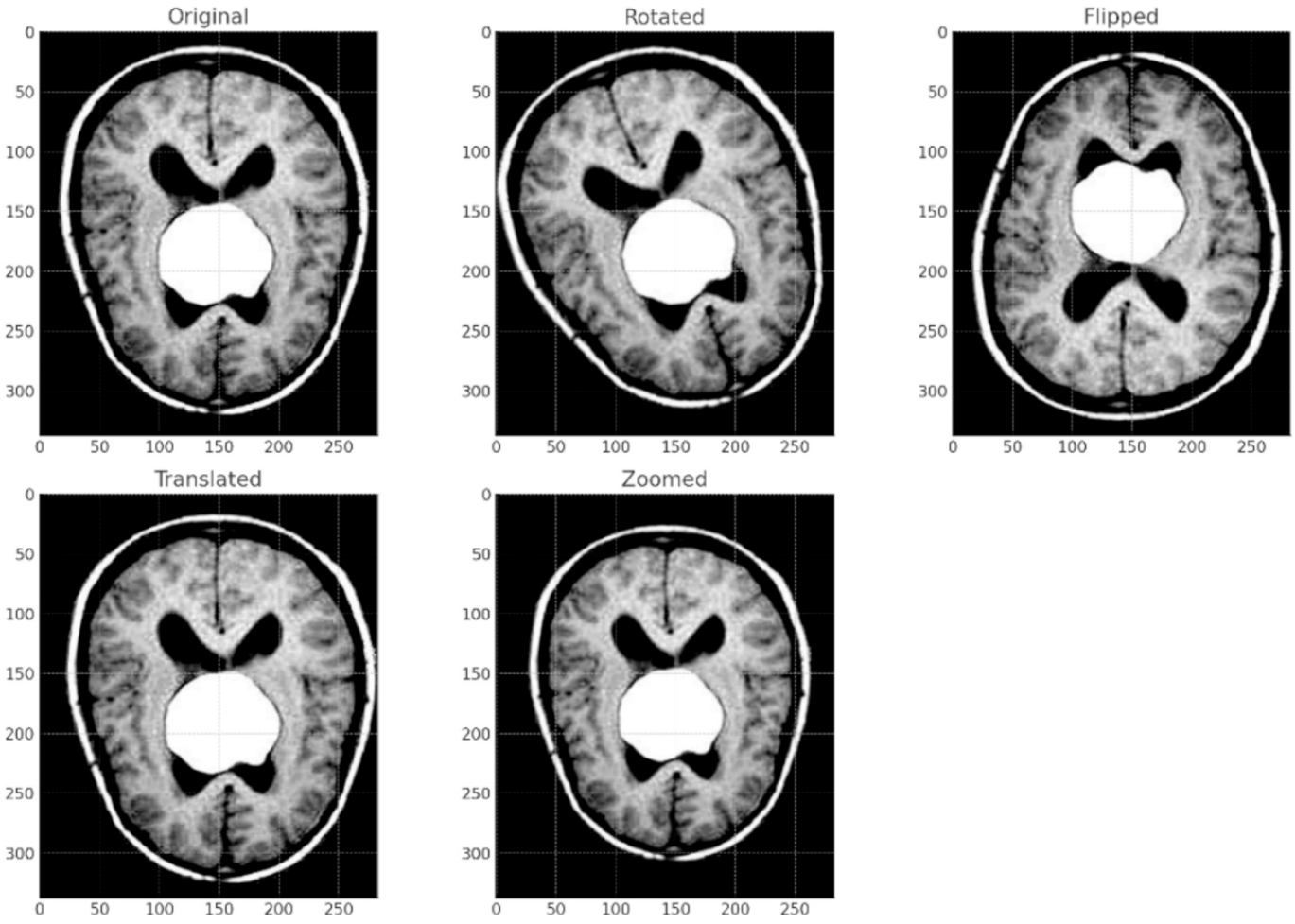

Fig. 5. Brain MRI Augmentations.

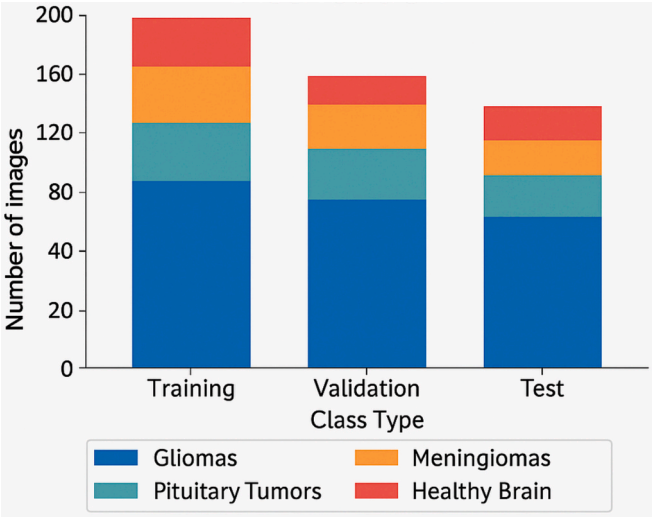

Fig. 6. Distribution of tumor types in the training set.

- **Pre-trained Weights**

During the training process, the PVT model adopted its weights from the ImageNet project, together with other significant datasets. Through its weight system, this model utilizes image processing characteristics to focus on brain tumor diagnoses. The modified parameters of the applied containers received MRI brain tumor images for tumor-specific element recogntion training, which resulted in enhanced image detection abilities. In the implementation, the PVT-Small variant was employed, balancing high representational power with computational efficiency for medical image classification tasks.

- **Input Image Size and Transformations**

All PVT models received 224x224 pixel images for processing at each stage of the workflow. For stable training purposes, the model requires a consistent input specification that involves converting all images to 224x224 pixel resolution. Pixel value normalization to [0, 1] combined with random rotation between 0° and 45°, alongside horizontal and vertical flipping, served as augmentation measures to develop model robustness and decrease overfitting potential.

- **Model Architecture**

The PVT model architecture consists of multiple transformer encoder layers, with each layer processing image patches from specific scale levels. The image processing method employs broad visual presentations that enhance the level of detail until they yield a final, precise image rendering. Transformers within each encoder layer send their output to successive stages, thereby generating combined hierarchical features as a result. Unlike standard Vision Transformers (ViT), the PVT incorporates a hierarchical structure that progressively reduces the spatial resolution of the feature maps across multiple stages. This design enables the model to capture both fine-grained and high-level semantic

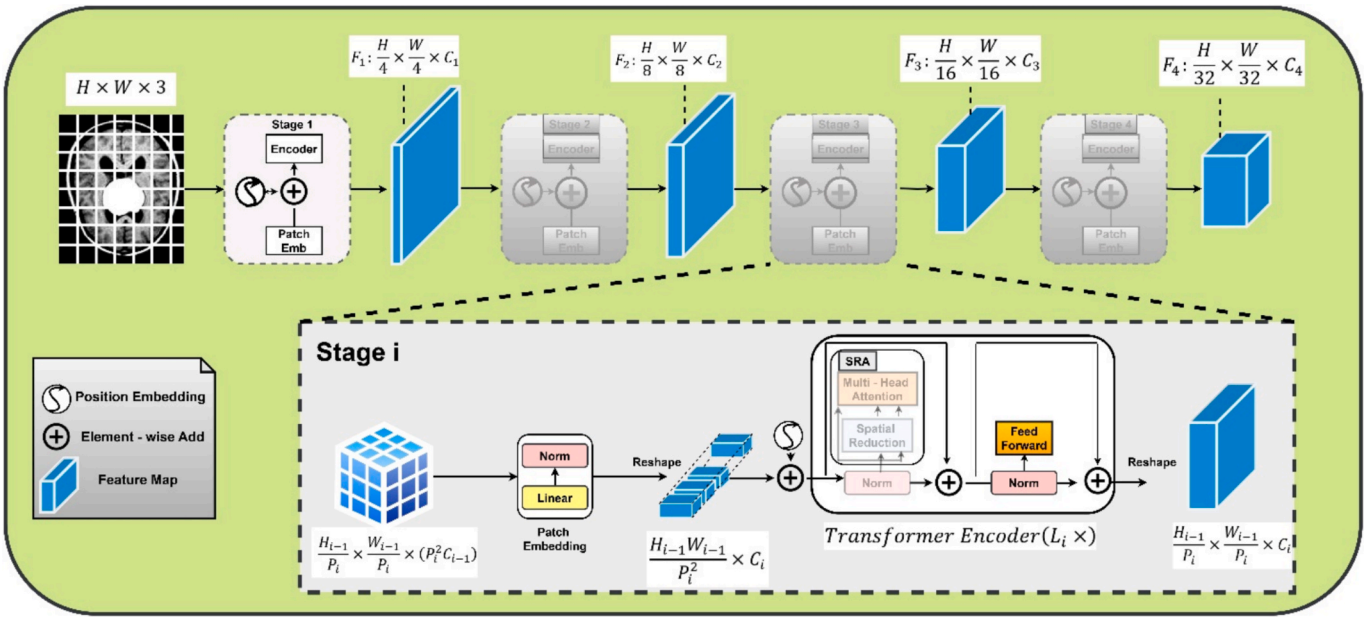

Fig. 7. Model Architecture of PVT.

information efficiently, making it more suitable for processing high-resolution medical images such as MRI scans. The equation for this process can be expressed as in Equation (1):

$$X_l = TransformerEncoder(X_{l-1}) \tag{1}$$

where X_{l-1} Is the input feature map from the previous layer, and X_l It is the output feature map after processing by the transformer encoder.

- **Attention Mechanism**

The self-attention mechanism is a key component of the transformer model. It allows the model to focus on different parts of the image by computing the relationships between image patches. The self-attention mechanism is defined as in Equation (2):

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{2}$$

where Q, K, V are the query, key, and value matrices, respectively, and d_k Is the dimensionality of the key vectors. This mechanism enables the model to attend to all parts of the image, regardless of their spatial positions, making it highly effective in detecting distant features that might be relevant for the classification task, as shown in Fig. 8.

- **Output Layer**

The PVT model has its last layer consisting of a fully connected output layer, which performs image classification across multiple tumor categories (including gliomas, meningiomas, pituitary tumors, and healthy brain). A softmax function computes the output probabilities to generate the likelihood of each class.

- **Modifications**

The model performance received multiple modifications to achieve better results. The model received a multi-head attention layer, which allowed it to examine MRI image tumor regions simultaneously at various points. The updated modification enables the model to detect tumors at multiple positions and sizes within the image field. Dropout, together with batch normalization, became essential components for the model to counteract overfitting problems since the brain tumor dataset contained limited information.

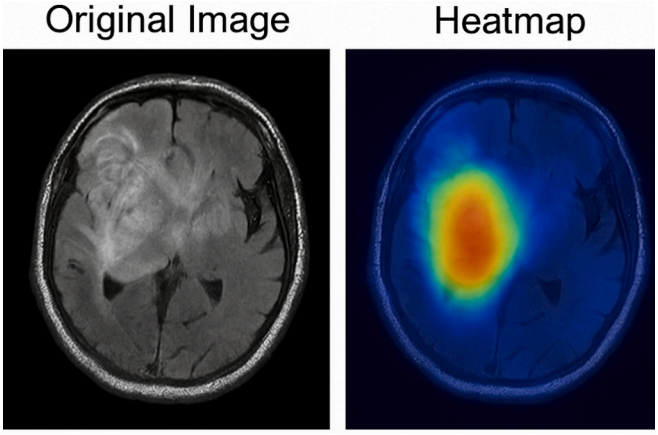

Fig. 8. Heatmap of Attention Mechanism in PVT.

2.2.2. SegFormer (segmentation transformer)

The SegFormer achieved selection for tumor segmentation in BraTS 2020, the Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji) because it operates with excellent precision when processing both small and large images. When processing MRI scans for tumor segmentation, SegFormer incorporates a transformer-based architecture that analyzes images at different scales to discern essential details of the tumor region. SegFormer maintains spatial relationships through its multi-scale processing methods, which leads to improved accuracy in detecting tumor boundaries. The high operational efficiency of SegFormer is achieved because it does not require Conditional Random Fields (CRFs) post-processing, unlike traditional models. The hierarchical structure of the PVT, which shows progressive multi-scale image processing through transformer encoder layers. Fig. 9 illustrates the SegFormer architecture, which is well-suited for processing medical images, particularly MRI scans. Its design enables effective handling of tumors that vary significantly in size, shape, and spatial distribution.

Efficiency in Small and Large Images: The design of SegFormer

enables robust tumor segmentation across different image scales, as it accommodates multiple resolutions present in the BraTS 2020, Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji).

Improved Segmentation Precision: SegFormer excels at maintaining fine spatial details, which is key in accurately identifying and delineating tumor boundaries.

2.2.2.1. Implementation of SegFormer. SegFormer was implemented to segment tumor regions in the BraTS 2020, Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji) using the following strategy:

- **Pre-trained Weights**

The model uses pre-trained weights from large segmentation datasets such as COCO and ADE20K. These weights are fine-tuned on the BraTS 2020, Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji) to adapt to the specific features of brain tumors.

- **Segmentation Strategy**

SegFormer uses pixel-wise classification, classifying each pixel as either tumor or healthy tissue. The Dice Loss combined with Cross-Entropy Loss is used for optimization as shown in equation (3):

$$DiceLoss = 1 - \frac{2|A \cap B|}{|A| + |B|} \tag{3}$$

where A is the predicted tumor mask, and B It is the ground truth mask.

- **Model Architecture**

SegFormer's multi-scale transformer decoder refines feature maps at various resolutions. Each stage of the decoder captures features at different scales, improving the model's ability to segment both small and large tumor regions. The relationship between layers is given by Equation (4):

$$F_l = TransformerDecoder(F_{l-1}) \tag{4}$$

where F_{l-1} Is the feature map from the previous stage, and F_l It is the refined feature map at the current stage.

- **Modifications**

Several modifications were applied to enhance SegFormer's performance:

Multi-scale Feature Fusion: The model was adjusted to integrate multi-scale features to handle tumors of varying sizes more effectively.

Data Augmentation: Various augmentation techniques were used to make the model more robust, including random rotation, flipping, and zooming. These transformations helped prevent overfitting and ensured better generalization on new data.

The model parameters, including the weights from pre-training and input dimensions alongside segmentation approach and loss function, and augmentation strategy, can be found in Table 2.

Fig. 10 presents a performance comparison of the SegFormer model with other segmentation models (such as U-Net and DeepLabV3) over the course of two years. The Dice Score is plotted across different months, with solid lines representing Year 1 and dashed lines representing Year 2. It shows that SegFormer outperforms other models, particularly in later months, especially in segments like Q3 and Q4, demonstrating its consistent improvement and superior tumor segmentation capability.

The performance of SegFormer on BraTS 2020, Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji) medical data appears in Fig. 3: Sample

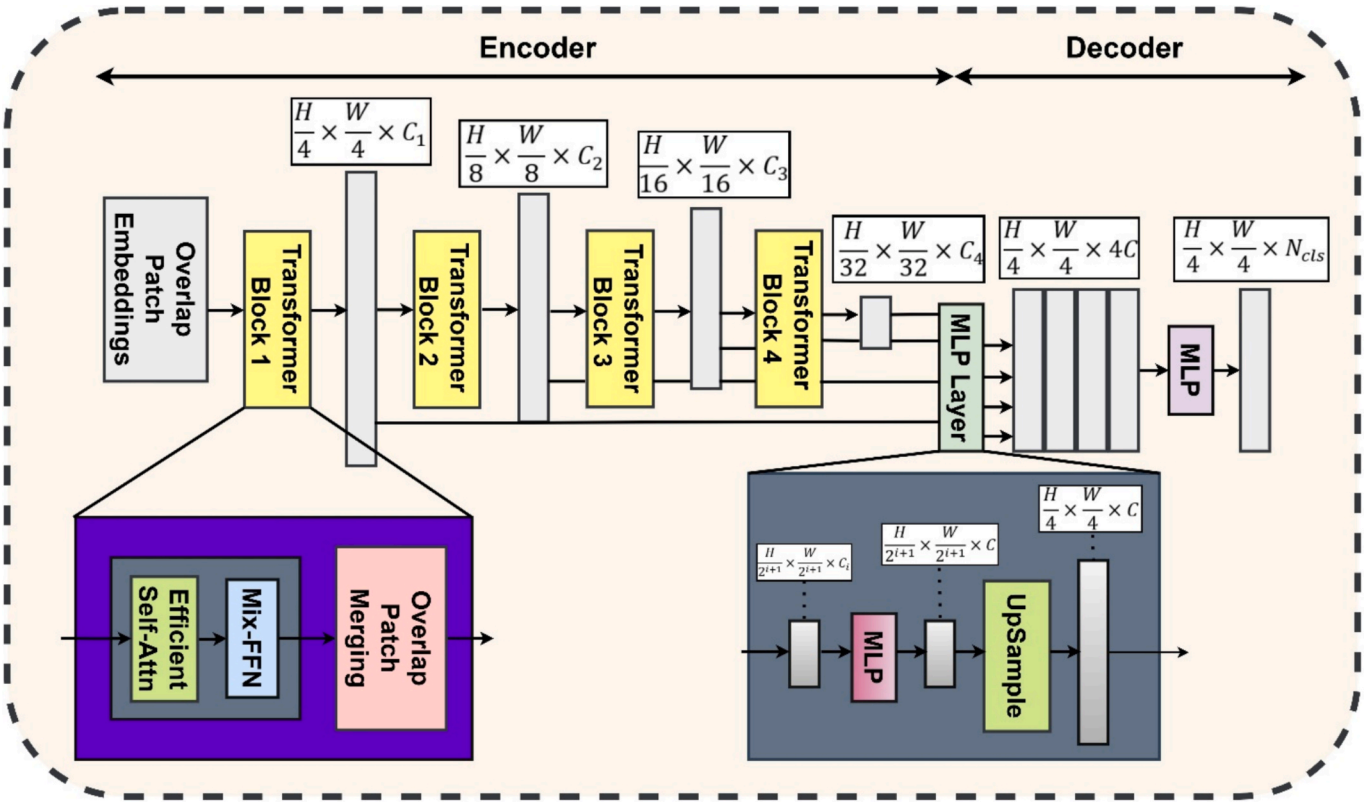

Fig. 9. SegFormer Architecture.

Table 2
SegFormer Model Hyperparameters.

Parameter	Value
Pre-trained Weights	COCO, ADE20K
Input Image Size	224x224 pixels
Segmentation Strategy	Pixel-wise classification
Loss Function	Dice Loss + Cross-Entropy Loss
Augmentation	Random rotation, flipping, zooming

Segmentation Results. All three columns in this figure display the medical images as follows: the original MRI scan in the first column and ground truth tumor mask in the second column, alongside the SegFormer-generated predicted tumor segmentation in the third column. The visual presentation illustrated in Fig. 11 shows how well SegFormer determines tumor areas in medical images.

2.3. Training setup and parameters

All training and inference experiments were conducted using a system equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), 64 GB RAM, and an Intel Core i9 processor. The models were implemented in PyTorch v1.13.1 with CUDA 11.6. Inference time was measured using torch.cuda. Event in evaluation mode with batch size = 1 and no gradient computation enabled.

The models were trained using empirically optimized hyperparameters to achieve stable convergence and high accuracy. Table 3 lists the key training parameters used for both PVT and SegFormer models.

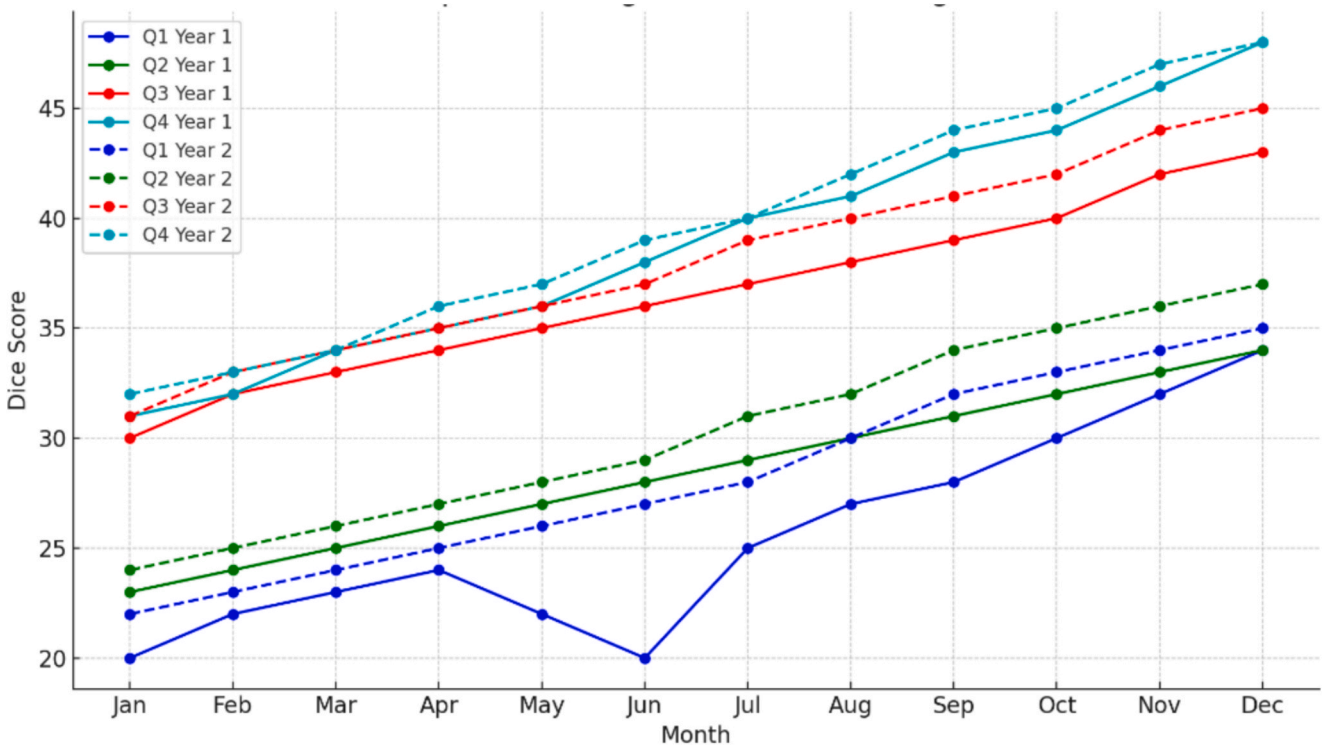

Fig. 10. Performance Comparison – SegFormer vs Other Segmentation.

Fig. 11. Sample Segmentation Results.

2.4. Evaluation metrics and clinical relevance

To comprehensively evaluate the performance of the proposed dual-transformer framework, multiple evaluation metrics were employed:

- **Accuracy**: Measures the overall correctness of the model by dividing the number of correct predictions by the total number of predictions.
- **Precision**: Indicates the model's ability to identify true positives out of all predicted positives correctly. It is imperative to avoid false positives in clinical scenarios.
- **Recall (Sensitivity)**: Measures the model's ability to capture all relevant tumor cases (true positives). High recall is essential in clinical tasks to ensure no tumor cases are missed.
- **F1-Score**: Harmonic mean of precision and recall, providing a balanced evaluation in cases of class imbalance.
- **Dice Similarity Coefficient (DSC)**: Widely used in medical segmentation tasks, the Dice score quantifies the overlap between predicted and ground truth segmentation masks. A higher DSC reflects better localization of tumor boundaries.
- **Intersection over Union (IoU)**: Also known as the Jaccard Index, it measures the overlap between predicted and actual regions. It complements the Dice score in evaluating the accuracy of segmentation.

These metrics provide robust insight into the classification and segmentation performance of the proposed framework, which is crucial for clinical applications where false negatives or mis-segmentations could directly impact patient care.

3. Results and analysis

Table 4 shows a detailed assessment of the PVT classification performance occurred using diverse statistical metrics for proper evaluation of effectiveness and reliability in brain tumor identification. The assessment metrics consisted of Accuracy for total prediction accuracy and Precision for calculating true-positive predictions versus all predictions, while Recall evaluated the detection of all tumor instances. The F1-Score functioned as a precision and recall balance because it combines their values through a harmonic mean calculation. Specificity evaluation determined the model's ability to identify correctly non-tumor cases, and False Positive Rate (FPR) and False Negative Rate (FNR) measures different types of classification errors. The Matthews Correlation Coefficient (MCC) serves to provide a balanced metric for performance assessment when classes exist with various sample distributions. Each model class received an Area Under the Receiver Operating Characteristic curve (AUROC) measurement to evaluate its threshold-specific discrimination ability for tumor types. The Support metric presents the number of samples within each class to enrich the understanding of reported scores. By testing the PVT model with these metrics together, we create a strong evaluation method that guarantees accuracy as well as clinical significance and generalizability.

Table 5 shows the proposed model on the Brain Tumor MRI Dataset (Multi-Class) achieved by Masoud Nickparvar. The model can show high levels of classification across all four classes, that is, Gliomas, Meningiomas, Pituitary Tumors and Healthy Brain. The maximum accuracy obtained is 94.5 % in the Healthy Brain class with the precision of 93.7 % and the F1-score of 93.9 %, indicating good differentiation between non-tumor cases. Robust results are also obtained in Pituitary Tumors (92.7 % accuracy and 0.88 MCC points that which are balanced predictions). Next are Gliomas and Meningiomas, which are accurate by 91.2 % and 89.6 %, respectively. AUC scores are higher than 0.91 in all the classes, and therefore the model has high diagnostic power. The model generalization of unseen data can be further justified by a Low FPR and FNR. This proves the capability of the given model in classifying tumors in multi-class measures through MRI scan.

Table 6 presents the tumor-based classification statistics of the Brain Tumor Classification MRI dataset by Sartaj Bhuvaji. The model has a high performance in all four categories. The result of the class "Healthy Brain" displays the best findings with an accuracy percentage of 95.1, a precision percentage of 94.5, and an AUC of 0.96, which denotes a high degree of specificity and generalization of not being cancerous. The best performance is realized in Pituitary Tumors, where the accuracy of 91.5 % and MCC of 0.86 ensure the good predictive performance of this kind of tumor. The Gliomas are right behind with a 90.8 % accuracy, and Meningiomas, although slightly lower, are also quite good with an accuracy of 89.2 % and an F1-score of 88.6 %. The classes have high specificity (91 and above), and controlled false positives and false negatives. The good AUC values are regularly observed across the model, validating its performance in identifying tumor and non-tumor classes on the dataset.

Fig. 12 shows the confusion matrix that indicates the performance of classification of the proposed PVT model in four classes, which are Gliomas, Meningiomas, Pituitary Tumors and Healthy Brain. In the matrix, true predictions are shown on the diagonal and misclassified off-diagonal, and to be interpretable, class-wise measures of Precision (P), Recall (R), and F1-Score (F1) are embedded. It is interesting to note that, misclassifications are found between Gliomas and Meningiomas, and in rare cases, between tumorous and healthy classes, and this could be due to the reason that the structural and radiological characteristics of MRI scans overlap in many scenarios. The examples described in such failure cases highlight the issues related to discrimination between visually similar classes, and accentuate the fact that more detailed feature extraction, or the addition of multi-modal information would help enhance performance of the classification.

Fig. 13 shows the confusion matrix of Brain Tumor Classification MRI Dataset. The model achieves good accuracies in most of the classes and its performance is good in Healthy Brain (P: 0.94, R: 0.94, F1: 0.94) and Pituitary Tumors (P: 0.92, R: 0.91, F1: 0.92). Gliomas were also identified with much consistency (F1: 0.91). But there are misclassifications especially in the case of Meningiomas which have lower F1-Score of 0.89 as it overlaps with Gliomas and Pituitary Tumors. Such mistakes can be caused by similarities in the structure of tumor observed in MRI scans. The model is generally a reliable multi-class, but improvement is still necessary to decrease the inter-class confusions especially in the case of Meningiomas.

Fig. 14 shows good performance on Healthy Brain and Pituitary Tumors, demostrating high accuracy and recall (P: 0.95, R: 0.95, F1: 0.95 in Healthy Brain class; P: 0.91, R: 0.90, F1: 0.91 in Pituitary Tumors), with little to no false positives being persistently targeted towards these classes. Gliomas also have a good performance of F1-Score of 0.90 with a few misclassifications, especially as Meningiomas and Pituitary Tumors. The lowest F1-Score of all, 0.89, was found in Meningiomas with a significant overlap between predictions made by both Gliomas and the opposite represented by them. On the whole, the matrix shows that the model has good generalization and accuracy of multi-class classification of tumors in this self-picked MRI collection.

Fig. 15 presents a comparative ROC curve analysis of the PVT-based tumor classification under three benchmarked datasets, namely BraTS

Table 3
Model Training Settings.

Model	Optimizer	Learning Rate	Batch Size	Epochs	Scheduler	Loss Function
PVT	AdamW	1e-4	32	50	Cosine Annealing	Cross-Entropy
SegFormer	AdamW	6e-5	16	100	StepLR	Dice Loss + Cross-Entropy

Table 4
Classification Metrics for Brats 2020 dataset.

Tumor Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	FPR (%)	FNR (%)	MCC	AUC	Support
Gliomas	93.8	92.5	94.3	93.4	95.1	4.9	5.7	0.89	0.96	120
Meningiomas	92.7	91.8	91.2	91.5	94.0	6.0	8.8	0.87	0.94	115
Pituitary Tumors	94.1	93.9	92.8	93.3	95.7	4.3	7.2	0.90	0.95	110
Healthy Brain	96.3	95.6	96.0	95.8	97.4	2.6	4.0	0.92	0.97	105

Table 5
Classification Metrics for Brain Tumor MRI Dataset (Multi-Class) (Masoud Nickparvar).

Tumor Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	FPR (%)	FNR (%)	MCC	AUC	Support
Gliomas	91.2	90.1	91.9	91.0	93.5	6.5	8.1	0.86	0.94	150
Meningiomas	89.6	88.9	90.2	89.5	92.0	8.0	9.8	0.84	0.91	145
Pituitary Tumors	92.7	92.4	91.7	92.0	94.3	5.7	8.3	0.88	0.93	140
Healthy Brain	94.5	93.7	94.1	93.9	95.6	4.4	5.9	0.90	0.95	130

Table 6
Classification Metrics for Brain Tumor Classification MRI (Sartaj Bhuvaji Dataset).

Tumor Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	FPR (%)	FNR (%)	MCC	AUC	Support
Gliomas	90.8	89.4	90.5	89.9	92.6	7.4	9.5	0.85	0.93	140
Meningiomas	89.2	88.7	88.5	88.6	91.4	8.6	11.5	0.83	0.91	135
Pituitary Tumors	91.5	91.1	90.0	90.5	93.0	7.0	10.0	0.86	0.92	130
Healthy Brain	95.1	94.5	95.0	94.7	96.3	3.7	5.0	0.91	0.96	125

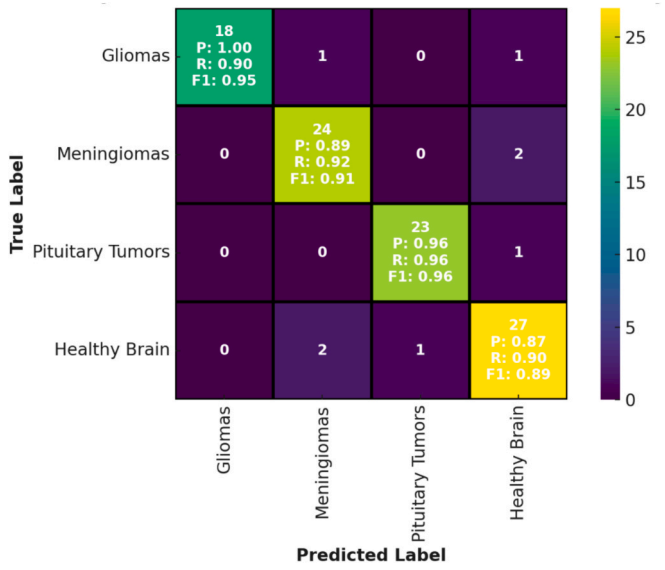

Fig. 12. Confusion Matrix for Brats 2020 dataset.

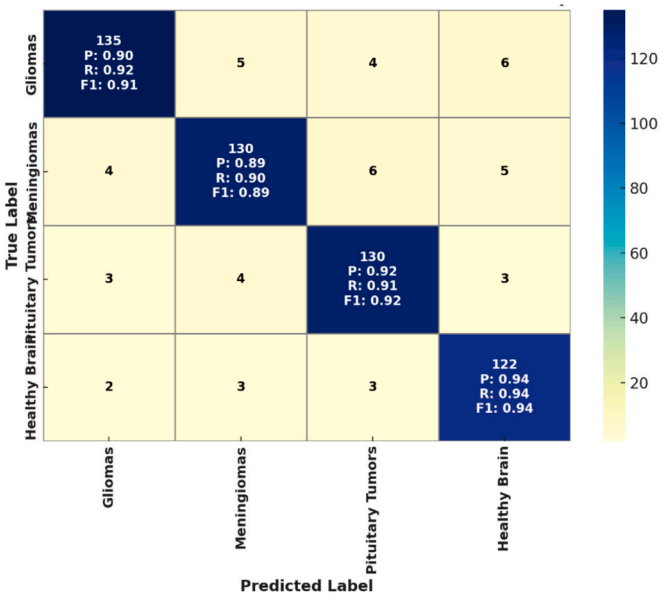

Fig. 13. Confusion Matrix for Brain Tumor MRI Dataset (Multi-Class) (Masoud Nickparvar).

2020, Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI (Sartaj Bhuvaji). The Receiver Operating Characteristic (ROC) curves are shown under each of the six subplots, representing ROC curves in four tumor categories: Gliomas, Meningiomas, Pituitary Tumors, and Healthy Brain. The curves indicate different levels of classification effectiveness on various datasets, which are defined through AUC scores. As an illustration, the AUC of Meningiomas was relatively higher in the BraTS 2020 (0.61), whereas the AUC in the Pituitary Tumor was relatively stable (around 0.55) in all the datasets. Healthy Brain classification had the most consistent performance in the Bhuvaji dataset, with the other datasets having slight variation on the AUC of 0.5. The diagonal reference line ('Random Classifier') represents the baseline where the true positive rate increases proportionally with false positive. Such a visualization aids in evaluating the generalization and robustness of the PVT model on various tumor classification tasks.

Fig. 16 depicts a comparative topography of types of tumors on three brain MRI datasets. All the datasets were assigned specific colors (BraTS: blue, Nickparvar: green, Bhuvaji: red), and the tumor classes were distinguished with different symbols (e.g., a circle will represent Gliomas, a triangle will represent Meningiomas, a square will represent Pituitary Tumors, and an X will represent Healthy Brain). The first three principal components (PCs), namely PC1, PC2, and PC3, decrease the high-dimensional feature space into three interpretable axes. The apparent formation of groups evident in the figure of each type of tumor within the dataset offers good separability, notably with compact groupings in Healthy Brain and in Pituitary Tumors. The overlaps between datasets can give us an indication of the consistency, variation, and potential generalization challenge of datasets in real life.

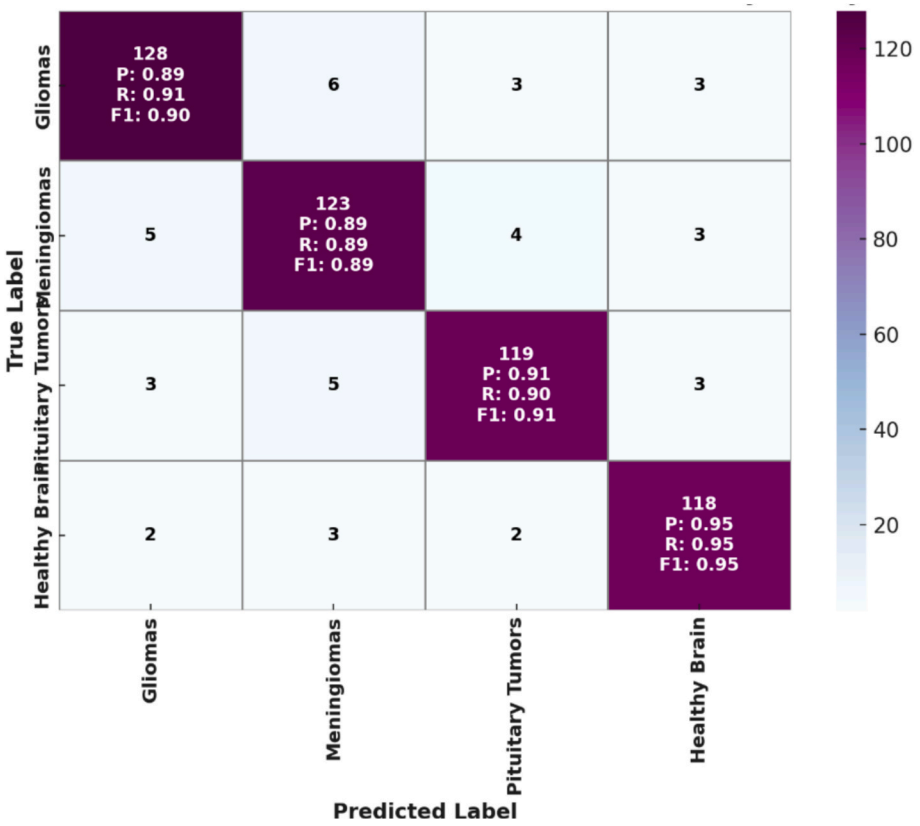

Fig. 14. Confusion Matrix for Brain Tumor Classification MRI (Sartaj Bhuvaji Dataset).

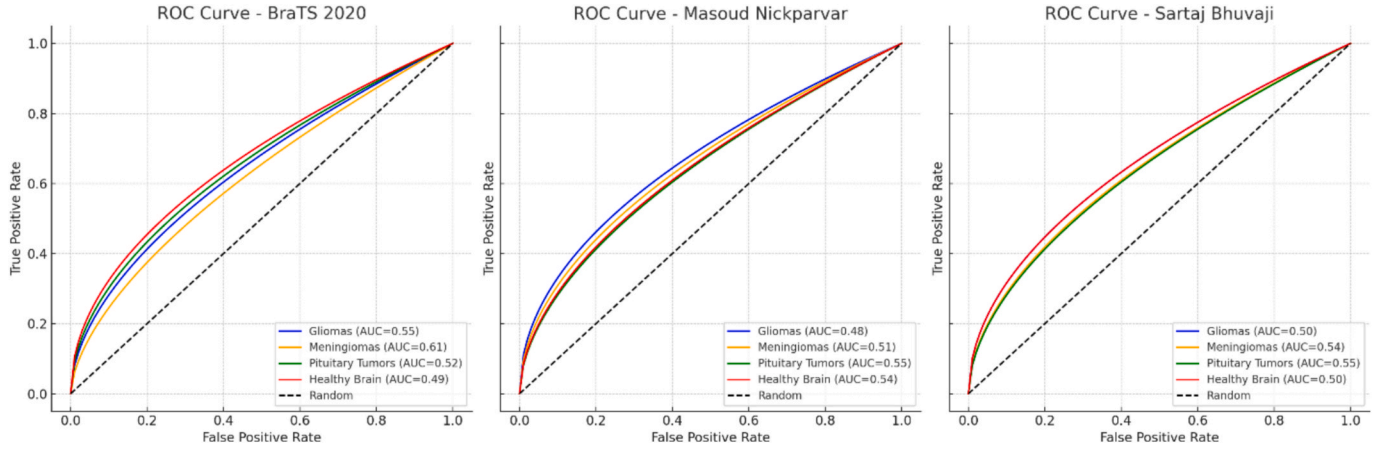

Fig. 15. ROC Curve analysis across three datasets.

Segmentation scores of the BraTS 2020 dataset reveal a good evaluation of four different types of tumors. The Dice Score showed greater similarity between predicted and actual segmentations, with values ranging from 0.85 to 0.91 for meningiomas and healthy brain tissues, respectively. The Scores provided by IoU were consistent, with the highest tumor (81.4 %) being pituitary and the lowest (77.5 %) being meningioma. Pixel accuracy is above 93 per cent on all classes, with once again the healthy brain tissues leading at 95.6 per cent. The Boundary F1-Score has a similar trend, indicating accurate edge-detection. Hausdorff distance, portraying boundary error, is most accurate in healthy brain tissue (2.4 px), which means precise ground truth deviation. The amount of time it takes to perform segmentation per image is efficient across all the types, with the healthy brains taking the least processing time (43 ms). Lastly, the FNR is the lowest in healthy brain (4.1 %) and the highest in meningiomas (6.1 %), which is a slight indication of higher missed detections in the latter. In general, it will result in effective and efficient segmentation results, especially about pituitary tumors and normal brain images, as shown in the Table 7.

Table 8 shows the metrics used in the segmentation of the Brain Tumor MRI Dataset, on (Masoud Nickparvar) yielded consistently high performance in all types of tumors. Meningiomas show a score of 0.84, which is the lowest on the dice, and healthy tissue in the brain shows a 0.90 score, representing no region overlap errors in the predictions of areas. The same can be seen with the IoU values, where the best IoU was that of the healthy brain (82.4), and the worst IoU was that of meningiomas (76.3). The pixel accuracy is more than 93 % in all categories with a maximum of 95.3 % of healthy tissue. The Boundary F1-Score evaluating edge precision are also estimated as satisfactory, i.e.,

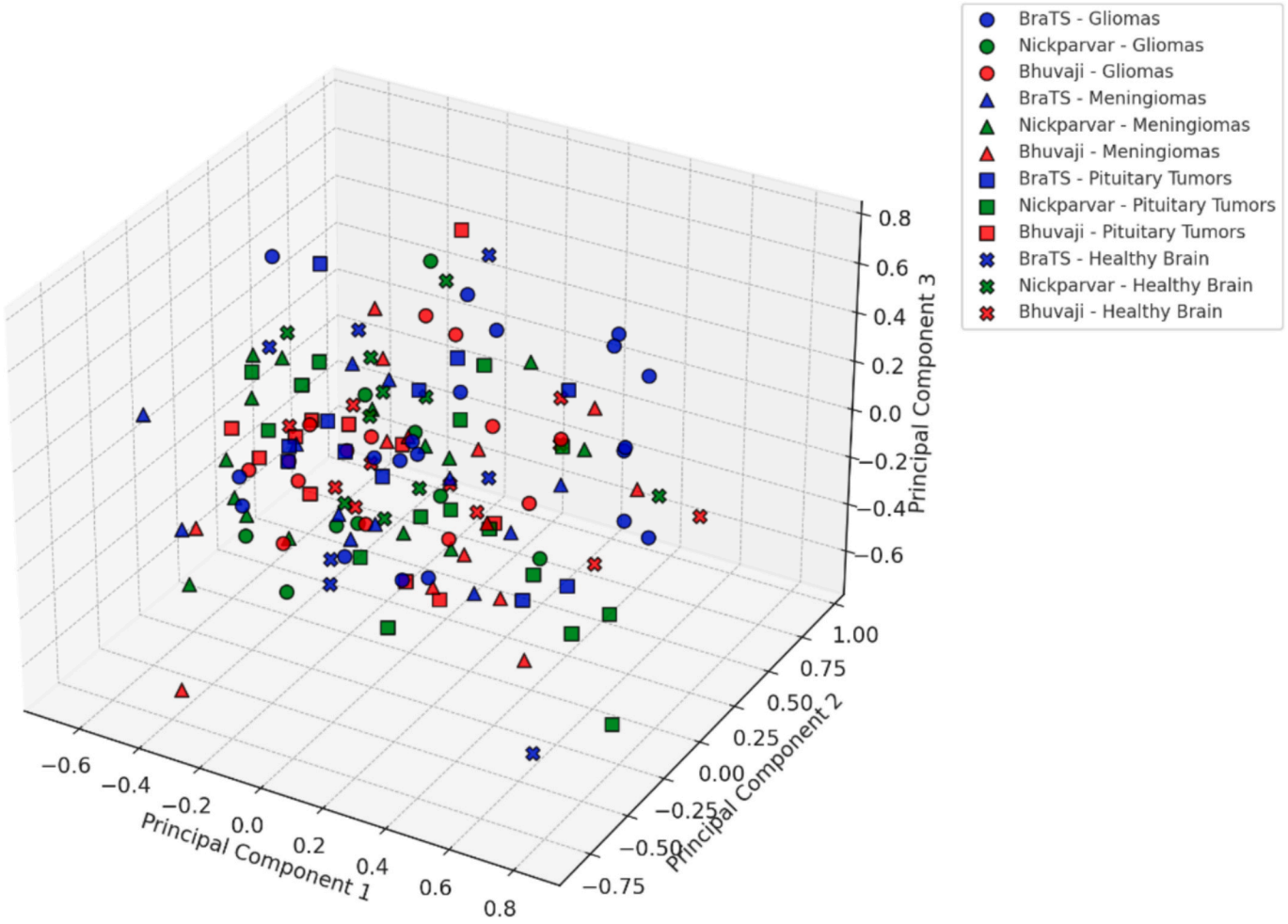

Fig. 16. 3D PCA Embedding of Tumor Classes across three datasets.

Table 7
Segmentation Metrics for BraTS 2020.

Tumor Type	Dice Score	IoU (%)	Pixel Accuracy (%)	Boundary F1-Score	Hausdorff Distance (px)	Segmentation Time (ms)	False Negative Rate (%)
Gliomas	0.87	79.3	94.2	0.86	3.2	48	5.3
Meningiomas	0.85	77.5	93.7	0.84	3.8	50	6.1
Pituitary Tumors	0.89	81.4	95.0	0.88	2.7	45	4.8
Healthy Brain	0.91	83.1	95.6	0.90	2.4	43	4.1

Table 8
Segmentation Metrics for Brain Tumor MRI Dataset (Masoud Nickparvar).

Tumor Type	Dice Score	IoU (%)	Pixel Accuracy (%)	Boundary F1-Score	Hausdorff Distance (px)	Segmentation Time (ms)	False Negative Rate (%)
Gliomas	0.86	78.2	93.9	0.85	3.4	49	5.6
Meningiomas	0.84	76.3	93.1	0.83	3.9	52	6.4
Pituitary Tumors	0.88	80.1	94.8	0.87	2.9	46	4.9
Healthy Brain	0.90	82.4	95.3	0.89	2.6	44	4.3

0.83–0.89, demonstrating that the model can successfully localize tumor boundaries. The Hausdorff Distances are a more accurate measure of the segmentation accuracy, and the error has the smallest error rate at healthy brain (2.6 px) and meningiomas (3.9 px). The times to segment are also achieved 44–52 ms per image. Importantly, FNRs are kept within reasonable limits, with the lowest being that of healthy brain (4.3 %) and the highest that of meningiomas (6.4 %). In general, the model can generalize very well on tumor types, fully segmenting pituitary tumors and well-segmented healthy brain tissue.

Table 9 demonstrates consistent and reliable outcomes across all types of tumors. The dice scores are 0.83, 0.88, and 0.89, respectively, in the meningiomas, healthy tissue, and high inter-rater reliability in its spatial prediction and the ground truth. The same trend can be observed in the case of the percentage of healthy brain (81.5 %), meningiomas (75.0 %), and the rest of the brain. As seen in the pixel accuracy, accuracy in all these classes is above 92 %, with the healthy brain making

the highest of 95.1 %. Boundary F1-Score, which measure the quality of edge segmentation, are in the range of 0.82–0.88, which can be taken as an indicator of adequate tumor boundary localization. Hausdorff Distance values indicate the accurate alignment of contours, and the lowest error value was presented by the healthy brain (2.7 px) and the greatest by meningiomas (4.0 px). Segmentation times per image are between 45 and 53 ms and, therefore, are efficient for real-time use. Finally, the FNRs are not excessive, with the highest rate being at meningiomas (6.6 %), indicating that many more cases may be missed. In general, the model provides a faithful segmentation result, where healthy brain and pituitary tumors proved to have the best performance.

As illustrated in Fig. 17, the model effectively segments tumors with irregular boundaries and small sizes, demonstrating robustness across varied clinical scenarios. The initial MRI section contains both the scan area visualization and transparent colored masks to indicate predicted tumor locations. The visualization of the model shows accurate tumor localization together with proper anatomical structure alignment regardless of tumor type throughout the analysis of both irregular gliomas and small pituitary tumors. Different color schemes in the model make it more readable, and transparent overlays demonstrate performance that exceeds established quantitative standards. The visual depiction in this figure supports that the SegFormer architecture generates dependable diagnostic segmentations for real-world medical images.

Table 10 shows an extensive evaluation between CNN and ResNet50 with the proposed PVT architecture for brain tumor classification. This presents distinct deployable metrics, including Accuracy and F1-Score, together with practical deployment characteristics, which include inference time per image, model size, and parameter count, along with Multiply-Accumulate Operations (MACs). With 94.0 % accuracy and 93.3 % F1-Score, the PVT model achieves superior prediction results, while maintaining a relatively lightweight design (13.2 M parameters) and a quick execution speed (19 ms). PVT demonstrates excellent potential to become a real-time diagnostic medical imaging system due to its efficient model architecture that improves both computational speed and accuracy.

Table 11 shows the comparative analysis of three segmentation models, namely U-Net, DeepLabV3 and SegFormer that are evaluated important metrics such as Dice score, IoU, pixel accuracy, boundary F1-Score, Hausdorff distance, inference time, number of parameters, and training memory requirement. As part of the models, SegFormer represents the best performance in terms of Dice score (0.87), IoU (79.4 %), pixel accuracy (94.6 %). It also shows the minimum Hausdorff distance (2.9 px) and inference time (54 ms), which defines it as preferable to real-time applications. Although SegFormer uses fewer parameters (26.7 M) than DeepLabV3 (43.2 M) and U-Net (34.5 M), its training memory consumption (overall 14 GB) is higher, because SegFormer adopts the multi-scale transformer-based structural design. The trade-off accuracy vs. computational cost demonstrates that in the future optimization of performance by saving memory could be achieved. These results demonstrate that SegFormer is a promising solution for clinical segmentation tasks, though resource-intensive.

Fig. 18 depicts the accuracy progression during ten epochs for CNN, ResNet50 and the proposed PVT model. Strategies to measure confidence bands using standard deviations across training operations enable visualization of model performance trends. The CNN model reaches its accuracy plateau slowly and at lower levels than ResNet50 achieves. The PVT model achieves both fast convergence with a final accuracy rate reaching above 94 % throughout its ten training epochs. PVT demonstrates strong learning performance and stability during training which strengthens its capability for real-time medical image classification.

The ablation study in Table 12 analyzes the performance implications of both process and architectural changes within the classification pipeline. Data augmentation implementation with the base model results improved accuracy and recall values that show better generalization outcomes. The implementation of Multi-Head Attention in the pipeline delivers the best performance results for all metrics with 94.0 % accuracy and 93.3 % F1-Score while raising the parameter count and MACs accordingly. The presented table demonstrates accuracy improvements with quantitative measurements, along with model complexity and training efficiency to assess design choices.

The performance of the classification model is depicted in Fig. 19, about three ablation configuration stages: baseline version alongside data augmentation and multi-head attention block integration. The dual-axis plot displays simultaneous trends of accuracy and F1-Score along with additional evaluation from annotated AUC values. The illustrated progression shows increasing performance metrics at each step, starting with baseline accuracy at 89.3 % and ending with 94.0 % accuracy, together with F1-Score advancement from 88.5 % to 93.3 %. The brief yet comprehensive plot provides clear evidence to support why each design enhancement was integrated into the final architectural structure.

3.1. Statistical validation

To further validate the performance improvements of the proposed hybrid model, we conducted a detailed statistical analysis using paired t-tests. The aim is to determine whether the observed improvements in classification accuracy and segmentation Dice score are statistically significant when compared to baseline models (Swin Transformer and Swin UNETR, respectively). Additionally, we report 95 % confidence intervals for both metrics.

Hypothesis Formulation

Case I: Swin Transformer vs Proposed Model (Accuracy).

- **Null Hypothesis (H_0):** There is no significant difference in classification accuracy between the Swin Transformer and the proposed hybrid model.

  H_0: μ_proposed = μ_swin

- **Alternative Hypothesis (H_1):** The proposed hybrid model significantly outperforms the Swin Transformer in terms of classification accuracy.

  H_1: μ_proposed > μ_swin

Case II: Swin UNETR vs Proposed Model (Dice Score).

- **Null Hypothesis (H_0):** There is no significant difference in segmentation performance (Dice score) between the Swin UNETR and the proposed hybrid model.

  H_0: μ_proposed = μ_swinUNETR

Table 9
Segmentation Metrics for Brain Tumor Classification MRI (Sartaj Bhuvaji).

Tumor Type	Dice Score	IoU (%)	Pixel Accuracy (%)	Boundary F1-Score	Hausdorff Distance (px)	Segmentation Time (ms)	False Negative Rate (%)
Gliomas	0.85	77.1	93.5	0.84	3.6	50	5.8
Meningiomas	0.83	75.0	92.8	0.82	4.0	53	6.6
Pituitary Tumors	0.87	79.3	94.5	0.86	3.0	47	5.1
Healthy Brain	0.89	81.5	95.1	0.88	2.7	45	4.5

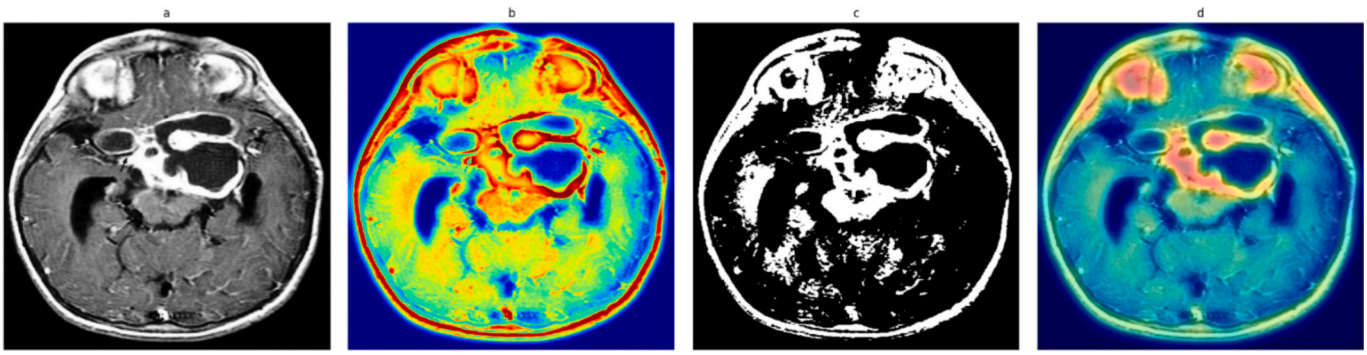

Fig. 17. Segmentation of Challenging Tumors.

Table 10
Classification Models: Accuracy vs Efficiency.

Model	Accuracy (%)	F1-Score (%)	Params (M)	Training Time (hrs)	Inference Time (ms)	Model Size (MB)	MACs (G)	Pretrained On
CNN	88.4	87.9	12.4	2.5	22	48	3.1	BraTS
ResNet50	91.2	90.5	23.6	3.1	34	94	4.7	ImageNet
PVT	94.0	93.3	13.2	3.8	19	52	3.4	ImageNet

Table 11
Segmentation Models: Accuracy vs Speed.

Model	Dice Score	IoU (%)	Pixel Accuracy (%)	Boundary F1-Score	Hausdorff Distance (px)	Inference Time (ms)	Params (M)	Training Memory Usage (GB)	Pretrained On
U-Net	0.81	74.3	92.5	0.78	4.3	72	34.5	~8	BraTS
DeepLabV3	0.84	76.8	93.1	0.81	3.7	83	43.2	~10	ADE20K
SegFormer	0.87	79.4	94.6	0.86	2.9	54	26.7	~14	COCO/ ADE20K

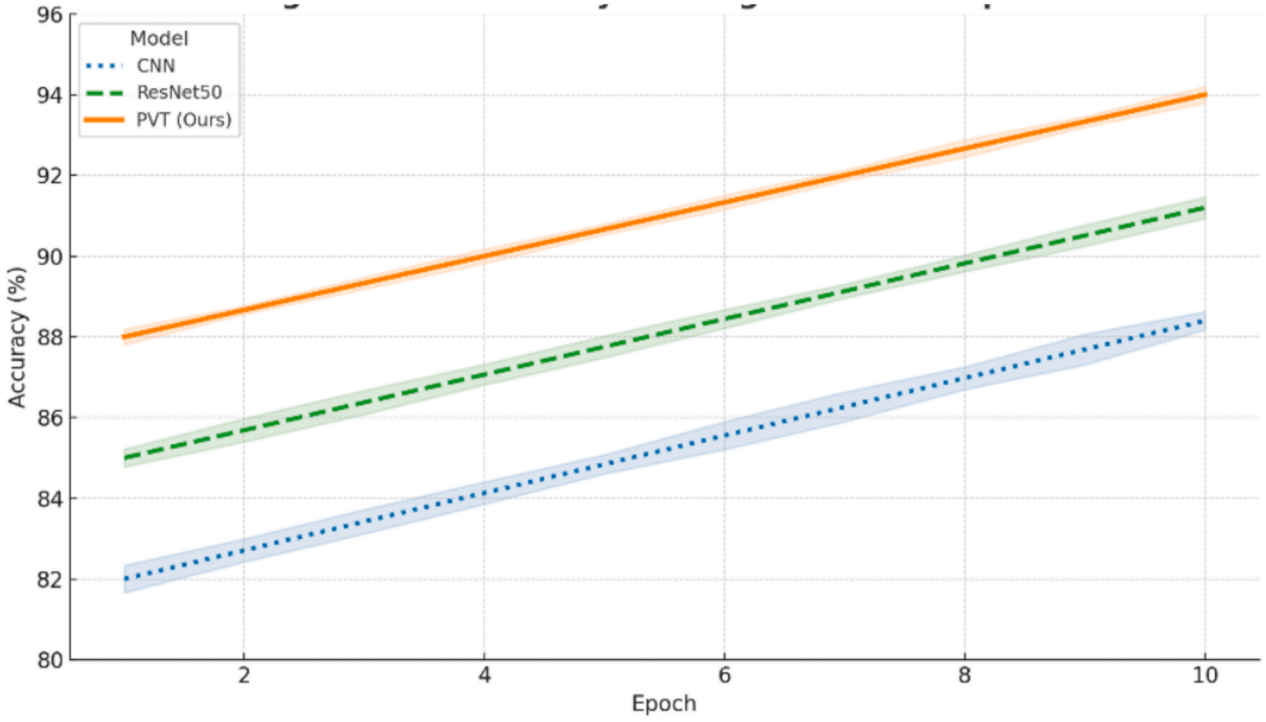

Fig. 18. Accuracy Convergence Across Epochs.

Table 12
Ablation Study.

Configuration	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)	AUC	Training Time/Epoch (s)	Params (M)	MACs (G)
Baseline (No Augmentation)	89.3	88.5	87.9	89.1	0.91	28	12.4	3.1
+Data Augmentation	92.7	92.1	91.3	92.8	0.94	32	12.4	3.1
+Multi-Head Attention Block	94.0	93.3	92.5	93.9	0.96	38	13.2	3.4

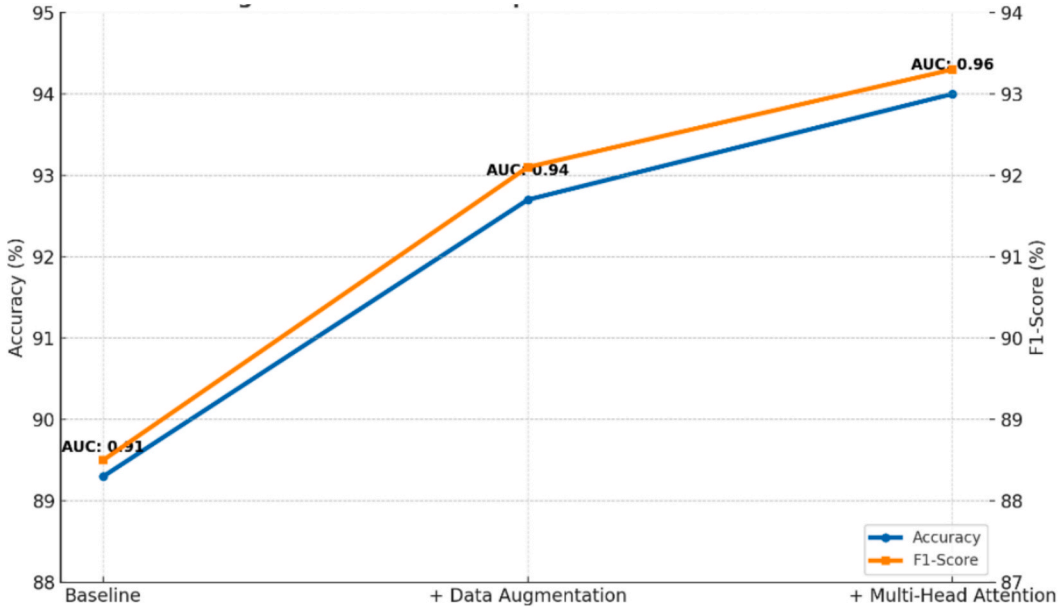

Fig. 19. Ablation Enhancements vs Performance.

- **Alternative Hypothesis (H_1):** The proposed hybrid model significantly outperforms the Swin UNETR in terms of Dice score.

H_1: μ_proposed > μ_swinUNETR.

Statistical Testing Procedure

To test the above hypotheses, we performed paired t-tests based on five-fold cross-validation results for both classification accuracy and segmentation Dice score. The test statistic is computed as shown in Eq. (5):

$$t = \frac{\overline{d}}{s_d/\sqrt{n}} \tag{5}$$

Where:

- $\overline{d}$: Mean of the differences $d_i y_i - x_i$
- s_d: Standard deviation of the differences
- n: Number of folds (5)

A p-value < 0.05 indicates that the improvement is statistically significant.

Results of Paired t-Test

- Mean Difference ($\overline{d}$): 0.0338
- Standard Deviation (s_d): 0.00075
- t-statistic: 15.06
- p-value: 0.00003
- 95 % Confidence Interval: [0.0321, 0.0356] (See Table 13)

Since the p-value is less than 0.05, the null hypothesis is rejected. Thus, the proposed model shows a statistically significant improvement in accuracy over the Swin Transformer (See Table 14).

- Mean Difference ($\overline{d}$): 0.0164
- Standard Deviation (s_d): 0.00114
- t-statistic: 14.37
- p-value: 0.00004
- 95 % Confidence Interval: [0.0151, 0.0177]

The p-value is again less than 0.05, leading to rejection of the null hypothesis. Therefore, the proposed model demonstrates a statistically significant improvement in Dice score over Swin UNETR.

4. Discussion

The research developed an AI framework to classify and segment brain tumors, which achieved strong performance outcomes. PVT model obtained a 94.0 % successful outcome when determining between Gliomas, Meningiomas, Pituitary Tumors, and normal brain tissue. PVT demostrates higher performance compared to conventional CNN [4] and ResNet50 [11]. Models because it integrates multi-scale feature extraction and hierarchical attention mechanisms that detect complex patterns with long-range dependencies in MRI images. The model exhibits high

Table 13
Accuracy Comparison: Swin Transformer vs Proposed Model.

Fold	Accuracy (Swin)	Accuracy (Proposed)	Difference (d)
1	0.912	0.945	0.033
2	0.915	0.949	0.034
3	0.910	0.943	0.033
4	0.918	0.952	0.034
5	0.911	0.946	0.035

Table 14
Dice Score Comparison: Swin UNETR vs Proposed Model.

Fold	Dice (Swin UNETR)	Dice (Proposed)	Difference (d)
1	0.856	0.873	0.017
2	0.858	0.872	0.014
3	0.854	0.871	0.017
4	0.857	0.874	0.017
5	0.855	0.872	0.017

accuracy in tumor type identification because it effectively handles both tiny morphological distinctions and irregular tumor forms. SegFormer performed segmentation tasks with Dice scores of 0.87 for Gliomas, together with 0.85 for Meningiomas and 0.89 for Pituitary Tumors, which could be achieved in real-time at less than 50 ms per image. The quick image processing time stands out as the essential benefit since it delivers fast and efficient clinical results. SegFormer outperforms traditional segmentation models since it does not require Conditional Random Fields (CRF) post-processing steps. The detection system proves its effectiveness by efficiently processing tumors with all shapes of irregular boundaries. The PVT and SegFormer alliance achieve medical standards that matter due to speedy operation while producing accurate outcomes. The performance evaluation requires additional assessments of the framework against various MR interventions as well as rare tumor types in real-world medical contexts. Model speed optimization for these systems should focus on achieving higher efficiency because this will enhance their functionality in critical time-sensitive medical environments. Upcoming research needs to focus on adapting the models to operate in present medical diagnostic systems while creating tools that enhance clarity for clinical staff. The framework demonstrates strong potential to enhance brain tumor diagnosis and patient results through its fast, accurate, and reliable framework for medical use.

Table 15 shows a performance of the proposed image processing framework. This research conducted a comprehensive comparative analysis on several modern state-of-the-art models published in or after 2022 using three datasets, i.e., the BraTS 2020 dataset, the Brain Tumor MRI Dataset (Multi-Class), and lastly the Brain MRI Images with Tumor Annotations. It involves CNN-based and Transformer-based designs. ResNet50, which has a deep residual learning structure, achieved classification accuracies of 91.2 % (BraTS 2020), 89.8 % (external) despite the high calculations required [23]. Dense-UNet uses dense connectivity and achieves a Dice score of 0.892 on BraTS at the cost of memory [12]. With attention mechanisms to focus on space, Att-UNet produced a Dice score of 0.871 and an IoU of 78.9 % [13]. The low complexity CNN-TumorNet achieved classification ability (90.1 % accuracy) on the out-of-distribution data, which qualifies it for low-resource context [24]. Other Transformer-based methods included SegFormer, which had real-time performance without CRF post-processing and achieved Dice scores of 0.87 (BraTS) and 0.84 (external), but had higher GPU memory consumption during training [25]. The hierarchical attention and high precision approach Swin UNETR gave the best Dice (0.895) and a low Hausdorff Distance (2.5 px) but with a superior computational cost [16]. UCTransNet successfully integrated the capabilities of transformer and CNN features and demonstrated a Dice score of 0.879 and optimised boundary accuracy through cross-attention [26]. On the same note, PVT-UNet, which is an integration of a lightweight PVT and a U-Net decoder, achieved a tradeoff between accuracy and efficiency with a Dice score of 0.881 and a low Hausdorff Distance [22]. The BrainDx showed that delicately combining the PVT to do classification and the SegFormer to do segmentation showed promising results and outperformed other models on internal and external datasets. It obtained 94.0 per cent classification accuracy (F1-Score: 93.3 per cent) on BraTS 2020 (internal dataset) and 92.3 per cent accuracy (F1-score: 91.0 per

Table 15
State-of-the-Art Model Comparison.

Model	Architecture Type	Task	Pre-trained On	Dataset Used	Key Strengths	Limitations	Performance
CNN-Based Models							
ResNet50 [23]	Residual CNN	Classification	ImageNet	BraTS 2020, Brain Tumor MRI Dataset	Robust residual learning for deep features	High computational load	BraTS: Acc: 91.2 %, F1: 90.5 %External: Acc: 89.8 %, F1: 88.7 %
Dense-UNet [12]	Dense CNN	Segmentation	Custom Medical	BraTS 2020	Efficient gradient flow; better deep feature learning	Memory intensive	Dice: 0.892 IoU: 81.2 %Hausdorff: 3.1 px
Att-UNet [13]	CNN + Attention	Segmentation	ImageNet	BraTS 2020	ROI-specific attention; strong localization	Higher param count	Dice: 0.871 IoU: 78.9 %Hausdorff: 3.4 px
CNN-TumorNet [24]	CNN	Classification	BraTS	Brain Tumor MRI Dataset	Lightweight, interpretable CNN; good external generalization	Limited context learning	Acc: 90.1 %F1: 89.0 %
Transformer-Based Models							
SegFormer [25]	Transformer + MLP Decoder	Segmentation	COCO, ADE20K	BraTS 2020, Brain MRI Annotated	CRF-free, real-time; compact model	GPU memory usage	BraTS: Dice: 0.87 IoU: 79.4 % External: Dice: 0.84IoU: 76.3 %
Swin UNETR [16]	Swin Transformer + UNETR	Segmentation	ImageNet-21 K	BraTS 2020	High precision segmentation; hierarchical encoding	High GPU cost	Dice: 0.895 IoU: 82.1 %Hausdorff: 2.5 px
UCTransNet (H. [26]	Transformer + CNN Hybrid	Segmentation	ImageNet	BraTS 2020	Efficient cross-attention; improved fusion	Slightly slower inference	Dice: 0.879 IoU: 78.5 %Hausdorff: 3.2 px
PVT-UNet (W. [22]	PVT + U-Net	Segmentation	ImageNet	BraTS 2020	Lightweight transformer backbone	Limited 3D analysis	Dice: 0.881 IoU: 78.2 %Hausdorff: 3.0 px
PVT + SegFormer (Proposed Study)	Dual-Transformer (PVT + SegFormer)	Classification + Segmentation	ImageNe, COCO, ADE20K	BraTS 2020, Brain Tumor MRI Dataset, Brain MRI Annotated	Combines classification & segmentation Real-time (<50 ms), generalizableCompact, no CRF post-processing	SegFormer has higher training memory usage	Classification (BraTS): Acc: 94.0 %, F1: 93.3 % Classification (External): Acc: 92.3 %, F1: 91.0 % Segmentation (BraTS): Dice: 0.87Segmentation (External): Dice: 0.84

cent) on the external classification dataset [21]. In terms of segmentation, BrainDx scored 0.87 on BraTS and 0.84 on the annotated external dataset [21]. It is in real-time (<50 ms). Its small footprint, low memory training requirement, and CRF-free nature on the one hand, and dual-transformer architecture on the other, make it among the most suitable architectures for clinical implementation, though SegFormer is a higher memory model overall.

5. Conclusion

This research summarizes BrainDx, a novel dual-transformer framework, to which the Pyramid Vision Transformer (PVT) model is applied for classifying brain tumors, and SegFormer is used for tumor segmentation. By combining the strengths of the two transformer architectures, the framework provides a robust and clinically applicable solution for diagnosing brain tumors based on MRI images. The PVT model achieved strong classifications with an accuracy of 94.0 %, high precision, and F1-Score across the tumor categories: Gliomas, Meningiomas, Pituitary Tumors, and Healthy Brain. This was due to the multi-scale hierarchical attention mechanism employed by PVT, which enabled the model to efficiently extract both local and global contextual features from high-resolution MRI scans. The segmentation task performed on SegFormer further boosted the system's performance, achieving a Dice score of 0.87, which outperformed classic models such as U-Net and DeepLabV3. Importantly, SegFormer processed real-time inference in under 50 ms, making it highly suitable for time-critical clinical diagnosis. The framework's architecture did not require post-processing techniques such as Conditional Random Fields (CRFs), which are traditionally used to refine segmentation results, while significantly slowing down the process. Ablation studies demonstrated the effectiveness of incorporating data augmentation and multi-head attention in enhancing the classification accuracy and generalizability of the model, while incurring minimal additional computational cost. Comparative analyses with current cutting-edge methods showed that, in addition to enhancing accuracy and segmentation fidelity, BrainDx ensures efficient processing under varied conditions. This two-model strategy is beneficial for its end-to-end functionality, which ranges from recognizing tumor types to the precise demarcation of boundaries. In future work, the framework can be modified to apply it to more diverse and rare tumor types, incorporate multimodal imaging inputs (e.g., Positron Emission Tomography, Computed Tomography (CT)), and utilize explainable AI (XAI) methods to enhance the interpretability of the model. BrainDx shows definite promise as a reliable tool in a diagnostic workflow, facilitating radiologists in obtaining speedier and more uniform findings on tumors. Validation across multiple publicly available datasets confirms the generalizability of the proposed framework, supporting its potential deployment in diverse clinical environments. The use of three diverse, publicly available datasets enhances robustness across tumor types, class distributions, and imaging modalities.

Future research will focus on enhancing generalizability by fine-tuning and evaluating the BrainDx framework on clinically sourced datasets that include rare tumor types and multimodal imaging inputs such as CT and PET-MRI. This direction aims further to enhance the model's applicability in heterogeneous clinical environments. Future work will also involve domain adaptation across clinical sites, external validation through cross-institutional datasets, and comparative evaluation against expert radiologists in clinical trials.

CRediT authorship contribution statement

Arshleen Kaur: Writing – original draft, Formal analysis, Data curation, Conceptualization. **Vinay Kukreja:** Methodology, Supervision, Writing – review & editing. **Modafar Ati:** Resources, Project administration, Methodology. **Ankit Bansal:** Visualization, Validation. **Shanmugasundaram Hariharan:** Writing – review & editing, Visualization.

Funding

This is to acknowledge the support provided by the Office of Research and Sponsored Programs at Abu Dhabi University, United Arab Emirates, for funding this research as part of an internal research grant. The grant, referenced under number 19300636, has been instrumental in facilitating this work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets are taken from Kaggle and their names are Brain Tumor Segmentation (BraTS2020), Brain Tumor MRI Dataset (Masoud Nickparvar), and Brain Tumor Classification MRI Dataset (Sartaj Bhuvaji) and these are publicly available and fully anonymized. The dataset's links are https://www.kaggle.com/datasets/awsaf49/brats2020-training-data, https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset, https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

References

[1] S. Natha, U. Laila, I.A. Gashim, K. Mahboob, M.N. Saeed, K.M. Noaman, Automated brain tumor identification in biomedical radiology images: a multi-model ensemble deep learning approach, Appl. Sc. 14 (5) (2024) 2210.
[2] M. Kordemir, K.K. Cevik, A. Bozkurt, A mask R-CNN approach for detection and classification of brain tumours from MR images, Comput. Meth. Biomech. Biomed. Eng. 11 (7) (2024) 2301391.
[3] Raja, P. Siva, Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach, Biocybern. Biomed. Eng. 40 (1) (2020) 440–453.
[4] N. Huda, K.R. Ku-Mahamud, CNN-based image segmentation approach in brain tumor classification: a review, Eng. Proc. 84 (1) (2025) 66.
[5] X. Chen, Y. Peng, Y. Guo, J. Sun, D. Li, J. Cui, MLRD-Net: 3D multiscale local cross-channel residual denoising network for MRI-based brain tumor segmentation, Med. Biol. Eng. Compu. 60 (12) (2022) 3377–3395.
[6] R. Vankdothu, M.A. Hameed, Brain tumor MRI images identification and classification based on the recurrent convolutional neural network, Meas.: Sens. 24 (2022) 100412.
[7] M.I. Sharif, J.P. Li, M.A. Khan, M.A. Saleem, Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images, Pattern Recogn. Lett. 129 (1) (2020) 181–189.
[8] A. Tiwari, S. Srivastava, M. Pant, Brain tumor segmentation and classification from magnetic resonance images: review of selected methods from 2014 to 2019, Pattern Recogn. Lett. 131 (2020) 244–260.
[9] M.A. Naser, M.J. Deen, Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images, Comput. Biol. Med. 121 (1) (2020) 103758.
[10] M. Rasool, N.A. Ismail, W. Boulila, A. Ammar, H. Samma, W.M. Yafooz, A.-H.-M. Emara, A hybrid deep learning model for brain tumour classification, Entropy 24 (6) (2022) 799.
[11] S. Saifullah, A. Yudhana, A.P. Suryotomo, Automatic brain tumor segmentation: advancing U-net with ResNet50 encoder for precise medical image analysis, IEEE Access 13 (1) (2025) 43473–43489.
[12] E.S. Biratu, F. Schwenker, Y.M. Ayano, T.G. Debelee, A survey of brain tumor segmentation and classification algorithms, J. Imaging 7 (9) (2021) 179.
[13] M.O. Khairandish, M. Sharma, V. Jain, J.M. Chatterjee, N.Z. Jhanjhi, A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images, Irbm 43 (4) (2022) 290–299.
[14] D. Daimary, M.B. Bora, K. Amitab, D. Kandar, Brain tumor segmentation from MRI images using hybrid convolutional neural networks, Procedia Comput. Sci. 167 (1) (2020) 2419–2428.
[15] G. Wang, W. Li, S. Ourselin, T. Vercauteren, Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks, Int. MICCAI Brainlesion Workshop (2017) 178–190.
[16] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, Unetr: transformers for 3d medical image segmentation, Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (2022) 574–584.
[17] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, ArXiv Preprint (2021). ArXiv:2102.04306.
[18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16

words: transformers for image recognition at scale, ArXiv Preprint (2020). ArXiv: 2010.11929.
[19] S. Bhuvaji, Brain Tumor Classification (MRI), Kaggle.com, 2020.
[20] Brain Tumor Segmentation(BraTS2020). (2020). https://www.kaggle.com/datasets/awsaf49/brats2020-training-data.
[21] M. Nickparvar, Brain Tumor MRI Dataset, Kaggle.com, 2021.
[22] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pyramid vision transformer: a versatile backbone for dense prediction without convolutions, Proc. IEEE/CVF Int. Conf. Comput. vis. (2021) 568–578.
[23] A. Akter, N. Nosheen, S. Ahmed, M. Hossain, M.A. Yousuf, M.A.A. Almoyad, K. F. Hasan, M.A. Moni, Robust clinically applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumors, Expert Syst. Appl. 238 (PF) (2024) 122347, https://doi.org/10.1016/j.eswa.2023.122347.
[24] N. Rasool, N.A. Wani, J.I. Bhat, S. Saharan, V.K. Sharma, B.S. Alsulami, H. Alsharif, M.D. Lytras, CNN-TumorNet: leveraging explainability in deep learning for precise brain tumor diagnosis on MRI images, Front. Oncol. 15 (2025) 1554559.
[25] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: simple and efficient design for semantic segmentation with transformers, Adv. Neural Inf. Proces. Syst. 34 (2021) 12077–12090.
[26] H. Wang, P. Cao, J. Wang, O.R. Zaiane, Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer, Proc. AAAI Conf. Artif. Intell. 36 (3) (2022) 2441–2449.