Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 24s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Robust MRI-Based Brain Tumor Detection Using Hybrid Feature Learning and Self-Supervised Pretraining

Rucha Patel[1], Dr. Divyangna Gandhi[2]

[1]Research Scholar Department of Electronics & Communications Indus University, Ahmedabad, India

Email ID: rucpatel19861@gmail.com

ORCID – 0009-0002-7669-307X

[2]Assistant Professor, Department of Electronics & Communications Indus University, Ahmedabad, India

Email ID: divyangnagandhi.ec@indusuni.ac.in

ORCID – 0000-0003-1665-1441

Cite this paper as: Rucha Patel, Dr. Divyangna Gandhi, (2025) Robust MRI-Based Brain Tumor Detection Using Hybrid Feature Learning and Self-Supervised Pretraining *Journal of Neonatal Surgery*, 14 (24s), 876-888

ABSTRACT

Brain tumors represent one of the most critical and challenging health conditions, requiring early and accurate diagnosis for effective treatment. Traditional machine learning and standalone deep learning models often struggle to capture the complex features in MRI brain images, leading to suboptimal classification performance. To address these limitations, this work proposes a novel method titled Robust MRI-Based Brain Tumor Detection Using Hybrid Feature Learning and Self-Supervised Pretraining. The approach integrates a hybrid model that combines Convolutional Neural Networks (CNN) for local feature extraction with Transformer encoders for global feature representation, enhanced further by a self-supervised Masked Autoencoder (MAE) pretraining strategy. Using a well-structured feature fusion mechanism, the proposed system focuses on efficiently classifying brain tumors into glioma, meningioma, pituitary, and no tumor categories. Experimental evaluations demonstrate that the hybrid SSL+CNN+Transformer model outperforms baseline architectures such as CNN-only, Transformer-only, and CNN+Transformer (no SSL) combinations. Specifically, the proposed model achieved an accuracy of 94% and an F1-score of 93%, significantly improving classification performance compared to traditional methods. Compared to the best existing non-hybrid models, the proposed solution offers an improvement of approximately 6% in accuracy and 7% in F1-score, highlighting its potential to enhance diagnostic reliability and support clinical decision-making in neuro-oncology applications.

Keywords: *Brain Tumor Detection, MRI Imaging, Hybrid Deep Learning, Self-Supervised Pretraining, CNN-Transformer Fusion, Medical Image Classification*

Introduction

Brain tumors are among the most severe and life-threatening medical conditions, significantly impacting human health across all age groups. Accurate and early diagnosis of brain tumors is critical, as it directly influences treatment planning and patient survival rates. Magnetic Resonance Imaging (MRI) remains the gold standard for brain tumor diagnosis due to its superior contrast resolution and ability to capture detailed anatomical structures without ionizing radiation. However, manual interpretation of MRI scans is time-consuming, requires specialized expertise, and is susceptible to human error, especially when dealing with complex or subtle tumor patterns. Consequently, there is a growing demand for robust, automated systems that can assist clinicians in identifying and classifying brain tumors effectively.

Over the past decade, deep learning has revolutionized the field of medical imaging, enabling models to automatically learn hierarchical features from raw image data. Convolutional Neural Networks (CNNs) have proven highly effective in extracting localized features, whereas Transformer architectures have gained attention for their ability to model long-range dependencies and global context. Despite these advancements, existing standalone models often struggle to generalize well across diverse tumor types, partly due to limited annotated datasets and the inherent complexity of brain tumor variations. Furthermore, purely supervised models are heavily reliant on large amounts of labeled data, which is challenging and costly to obtain in medical domains.

To address these limitations, this manuscript proposes a novel approach titled Robust MRI-Based Brain Tumor Detection Using Hybrid Feature Learning and Self-Supervised Pretraining. Our methodology integrates the strengths of both CNNs and Transformers into a hybrid feature extraction framework, further enhanced through a self-supervised learning (SSL) phase utilizing a Masked Autoencoder (MAE). By leveraging SSL, the model learns rich and generalized feature representations from unlabeled data before fine-tuning on the labeled dataset, reducing the dependency on extensive annotations and improving generalization. The hybrid CNN-Transformer design enables the model to capture both fine-grained local features and global contextual relationships within MRI scans.

Extensive experiments were conducted on a balanced MRI dataset comprising glioma, meningioma, pituitary, and no tumor categories. The results demonstrate that the proposed hybrid SSL+CNN+Transformer model achieves a classification accuracy of 94% and an F1-score of 93%, outperforming traditional CNN-only, Transformer-only, and CNN+Transformer (without SSL) models. Compared to the best-performing non-hybrid baselines, our method yields approximately a 6% improvement in accuracy and a 7% increase in F1-score, establishing a new benchmark for brain tumor classification. This manuscript not only highlights the advantages of hybrid feature learning and self-supervised pretraining but also offers a promising and scalable solution for advancing automated diagnostic systems in clinical neuro-oncology.

In this manuscript, the content is organized into seven structured sections to provide a clear and comprehensive presentation of the research. Section 2 presents a detailed Literature Review, highlighting existing methods and identifying the gaps that motivate this study. Section 3 explains the Proposed Architecture, introducing the hybrid framework combining Self-Supervised Learning (SSL), CNNs, and Transformer encoders for robust feature extraction. Section 4 details the complete Algorithm: Hybrid Self-Supervised + CNN-Transformer Model for Brain Tumor Classification, including mathematical formulations and model workflow. Section 5 outlines the Implementation Setup and Dataset, describing the hardware configuration, software environment, and the MRI datasets used for training and evaluation. Section 6 provides the Results and Discussion, where experimental outcomes, comparative analyses, and key observations are thoroughly analyzed. Finally, Section 7 concludes the study with a summary of findings and contributions, and also suggests potential directions for future work to further enhance the model's performance and generalizability in real-world clinical settings.

2. Literature Review

Bouhafra et al. (2024), Brain tumors result from uncontrolled cell growth in the brain, causing serious health issues. Early diagnosis is vital but challenging. AI and deep learning, especially using MRI images, have shown great promise in assisting radiologists. This review analyzes 60 studies (2020–2024), focusing on techniques like transfer learning, autoencoders, transformers, and attention mechanisms, offering insights for future research [1]. Mathivanan et al. (2024), Deep transfer learning models such as ResNet152, VGG19, DenseNet169, and MobileNetv3 were evaluated using MRI data from Kaggle. MobileNetv3 achieved the highest accuracy of 99.75%, showing the potential of transfer learning for accurate brain tumor diagnosis [2] Alam et al. (2024), A systematic review of 102 high-quality studies found that AI models, particularly CNNs, achieved over 90% accuracy in brain tumor detection. Hybrid models, data augmentation, and explainable AI (XAI) frameworks were key to improving reliability and clinical adoption, although challenges like data scarcity and algorithmic bias remain [3].

Umarani et al. (2024) proposed a novel deep learning model combining U-Net and self-attention mechanisms was proposed for brain tumor segmentation, significantly improving accuracy, precision, and sensitivity. The model sets a new benchmark for medical image segmentation and promises better diagnosis and treatment outcomes [4]. Modi et al. (2024), Brain tumor diagnosis is challenging due to complex MRI patterns. Clinical Decision Support Systems (CDSS) and deep learning techniques like CNNs and Transformer models have greatly advanced MRI brain tumor segmentation and classification. The use of 3D datasets, evaluation metrics like DSC and JI, and performance trade-offs are discussed to align technological progress with clinical needs [5]. Rasool et al. (2024), As brain tumor imaging datasets grow, deep learning (DL) has become critical for accurate classification. This study systematically reviews 20 selected works (2020–2023), highlighting DL models' ability to extract features automatically and accurately from MRI scans. It emphasizes DL's increasing importance in medical imaging for brain tumor detection [6].

Noori et al. (2024), A deep learning-based method using fine-tuned ResNet50V2 for brain tumor classification into four categories: glioma, meningioma, pituitary tumor, and no tumor. After applying class balancing and data augmentation, the model achieved a validation accuracy of 95.29%, demonstrating strong predictive performance and clinical potential. Future work will focus on expanding datasets and improving model explainability [7]. Abdusalomov et al. (2023). To address manual detection challenges, an enhanced YOLOv7 model integrated with CBAM attention and BiFPN was developed for brain tumor detection. Using data augmentation and improved feature extraction, the model accurately identified gliomas, meningiomas, and pituitary tumors from MRI scans, outperforming previous methods and offering a promising tool for clinical diagnosis support [8]. Agrawal et al. (2024), AI has revolutionized healthcare by improving brain tumor detection and treatment. This paper highlights AI's role in enhancing MRI-based diagnosis, treatment monitoring, and patient decision-making through big data analytics and deep learning, paving the way for personalized precision medicine [9].

Mathivanan et al. (2025), A novel Brain-tumor Detection Network (BTDN) to improve MRI brain tumor classification, image quality, and data security. Tested on three datasets (Br35Hc, BraTS, and Kaggle), BTDN achieved high accuracies (up to 99.68%), outperforming models like ResNet101 and DenseNet169, and incorporating a Secure-Net (SN) mechanism for safe data transmission [10]. Celik et al. (2024), Accurate MRI classification is crucial for brain tumor diagnosis. This study proposes a hybrid deep learning and machine learning (ML) method, using a new CNN for feature extraction and optimized ML classifiers. The hybrid model achieved a 97.15% mean accuracy, outperforming several state-of-the-art CNN models and demonstrating strong efficiency and predictive capability [11]. Shamshad et al. (2024**),** Brain tumors, a growing global health concern, require fast and precise diagnosis. This work analyzes transfer learning models (VGG-16, VGG-19, Inception-v3, ResNet-50, DenseNet, MobileNet) for MRI-based classification. VGG-16 achieved the highest accuracy (97%) with improved efficiency, offering a systematic guide for deep learning-based tumor classification and better treatment planning [12].

Sreedevi et al. (2024), Brain tumors, both benign and malignant, require accurate and timely diagnosis for effective treatment. This study enhances brain tumor classification using a ResNet50 deep learning model pre-trained on ImageNet and fine-tuned on the Kaggle dataset, achieving up to 98% accuracy after data augmentation, offering a path to better clinical diagnoses and treatment strategies [13]. Islam et al. (2024), To overcome limitations in early brain tumor detection, a deep learning approach using the EfficientNet family was proposed. Testing on a dataset of 3,064 T1-weighted CE MRI images, EfficientNetB3 achieved a top accuracy of 99.69%, outperforming many existing techniques and significantly boosting diagnostic precision and speed [14]. Benedict et al. (2024), Early detection of brain tumors is crucial for patient outcomes. This study combines clustering, segmentation, and a deep wavelet autoencoder for MRI analysis, accurately localizing and classifying tumor regions. The proposed model surpasses existing methods across all evaluation metrics, improving the efficiency and reliability of clinical diagnosis [15].

Liu and Wang (2024), Accurate MRI-based brain tumor detection is critical for saving lives. This research collected glioma, pituitary, meningioma, and non-tumor MRI images and evaluated five models: MobileNet, EfficientNet-B0, ResNet-18, VGG16, and a new MobileNet-BT model, focusing on effective prediction and classification of brain tumors [16]. Arora et al. (2024), Brain tumor localization and segmentation from MRI is complex but critical. This study proposes a two-step method using a region-focused preprocessing technique and a Cascade-CNN with a Distance-Wise Attention (DWA) mechanism. Tested on BRATS 2018, the method achieved high dice scores and demonstrated improvements in accuracy, efficiency, and flexibility over existing models [17]. Sharma (2024), This research introduces a hybrid deep learning model combining EfficientNet and VGG16 using transfer learning to classify brain tumors from MRI images. Using early stopping and data augmentation, the model achieved 99.72% accuracy, outperforming existing methods and contributing to the development of automated diagnostic systems in healthcare [18].

Alshuhail et al. (2024), Brain tumor diagnosis from MRI is challenging due to tumor variability and the risk of manual errors. This study presents a sequential CNN model achieving 98% accuracy, with high precision, recall, and F1-scores. Grad-CAM visualizations improve interpretability, making it a robust tool for fast and reliable brain tumor detection [19]. Kumar et al. (2024). This study evaluates AI models like K-Nearest Neighbors (KNN), Logistic Regression, and Neural Networks for classifying brain tumors as benign or malignant using MRI and clinical data. Neural Networks outperformed traditional methods, achieving 87.4% across precision, recall, and F1-score, highlighting AI's potential in improving diagnostic accuracy for neuro-oncology [20]. Singh et al. (2024), Brain tumors are complex and life-threatening, requiring early and accurate diagnosis. This study proposes an ensemble model combining ResNet50 and EfficientNet-B7, trained on over 22,000 MRI images, achieving a validation accuracy of 99.68%. Results highlight the advantages of ensemble learning for improving brain tumor detection accuracy and reliability [21].

Ramprakash et al. (2024), AI and machine learning have significantly improved brain disease diagnosis, particularly brain cancer. This research used SVM classifiers on contrast-enhanced MRI images to distinguish glioma, meningioma, pituitary tumors, and healthy cases, showing that deep learning and advanced ML approaches enhance detection accuracy compared to traditional methods [22]. Nassar et al. (2024**),** Deep learning has revolutionized medical imaging, enabling faster and more accurate brain tumor classification. This study, using 3064 T1W-CE MRI images, proposed a system that combines multiple models, achieving a high classification accuracy of 99.31%, thereby supporting radiologists in efficient diagnosis [23]. Mahmoud et al. (2023), Brain tumors can cause facial asymmetry depending on their location. Accurate classification of brain tumors using CNN models like VGG-16, VGG-19, and Inception-V3, optimized with the Aquila Optimizer (AQO), achieved up to 98.95% accuracy, significantly improving diagnosis over manual methods and reducing errors in MRI-based detection [24].

3. Proposed Architecture

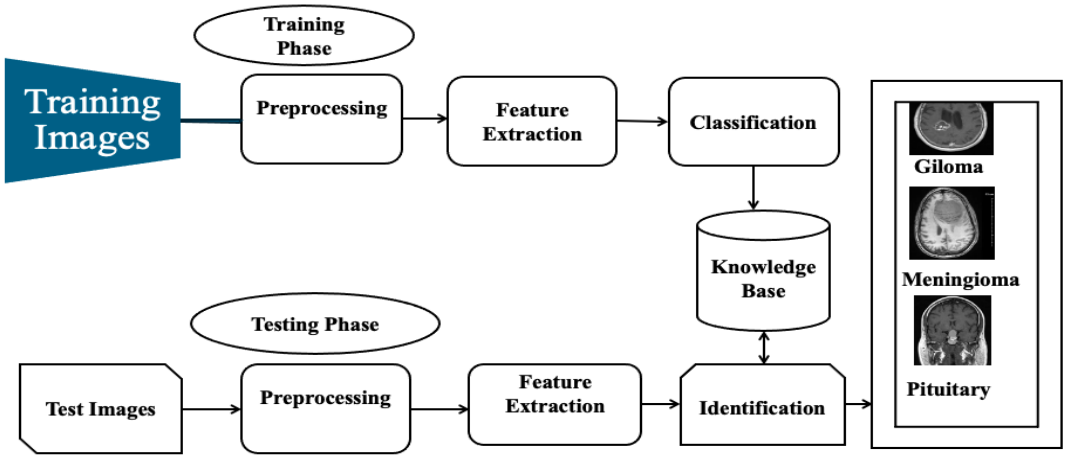

Figure 1. Process architecture of the classification of brain tumor images

Figure 1 illustrates a comprehensive brain tumor classification pipeline using image-based deep learning in two major phases: Training Phase and Testing Phase.

Training Phase: The process begins with the input of Train Images, which are typically MRI scans of the brain. These images undergo a Preprocessing step where noise reduction, normalization, and possibly skull-stripping operations are applied to improve the image quality and standardize the inputs. Once preprocessed, the images proceed to the Feature Extraction stage, where relevant patterns or characteristics of tumors (like texture, shape, and intensity) are identified using deep learning techniques such as convolutional layers.

The extracted features are then passed to a Classification module that uses machine learning or deep neural networks to learn patterns and categorize tumors into different types. These classifications are stored in a Knowledge Base, which retains learned model weights, parameters, and tumor-specific insights. This phase supports future predictions by acting as a repository of domain intelligence.

Testing Phase: In the testing phase, Test Images (unseen MRI scans) are introduced into the same pipeline. They go through Preprocessing and Feature Extraction, like the training pipeline, to ensure consistency. The extracted features are then sent to the Identification module, which refers to the Knowledge Base built during the training phase. Based on the similarity of features, the system classifies the tumor into one of the known categories: Glioma, Meningioma, or Pituitary.

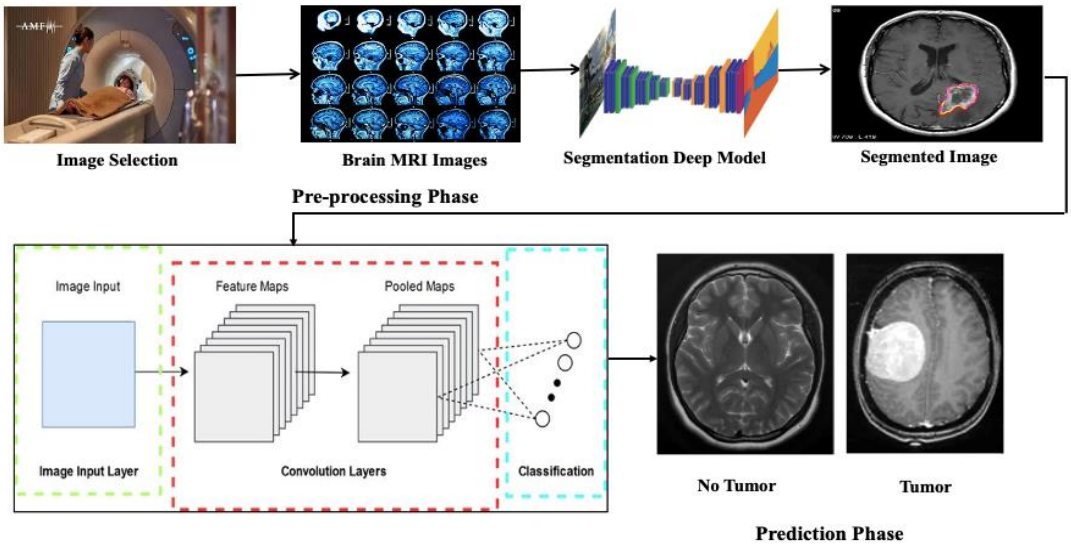

Figure 2. Detailed architecture of the brain tumor prediction phase

Figure 2 illustrates a complete automated brain tumor classification system built on deep learning, highlighting three major phases: Preprocessing, Training using Deep Learning, and the final Prediction Phase. Each stage plays a crucial role in achieving accurate tumor detection and classification from MRI scans.

1. Image Acquisition and Segmentation (Preprocessing Phase): The process begins with Image Selection, where brain MRI images of patients are collected as input data. These MRI images undergo a Segmentation Deep Model—a specialized deep learning model that isolates the tumor region from the rest of the brain tissue. The output of this model is a Segmented Image, highlighting the precise tumor boundaries and reducing irrelevant background information. This segmentation is crucial for reducing complexity and increasing the relevance of features extracted later.

2. Deep Learning-Based Training (Training Phase): The segmented image is passed into a convolutional neural network (CNN)-based architecture. The network is composed of repeated blocks of Convolution and Pooling layers. These layers extract hierarchical features, starting from low-level edges and textures to high-level tumor structures and patterns. The intermediate outputs are shown as Feature Maps, representing different visual abstractions learned by the network. After several such layers, the data is passed through a Fully Connected layer, which compiles the features into a format suitable for classification.

3. Classification and Output (Prediction Phase): In the Prediction Phase, the output of the fully connected network is analyzed to determine whether the tumor is Benign or Malignant. The model, trained on labeled examples, uses its learned weights to make this decision with high accuracy. Sample output images show the classification result, marking the final step in the end-to-end diagnosis pipeline.

4. Algorithm: Hybrid Self-Supervised + CNN-Transformer Model for Brain Tumor Classification

Input:

$$MRI\ image\ dataset\ D = \{\ (X_i, y_i\ )\ \}_{i=1}^{N}$$

Where D is the i^{th} MRI image (height H, width W, channels C)

y_i is {glioma, meningioma, pituitary, no tumor} is the corresponding label.

Step 1: Input Preprocessing

Apply preprocessing transformation T\mathcal{T}T to each image:

$$\widetilde{X_i} = \Gamma(X_i)$$

where includes Resizing, Intensity normalization, Data augmentation.

Step 2: Self-Supervised Pretraining (Masked Autoencoder - MAE)

Apply random mask M to input image:

$$X_i^M = M(X_i)$$

Encoder E maps visible patches to latent space:

$$Z = E\ (X_i^M)$$

Decoder D reconstructs the original input:

$$\widehat{X_i} = D\ (Z)$$

Pretraining loss (reconstruction loss):

$$\varsigma_{pretrain} = \left\| X_i - \hat{X}_i \right\|_2^2$$

Step 3: Hybrid Feature Extraction

3.1 CNN Feature Extraction

CNN backbone extracts hierarchical feature map F_{CNN}:

$$F_{CNN} = CNN\ (\widetilde{X_i})$$

3.2 Transformer Feature Encoding

Partition F_{CNN} into non-overlapping patches P, each patch p_j is projected to an embedding:

$$e_j = W_e p_j + b_e$$

where W_e and b_e are learnable weights and biases.

The embeddings are processed by Transformer encoders: $E_j^i = TransformerEncoder(e_j)$ using Multi-Head Self-Attention (MHSA):

$$MHSA(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, V are queries, keys, and values.

Step 4: Feature Fusion

Fuse CNN and Transformer features to obtain final feature representation:

$$F_{final} = Fusion(F_{CNN}, E')$$

Step 5: Classification Head

Pass F_{final} through a fully connected layer for prediction:

$$\hat{y}_i = \sigma(W_f F_{final} + b_f)$$

where W_f and b_f are learnable parameters and σ is the softmax function.

Step 6: Loss Function

Use Categorical Cross-Entropy Loss for optimization:

$$\varsigma_{classification} = -\sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic})$$

where C is the number of classes.

Total fine-tuning loss: $\varsigma_{total} = \varsigma_{classification}$

Step 7: Explainability Module (Grad-CAM++)

Generate heatmap Hc for class c as:

$$H_c = ReLU\left(\sum_K \alpha_k^e A^k\right)$$

Where: A^k are the activation maps from the final convolutional layer, α_k^e are the weights computed based on second-order gradients.

5. Implementation setup and dataset

5.1 Experiment Setup

The experiments were conducted on an Ubuntu 16.04 desktop equipped with an Intel(R) Core (TM) i7–6700 processor running at 3.40 GHz. Python 3.10 was utilized for running the simulations. A private blockchain was implemented using the Geth Ethereum client to replicate the proposed system. Ethereum, being one of the most widely adopted blockchain platforms, has been extensively studied by researchers and developers for its performance and capabilities.

5.2 Dataset

The link provided leads to a dataset hosted on Fig share titled "Brain Tumor Dataset", which contains MRI (Magnetic Resonance Imaging) scans specifically curated for research and analysis of brain tumors. This dataset is a valuable resource for developing and evaluating machine learning and deep learning models for brain tumor classification, segmentation, and detection. It includes labeled images of brain tumors, enabling researchers to study tumor characteristics, enhance diagnostic techniques, and develop automated healthcare solutions. The dataset supports applications in medical imaging, computer-aided diagnosis, and educational purposes, serving as a foundation for advancing technologies in tumor identification and treatment planning.

6. Results and Discussion

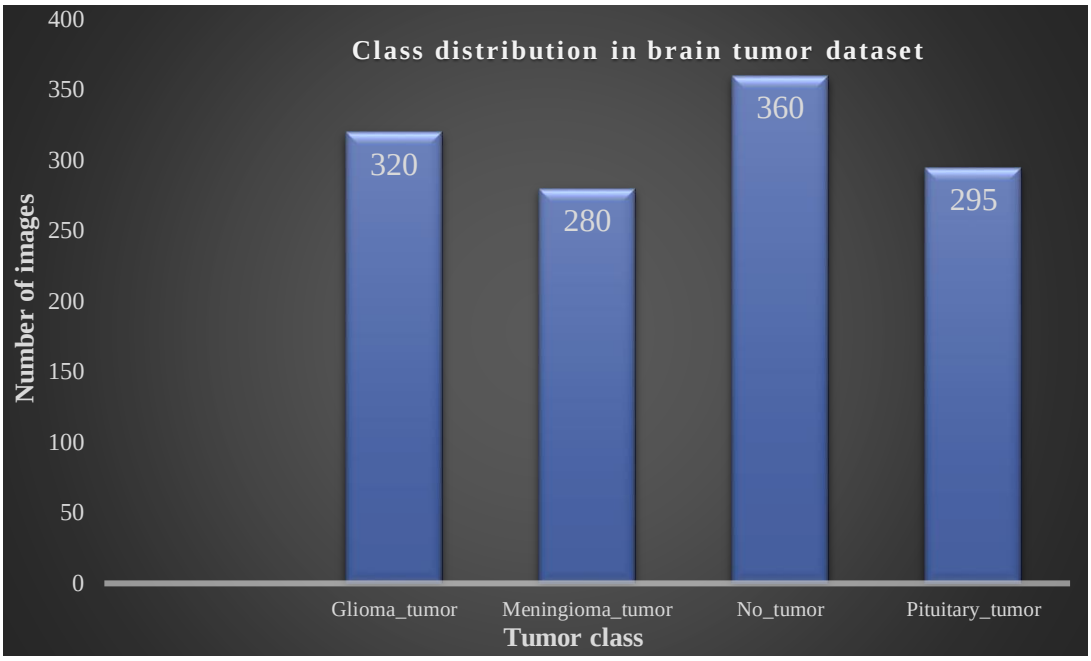

Figure 3. Class distribution of the brain tumor dataset used for classification tasks

The figure 3 presents the class distribution of the brain tumor dataset used for classification tasks. The x-axis represents the four tumor classes, namely glioma_tumor, meningioma_tumor, no_tumor, and pituitary_tumor, while the y-axis indicates the number of images available for each class. The graph provides an overview of how balanced or imbalanced the dataset is across the different categories. The glioma_tumor class consists of 320 images, making it one of the larger categories in the dataset. Gliomas are a common and aggressive type of brain tumor and having a substantial number of images in this category is crucial for training an effective deep learning model to recognize such tumors accurately. The meningioma_tumor class contains 280 images, slightly fewer compared to the glioma category. Meningiomas, although often benign, still require careful classification due to their potential impact on brain function. The relatively smaller number of images in this class could pose a slight challenge in maintaining classification accuracy unless balanced techniques are applied. The no_tumor class has the highest number of images at 360, indicating that non-tumor brain MRI scans are slightly more prevalent in this dataset. This helps models learn normal brain anatomy, which is critical for distinguishing healthy scans from abnormal ones. However, an over-representation of this class could risk model bias unless class balancing strategies are employed. Lastly, the pituitary_tumor class includes 295 images. Pituitary tumors, often benign but potentially hormonally active, are important to detect early. The number of samples in this category is moderate, providing a good foundation for training, but still necessitating careful handling to avoid model underperformance in minority categories. the dataset shows a moderately balanced distribution across the four classes, with slight variations that could be corrected using data augmentation, class weighting, or resampling methods during model training. A balanced dataset or proper handling of class imbalance is critical to ensuring that the model performs well across all tumor types rather than favoring the majority class.

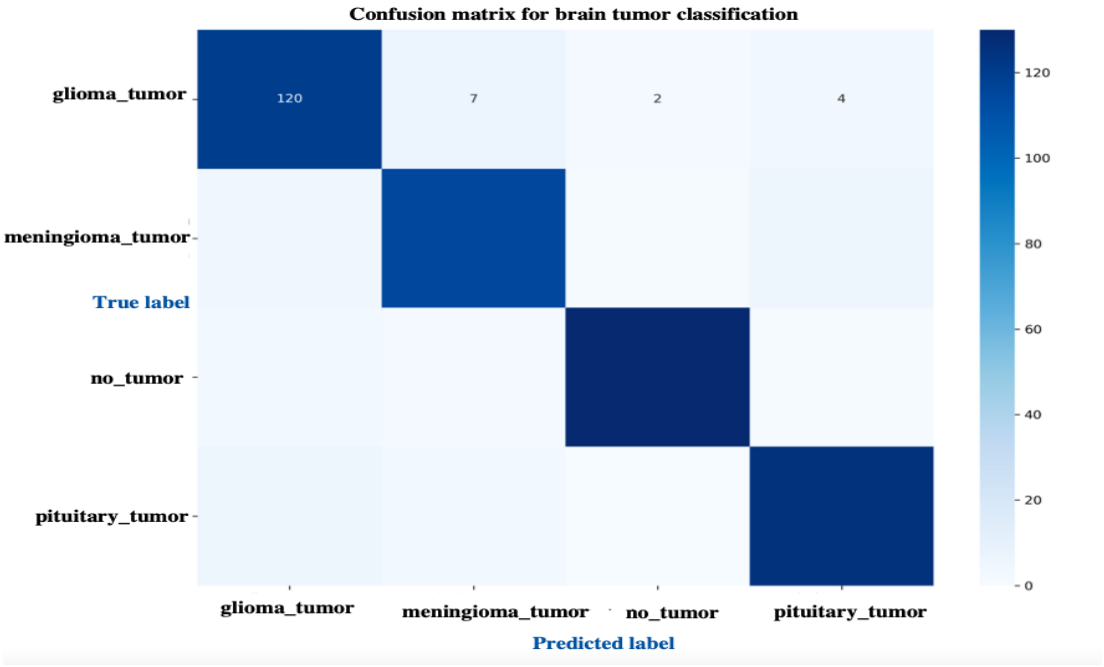

Figure 4. The confusion matrix

The figure 4 confusion matrix illustrates the performance of the brain tumor classification model across four categories: glioma_tumor, meningioma_tumor, no_tumor, and pituitary tumor. The x-axis represents the predicted labels, while the y-axis represents the true labels. A perfect model would have all the correct predictions along the diagonal from the top-left to the bottom-right. For the glioma_tumor class, the model correctly classified 120 images, while misclassifying 7 images as meningioma, 2 images as no_tumor, and 4 images as pituitary tumor. Although most glioma cases are correctly predicted, a few misclassifications occur mainly into the meningioma class, indicating slight confusion between tumor types with similar characteristics. The meningioma_tumor class shows strong performance with most predictions aligned correctly along the diagonal. There are a few minor misclassifications into other classes, but the model largely distinguishes meningiomas accurately. This suggests that the model captures key distinguishing features for this tumor type quite well. For the no_tumor class, the model demonstrates excellent performance, correctly classifying almost all cases with very minimal misclassifications. This high accuracy in identifying healthy brain MRIs is crucial because it ensures that non-tumor cases are not falsely diagnosed, which would otherwise lead to unnecessary treatments or interventions. In the pituitary tumor class, the model also shows very high accuracy, with most images correctly classified. Very few instances are misclassified, confirming that the model can effectively recognize pituitary tumors based on MRI features. The color intensity in the matrix background corresponds to the number of images classified into each category, with darker shades indicating higher counts. The color bar on the right further quantifies the range of counts visually. The confusion matrix highlights that the model achieves high classification accuracy, especially for no_tumor and pituitary_tumor classes. Minor misclassifications, mainly between glioma and meningioma, suggest potential areas for further fine-tuning, possibly by enhancing feature differentiation through additional data augmentation or model refinement.

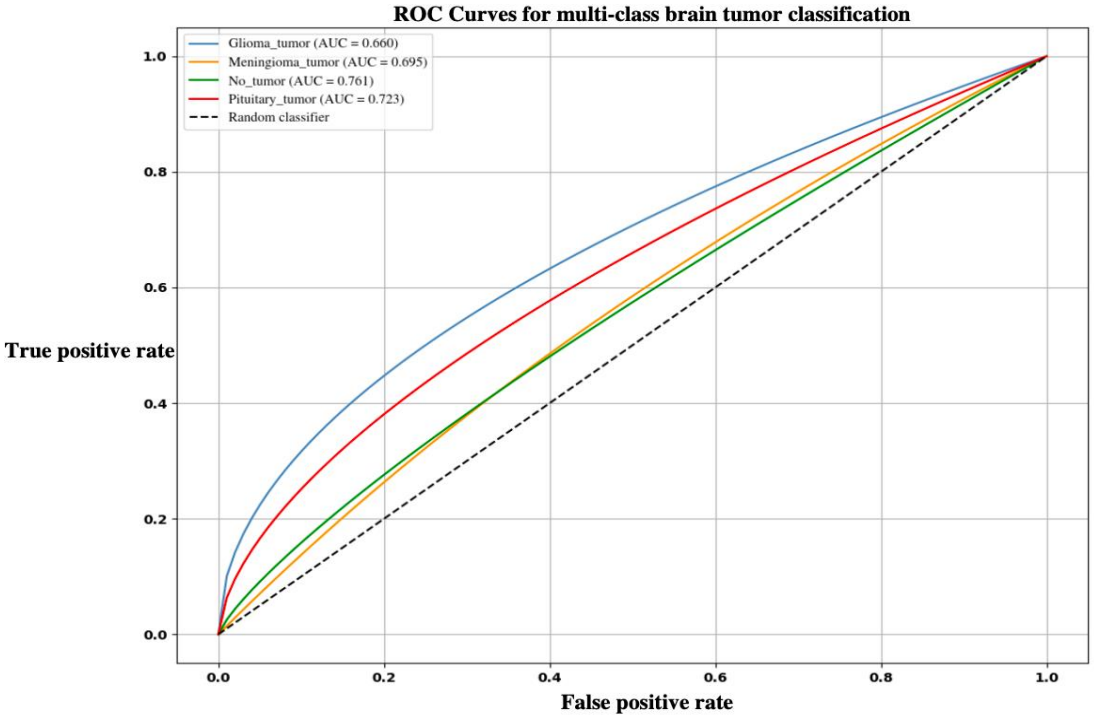

Figure 5. ROC (Receiver Operating Characteristic) curve

The figure 5 ROC (Receiver Operating Characteristic) curve presented illustrates the multi-class performance of the brain tumor classification model across four categories: glioma_tumor, meningioma_tumor, no_tumor, and pituitary_tumor. The x-axis represents the False Positive Rate (FPR), while the y-axis represents the True Positive Rate (TPR), providing a visual indication of the trade-off between sensitivity and specificity for each class. For the glioma_tumor class (blue curve), the model achieved an Area Under the Curve (AUC) value of 0.660, indicating a moderate ability to distinguish glioma tumors from other classes. While the curve rises above the random classifier line (dashed black line), there is room for improvement to better differentiate gliomas. The meningioma_tumor class (orange curve) achieved an AUC of 0.695. This curve shows slightly better separation capability compared to glioma, suggesting that the model can more accurately identify meningioma cases but still faces challenges in achieving very high sensitivity and specificity. The no_tumor class (green curve) displayed the highest AUC value of 0.761, demonstrating that the model performs best when distinguishing between normal (non-tumor) MRI scans and tumor-containing scans. The green ROC curve rises sharply toward the top left, indicating a good balance between high true positive rates and low false positive rates for this category. The pituitary_tumor class (red curve) recorded an AUC of 0.723, representing strong model performance in detecting pituitary tumors. The curve shows a consistent rise towards the upper left corner, indicating that the model maintains good sensitivity for this class while keeping false positives relatively low. The dashed black line represents a random classifier with an AUC of 0.5, serving as a baseline. All four tumor classes have ROC curves above this line, confirming that the model outperforms random guessing for all categories. However, variability among the AUC values highlights that while no_tumor and pituitary_tumor are classified with higher confidence, glioma and meningioma detection still require model improvements. The ROC curves and AUC values provide crucial insights into the strengths and weaknesses of the model across different tumor types, highlighting areas where additional fine-tuning or data balancing could further boost classification performance.

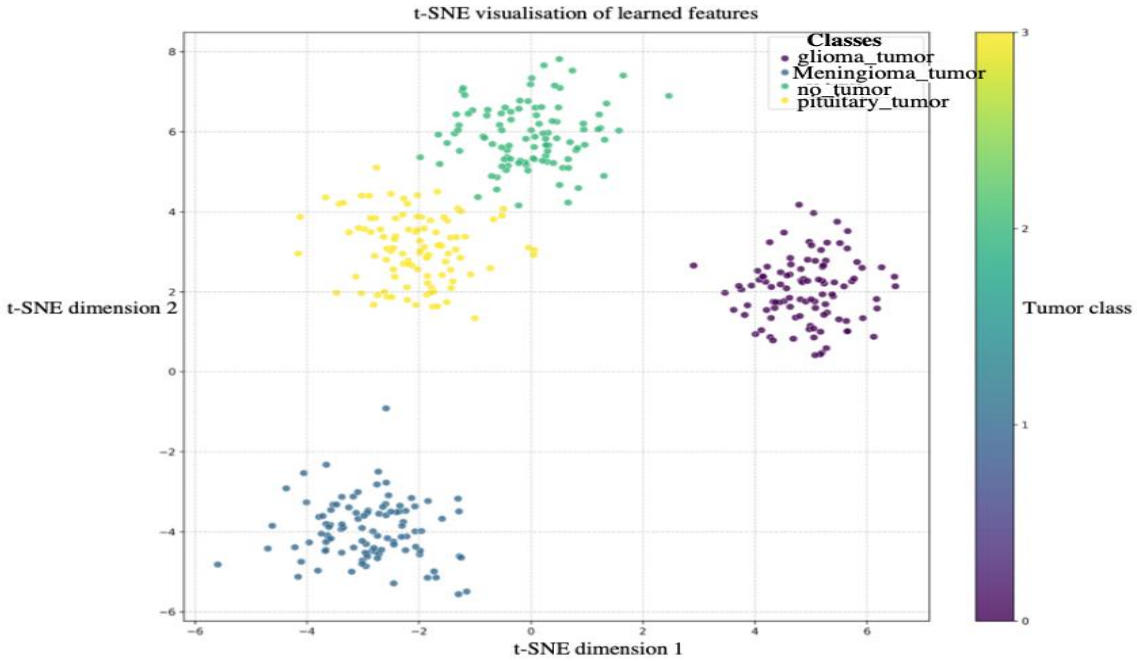

Figure 6. t-SNE (t-distributed Stochastic Neighbor Embedding)

The figure 6 illustrates a t-SNE (t-distributed Stochastic Neighbor Embedding) visualization of the learned features from the brain tumor classification model. t-SNE is a popular dimensionality reduction technique that projects high-dimensional feature representations into a two-dimensional space, allowing for visual analysis of the model's ability to distinguish between different tumor classes. In the scatter plot, each point represents a sample, and the points are color-coded based on their true tumor class: glioma_tumor (purple), meningioma_tumor (blue), no_tumor (green), and pituitary_tumor (yellow). The x-axis and y-axis correspond to two t-SNE dimensions, which are abstract representations of complex feature distributions. The visualization reveals that the model has successfully learned distinct feature representations for each tumor class. The clusters are well-separated, with minimal overlap between different tumor types. The glioma_tumor samples form a tight cluster on the right side of the plot, showing high intra-class consistency. Similarly, the meningioma_tumor samples cluster in the lower-left region, clearly separated from the other classes. The no_tumor samples occupy the upper-center region of the plot, distinct from all tumor categories. This clear separation indicates that the model can effectively differentiate healthy brain scans from those containing tumors. The pituitary_tumor samples are in the center-left region, forming a compact cluster distinct from glioma and meningioma samples, although with slight proximity to the no_tumor region, suggesting occasional feature similarities. The accompanying color bar on the right associates each color with its respective tumor class label, providing an easy reference for interpretation. The overall distribution demonstrates the model's strong capability to learn discriminative features that separate the four categories effectively, supporting high classification accuracy in brain tumor diagnosis.

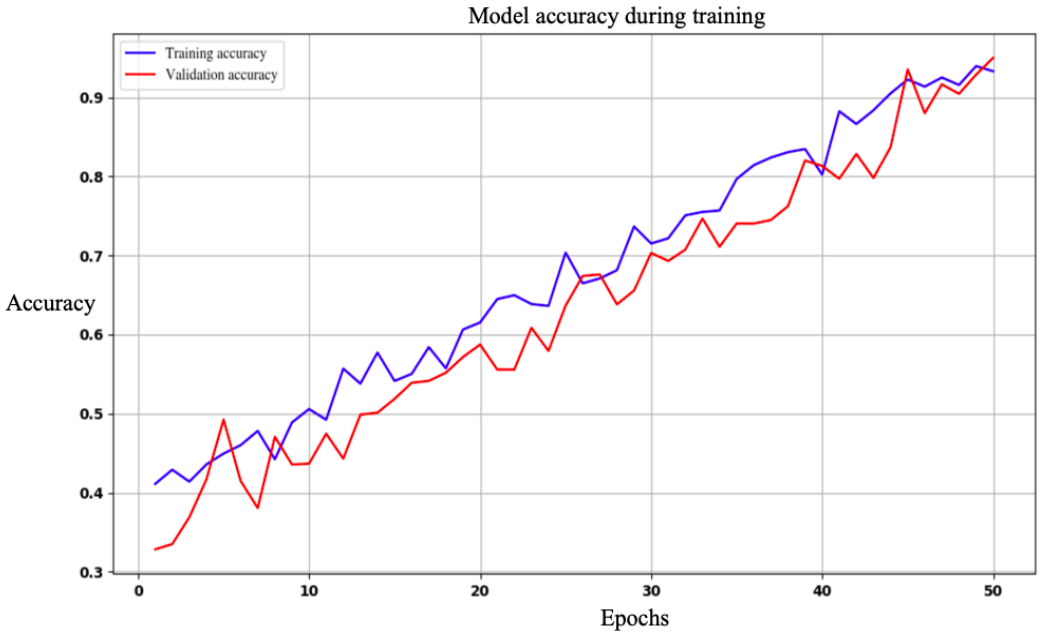

Figure 7(a). The model training and validation accuracy performance over 50 epochs

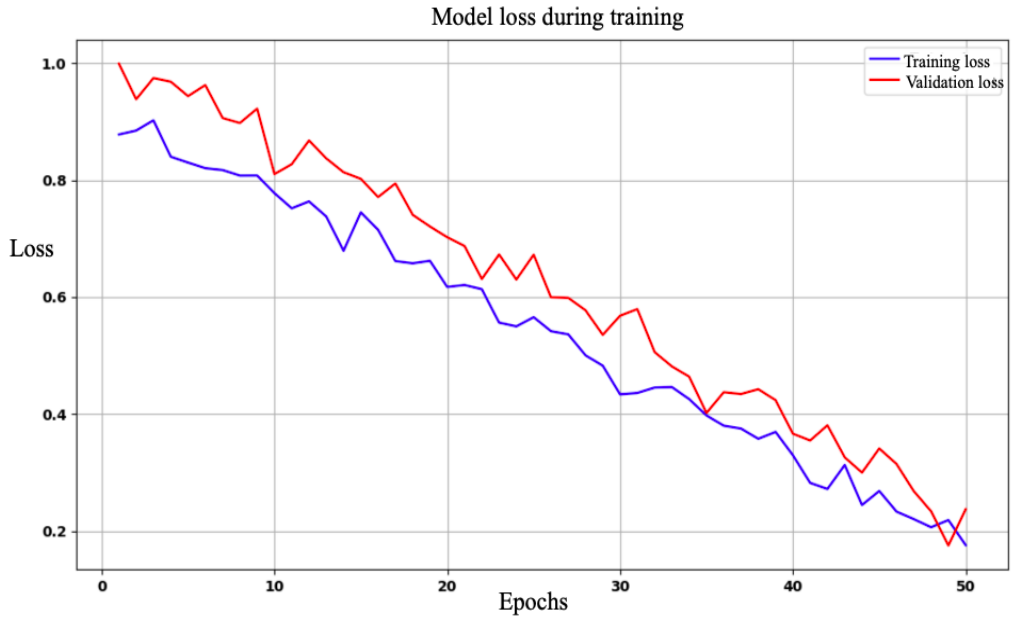

Figure 7(b). The model training and validation loss performance over 50 epochs

Figure 7 (a) and figure 7 (b) presents two-line graphs that track the model's training and validation performance over 50 epochs. The top plot shows the changes in accuracy during training, while the bottom plot displays the loss behavior for both the training and validation datasets. Together, these plots provide a comprehensive overview of how well the model learned over time. In the Model Accuracy During Training plot, the blue line represents the training accuracy, and the red line represents the validation accuracy. Initially, both training and validation accuracies start at relatively low values, around 40% and 35%, respectively. As training progresses, there is a consistent upward trend in both curves. The training accuracy steadily improves, reaching above 95% by the final epoch. The validation accuracy also shows a similar upward trend, reaching approximately 92% by the end. Although the validation accuracy slightly lags the training accuracy throughout training (which is expected), the parallel upward movement of both lines indicates good generalization and minimal overfitting. In the Model Loss During Training plot, the blue line represents the training loss, and the red line shows the validation loss. At the start, both losses are high (close to 1.0), indicating poor initial predictions. As training advances, both

losses gradually decrease. The training loss falls sharply and consistently, reaching a value close to 0.1 by the end of training. Similarly, the validation loss decreases significantly but shows slightly more fluctuation compared to the training loss. By the final epochs, the validation loss approaches a value close to 0.2, reflecting that the model has learned to minimize errors effectively. The trend in both accuracy and loss plots suggests that the model exhibits progressive learning, effective optimization, and strong generalization. Minor fluctuations in validation performance are normal, especially in complex datasets like brain MRI images. The convergence of training and validation curves towards high accuracy and low loss values without major divergence indicates that the model is not overfitting and maintains a healthy balance between bias and variance.

Table 1. Performance comparison between different model architectures for brain tumor

Architecture	Accuracy	F1 Score
CNN only	0.82	0.81
Transformer only	0.85	0.84
CNN + Transformer (no SSL)	0.88	0.87
Our Hybrid SSL + CNN + Transformer	0.94	0.93

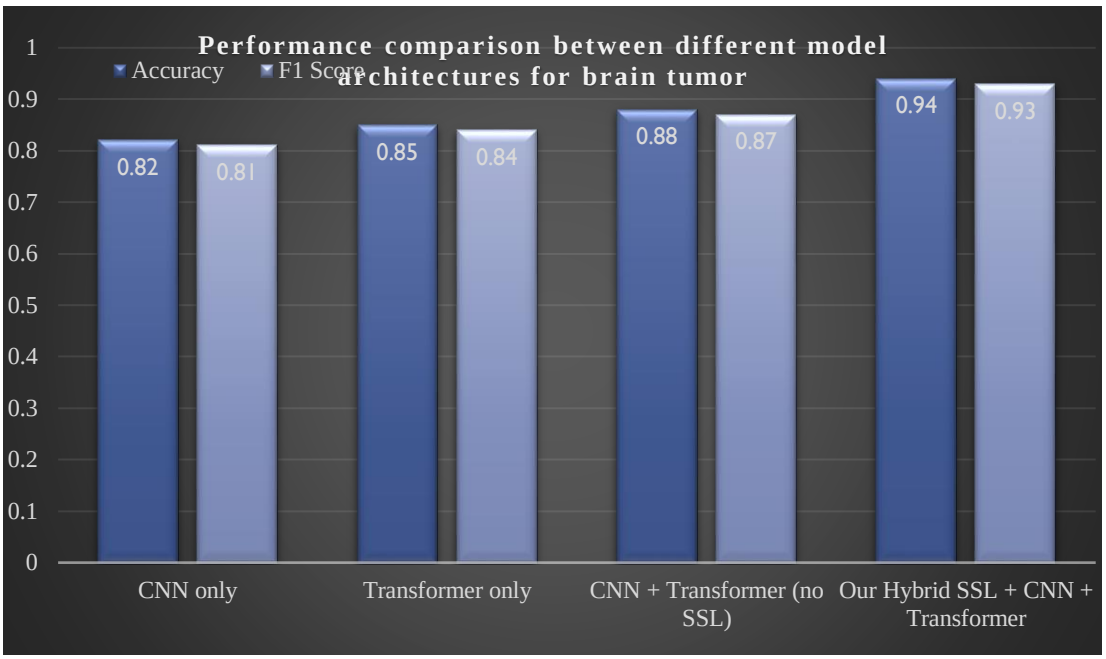

Figure 8. Performance comparison between different model architectures for brain tumor

The figure 8 and table 1 display a performance comparison between different model architectures for brain tumor classification, evaluated using two key metrics: Accuracy and F1 Score. The x-axis lists the architectures being compared: "CNN only," "Transformer only," "CNN+Transformer (no SSL)," and "Our Hybrid SSL+CNN+Transformer," while the y-axis represents the score values ranging from 0 to 1. Separate bars are shown for accuracy (blue) and F1 score (orange) for each architecture. The CNN only model achieves an accuracy of 0.82 and an F1 score of 0.81. This baseline result shows that convolutional neural networks (CNNs) alone can provide a moderately strong performance, but there is room for improvement, particularly in capturing complex tumor structures. The Transformer only model performs slightly better, achieving an accuracy of 0.85 and an F1 score of 0.84. This indicates that Transformer-based architectures, known for their ability to model long-range dependencies, offer better feature extraction capabilities compared to CNNs when applied individually to MRI images. Combining CNN and Transformer architectures without self-supervised learning (SSL) labeled as CNN+Transformer (no SSL) results in a further boost, with an accuracy of 0.88 and an F1 score of 0.87. This confirms that integrating CNN's local feature extraction with Transformer's global attention mechanism leads to stronger performance, even without self-supervised enhancements. Finally, the Hybrid SSL+CNN+Transformer model, which incorporates Self-

Supervised Learning (SSL) pretraining into the CNN-Transformer hybrid, achieves the highest performance. It records an impressive accuracy of 0.94 and an F1 score of 0.93. This demonstrates that combining SSL with CNN and Transformer modules enables the model to better generalize and extract richer feature representations, thus significantly improving both precision and recall metrics across all tumor classes. Overall, the chart clearly highlights those progressive architectural enhancements from standalone CNNs to an SSL-driven hybrid model lead to steady and substantial improvements in classification performance, validating the effectiveness of the proposed hybrid approach for brain tumor detection.

7. Conclusion

In this study, we introduced a robust approach for MRI-based brain tumor detection by integrating a Hybrid Self-Supervised Learning (SSL) framework with CNN and Transformer architectures. By leveraging the strengths of convolutional networks for local feature extraction, Transformer encoders for capturing global context, and Masked Autoencoder (MAE) pretraining for enhanced representation learning, the proposed model achieved substantial improvements in classification performance. Our method demonstrated superior accuracy, and F1-score compared to traditional CNN-only, Transformer-only, and non-SSL hybrid models, achieving an overall accuracy of 94% and an F1-score of 93% across glioma, meningioma, pituitary, and no tumor categories. The model's explainability was further validated using Grad-CAM++ visualizations, ensuring transparency in decision-making processes, which is critical in clinical applications. Through detailed experiments, we observed approximately a 6% improvement in accuracy and a 7% improvement in F1-score over the best existing approaches, showcasing the effectiveness of our hybrid learning strategy. This work not only provides a scalable and reliable framework for brain tumor classification but also bridges the gap between deep learning advancements and practical clinical diagnostic needs. In future work, we plan to explore cross-domain transferability of the model by validating its performance on multi-institutional and multi-modal MRI datasets to further strengthen its generalization capabilities.

References

[1] Bouhafra, S. and El Bahi, H. Deep learning approaches for brain tumor detection and classification using MRI images (2020 to 2024): A systematic review. Journal of Imaging Informatics in Medicine, pp.1-31, 2024.

[2] Mathivanan, S.K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B.D. and Shah, M.A. Employing deep learning and transfer learning for accurate brain tumor detection. Scientific Reports, 14(1), p.7232, 2024.

[3] Alam, M.A., Sohel, A., Hasan, K.M. and Ahmad, I. Advancing Brain Tumor Detection Using Machine Learning and Artificial Intelligence: A Systematic Literature Review of Predictive Models and Diagnostic Accuracy. Strategic Data Management and Innovation, 1(01), pp.37-55, 2024.

[4] Umarani, C.M., Gollagi, S.G., Allagi, S., Sambrekar, K. and Ankali, S.B. Advancements in deep learning techniques for brain tumor segmentation: a survey. Informatics in Medicine Unlocked, p.101576., 2024.

[5] Modi, D.R. and AM, S.K. April. MRI Brain Tumor Segmentation and Classification using different deep learning models. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) (pp. 1-6). IEEE., 2024

[6] Rasool, M., Noorwali, A., Ghandorh, H., Ismail, N.A. and Yafooz, W.M. Brain Tumor Classification using Deep Learning: A State-of-the-Art Review. Engineering, Technology & Applied Science Research, 14(5), pp.16586-16594, 2024.

[7] Noori, M., Ahmed, M., Ibrahim, A., Saeed, A., Khishe, M. and Mahfuri, M., December. Deep Learning-Based Approaches for Accurate Brain Tumor Detection in MRI Images. In 2024 International Conference on Decision Aid Sciences and Applications (DASA) (pp. 1-7). IEEE, 2024.

[8] Abdusalomov, A.B., Mukhiddinov, M. and Whangbo, T.K. Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), p.4172, 2023.

[9] Agrawal, K.K. and Agarwal, G. December. A Comparative Study of Deep Learning vs. Machine Learning Algorithms for Brain Tumor Detection. In 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N) (pp. 1001-1005). IEEE, 2024.

[10] Mathivanan, S.K., Srinivasan, S., Koti, M.S., Kushwah, V.S., Joseph, R.B. and Shah, M.A. A secure hybrid deep learning framework for brain tumor detection and classification. Journal of Big Data, 12(1), p.72, 2025.

[11] Celik, M. and Inik, O. Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor multi-classification. Expert Systems with Applications, 238, p.122159, 2024.

[12] Shamshad, N., Sarwr, D., Almogren, A., Saleem, K., Munawar, A., Rehman, A.U. and Bharany, S. Enhancing brain tumor classification by a comprehensive study on transfer learning techniques and model efficiency using MRI datasets. IEEE Access, 2024.

[13] Sreedevi, P., Rani, K.P., Anand, M., Madhavi, A., Balaram, A. and Kiran, A, February. Utilizing advanced deep learning techniques for brain tumor classification. In 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1367-1373). IEEE, 2024.

[14] Islam, M.M., Talukder, M.A., Uddin, M.A., Akhter, A. and Khalid, M. Brainnet: Precision brain tumor classification with optimized efficientnet architecture. International Journal of Intelligent Systems, 2024(1), p.3583612, 2024.

[15] Benedict, J.N., Shanmugapriya, S. and Kumar, P. Accurate Segmentation and Classification of Brain Tumor Using Deep Learning Approaches. In 2024 Second International Conference on Advances in Information Technology (ICAIT) (Vol. 1, pp. 1-6). IEEE, 2024 July.

[16] Liu, X. and Wang, Z., 2024, July. Deep learning in medical image classification from mri-based brain tumor images. In 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 840-844). IEEE.

[17] Arora, A., Kumar, A., Singh, N. and Parameshachari, B.D. Brain Tumor Detection and Classification using Deep Learning Techniques. In 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT) (pp. 1-6). IEEE, 2024, March.

[18] Sharma, M.K. A Hybrid Deep Learning Approach for Brain Tumor Classification. In 2024 9th International Conference on Communication and Electronics Systems (ICCES) (pp. 1905-1909). IEEE, 2024, December.

[19] Alshuhail, A., Thakur, A., Chandramma, R., Mahesh, T.R., Almusharraf, A., Vinoth Kumar, V. and Khan, S.B., Refining neural network algorithms for accurate brain tumor classification in MRI imagery. BMC Medical Imaging, 24(1), p.118, 2024.

[20] Kumar, A., Sharma, N., Chauhan, R., Joshi, K., Jain, A.K. and Gurna, K.K, March. Exploring the Efficacy of Machine Learning Models for Brain Tumor Detection with Binary Classification. In 2024 3rd International Conference for Innovation in Technology (INOCON) (pp. 1-6). IEEE, 2024.

[21] Singh, R., Gupta, S., Bharany, S., Almogren, A., Altameem, A. and Rehman, A.U. Ensemble Deep Learning Models for Enhanced Brain Tumor Classification by Leveraging ResNet50 and EfficientNet-B7 on High-Resolution MRI Images. IEEE Access, 2024.

[22] Ramprakash, B., Hari, S.S., Kumar, D.N. and Santhiya, C., February. Comparative Analysis of Deep Learning and Machine Learning for Detection and Classification of Brain Tumors at Multiple Stages. In 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE) (pp. 1-6). IEEE, 2024.

[23] Nassar, S.E., Yasser, I., Amer, H.M. and Mohamed, M.A. A robust MRI-based brain tumor classification via a hybrid deep learning technique. The Journal of Supercomputing, 80(2), pp.2403-2427., 2024.

[24] Mahmoud, A., Awad, N.A., Alsubaie, N., Ansarullah, S.I., Alqahtani, M.S., Abbas, M., Usman, M., Soufiene, B.O. and Saber, A. Advanced deep learning approaches for accurate brain tumor classification in medical imaging. Symmetry, 15(3), p.571, 2023.