Fusion of Convolutional Architecture and Transformer Models for Enhanced Brain Tumor Classification

V. Sabitha[1,2*], Jagannath Nayak[3] , P. Ramana Reddy[4]
[1]Research Scholar,Department of Ece, Jntua, Ananthapuramu, Andhra Pradesh, 515002 India
2Department of ECE, Vaagdevi College of Engineering, Warangal, Telangana, 506002 India
[3]Professor, Department of Ece, Jntuace, Ananthapuramu, 515002, India
[4]Chess, Drdo Ministry of Defence, Government of India, Hyderabad, 500069 India
sabithavem@gmail.com[2]jnayakdr@gmail.com[3]prrjntu@gmail.com
[*]Corresponding author

Keywords: brain tumor, transformer model, CNN, MRI, deep learning

Recieved: March 10, 2025

Early detection of brain tumors based on MRI images has shown significant advancements with the advent of deep learning methods. However, achieving high accuracy and robustness in classification remains a challenge due to the complex and mixed nature of brain tumors and the clarity of samples. This study proposes a novel approach that integrates convolutional architectures with the transformer approach, which can lead to an optimal model. The convolutional neural networks (CNNs) excel in capturing local features and spatial hierarchies, while the transformer approach captures long-term dependencies and contextual information. By integrating these two robust architectures, our proposed model leverages the strengths of both to achieve superior performance. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) dataset is used to evaluate our model, which consists of 7023 samples across four classes. We compare the performance of the fusion model with that of the prescribed models. The results demonstrate that the fusion model significantly outperforms the standalone models, achieving a classification accuracy of 91.8%. The proposed approach also shows improved robustness in handling various tumor types and sizes, highlighting its potential for clinical application.

Povzetek: Za klasifikacijo možganskih tumorjev iz MRI (BRATS, 7023 vzorcev, 4 razredi) so uporabili hibridni fuzijski model, ki združi CNN (lokalne značilke) in transformer (globalni kontekst) za robustnejšo klasifikacijo heterogenih tumorjev.

1 Introduction

Brain tumors are the most challenging and life-threatening situations, requiring accurate diagnosis and effective treatment planning. Automatic early detection of tumors will overcome the threatening situations. Magnetic Resonance Imaging (MRI) samples are used for tumor detection and classification due to their superior contrast resolution and non-invasive nature. The early detection of tumors from MRI samples by Tampu, I. E., et al. (2024) [14]is crucial for determining appropriate treatment strategies and predicting patient outcomes. Traditional methods for brain tumor classification are mainly based on manual inspection and human analysis, which is a time-consuming process. As the number of patients increases day by day, manual detection becomes prone to variability, necessitating the development of an automated system. Many researchers have worked on deep learning on medical images to diagnose diseases, as seen in Odusami, M. (2024) [17].

In recent years, the use of deep learning (DL)in the field of medical image analysis has offered automated and highly accurate solutions for various diagnostic tasks. CNNs, Nobel, S. N., et al (2024) [3] in particular, have shown remarkable success in extracting hierarchical features from medical images and achieving high performance in classification tasks. However, despite their efficacy, CNNs have some limitations. For instance, these models have captured complex patterns and sequential patterns from an image, which are necessary for accurately classifying complex and heterogeneous brain tumors.

Transformers, a cutting-edge approach implemented for text-based data, have demonstrated their capability to capture sequential patterns and global patterns through self-attention mechanisms (Katran, L. F., et al., 2024) [4]. Their application to vision tasks has opened new avenues for enhancing image analysis performance. While transformers are capable of capturing long-term dependencies, they may struggle with capturing fine-grained local features due to their inherently global nature (Srinivas, B., et al, 2024) [11].

This paper proposes a novel approach combining CNN and transformer methods to enhance the strengths of both paradigms for improved brain tumor classification. By combining CNNs' ability to capture local features and transformers' proficiency in modeling global context, the

proposed hybrid model aims to achieve superior classification performance. This fusion approach is expected to address the limitations of standalone CNN and transformer models, providing a more robust and accurate classification framework. According to Chen, C., et al. (2023) [19], many of the systems implemented a transformer model to detect brain tumors.

The Multimodal BRATS dataset, a widely recognized and comprehensive dataset, is utilized to evaluate the performance of the proposed model. Extensive experiments are conducted to compare the performance of the fused model against state-of-the-art CNN and transformer-based models individually. Our results demonstrate that the fusion model outperforms the other models.

The paper is organized as follows: Section 2 reviews related work in brain tumor classification using DL. Section 3 describes the proposed fusion model architecture. Section 4 presents the experimental setup, including dataset details. Section 5 explores the experimental results analysis and comparison with prescribed models. Finally, Section 6 concludes the paper.

2 Related work

Hekmat et al (2025) [1] implemented an attention-based architecture for brain tumor detection. The model uses attention mechanisms to fuse different feature representations effectively, enhancing the accuracy of tumor detection in MRI scans. By clinicians to better understand the decision-making process. Extracted features from key regions of interest within MRI images, this method outperforms traditional CNN. Benzorgat, N. et al (2024) [2] proposed brain tumor classification by combining an ensemble of models with a transformer. With transformers, which capture global dependencies, and DL models that specialize in local features? The integrated model got an accuracy of 0.97. Nobel, S. N., et al. (2024) [3] proposed a hybrid model, a mixed convolutional-transformer model, aimed at diagnosing glioma subtypes rapidly and accurately. They combined CNN layers, which efficiently capture spatial information, with transformers to handle long-range dependencies. This hybrid model significantly improves the accuracy by 0.98. Mzoughi, H et al (2024) [5] Combined Vision Transformers (ViT) with Deep-CNN for classification of tumor images, incorporating explainable AI (XAI) for interpretability. The integration of the ViT and D-CNN models will learn both global and local features effectively, achieving an accuracy of 0.96. Alzahrani, S. M., and Qahtani, A. M. (2024) [6] worked with tripartite attention for multi-class brain tumor detection in highly augmented MRIs. They improved the generalization of models trained on augmented datasets by distilling knowledge from larger models into more compact ones. And got an accuracy of 0.97. Nguyen-Tat, T. B., (2024) [7] Proposed a hybrid approach for brain tumor segmentation that combines UNet, attention mechanisms, and transformers. This method integrates the strengths of each technique, with UNet efficiently capturing spatial features, transformers handling long-range dependencies, and attention mechanisms focusing on relevant regions. As a result, they achieved an accuracy of 0.91.

Gasmi, K., et al. (2024) [8] proposed an enhanced brain tumor diagnosis model that combines DL with a weight selection technique. This method aims to optimize the learning process by selecting the most relevant features and assigning them appropriate weights. Rasheed, Z., et al. (2024) [9] implemented a hybrid CNN model with an attention method for brain tumor identification. We improved the performance of CNNs by focusing on complex patterns from images using attention layers, achieving an accuracy of 0.97. Pacal, I. (2024) [10] proposed a Transformer method by adding a multi-layer perceptron and self-attention methods for diagnosing tumors automatically. The Transformer is known for its efficient handling of high-resolution images and is combined with a residual MLP to improve feature learning and classification accuracy. Kang, M., et al (2024) [12] Implemented a CNN-transformer network for brain tumor segmentation in cases with incomplete modalities. The method aims to address the challenge of missing or incomplete MRI data by distilling features from available modalities and utilizing the CNN-transformer architecture to refine the segmentation.

Asiri, A. A et al. (2024) [13] implemented the Swin Transformer for accurate brain tumor classification and performance analysis. The Swin Transformer can handle high-resolution images, and it is applied to the classification task to improve diagnostic accuracy. The paper also focuses on performance analysis, comparing the results with other state-of-the-art methods. Tabatabaei, S., et al. (2023) [15] proposed an attention method and DL architecture for tumor classification. The attention mechanism with the DL method will enable the model to focus on complex areas of samples, improving the accuracy of tumor classification. The model combines the benefits of attention-based transformers with traditional methods, leading to enhanced performance in tumor detection. Aloraini, M., et al. (2023) [16] implemented a transformer with CNN for effective brain tumor classification using MRI images. This hybrid model uses the strengths of both approaches: CNNs for local feature extraction and transformers for global dependency modeling. This combination leads to enhanced tumor classification accuracy**.** Sun, X., et al (2024) [18] implemented aEF-UV method for a feature-enhancement of U-Net and ViT for tumor segmentation. This approach uses the strengths of U-Net for segmentation and ViT for capturing long-range dependencies in the image. The fusion of these models enhances feature extraction and segmentation accuracy, particularly in complex brain tumor cases. Saleh et al. (2024) [20] implemented a multimodal approach for semantic segmentation in brain tumor images, integrating advanced models and optimal filters via advanced 3D segmentation methods. They used multiple imaging modalities to improve the segmentation accuracy by capturing complementary information from different sources. Zebari, N. A., et al. (2024) [21] proposed a DL model for detecting brain tumors from image samples.

And integrated multiple DL techniques to enhance the performance by fusing different features from various sources of samples.

Zakariah, M., et al. (2024) [22] proposed a Dual ViT with DSUNET for brain tumor segmentation. The feature fusion mechanism will demonstrate the model's ability to capture various patterns from MRI images by leveraging the strengths of Vision Transformers and deep segmentation networks. The dual model ensures that the spatial and contextual features are well-represented, leading to improved segmentation results. Nazir, K., et al. (2023) [23] implemented a 3D Convolutional method for tumor segmentation in MRI imaging. The feature pyramid network structure is enhanced with Kronecker convolutional layers, which capture features and improve segmentation accuracy. The 3D nature of the model allows it to handle volumetric data, which is particularly important for brain tumor segmentation in medical imaging. Ramamoorthy, H., et al. (2023) [24] implemented TransAttU-Net, a deep neural network for brain tumor segmentation in MRI images. The model combines a basic method with an attention method to improve the segmentation of tumors by emphasizing relevant features. The combination of attention systems enables the model to focus on tumor areas in images, which is potentially important for better segmentation results. Ramakrishnan, A. B., et al (2024) [25] proposed a hybrid CNN architecture for improved accuracy. We utilized oneAPI optimization techniques to adjust the weights and enhance the performance of the hybrid CNN model. By combining CNNs with optimization frameworks, the model achieves efficient classification while maintaining high accuracy.

3 Methodology

A CNN-Transformer Fusion Model is implemented to extract the spatial feature extraction capabilities of CNNs and the global contextual understanding of transformers for accurate brain tumor classification. The method involves three key components: feature extraction, sequence modeling, and classification, all underpinned by rigorous mathematical formulations as shown in Figure 1.

Feature Extraction: The input image is represented as $X \in R^{c_i * H * W}, where c_i = 3\ for$RGB color encoding, H is height and W is the width. CNN extracts the spatial features depth wise separable convolutions, producing a feature map$F \in R^{c_i * H * W}$ with equation (1).

$$F = \emptyset_{CNN}(X) \qquad (1)$$

Where $C_{out} = 1280, and\ H\ and\ W\ are\ redused\ spatial\ features.$ by aggregating all spatial features, applied global average pooling method with equation (2), for compacting all features (f_c).

$$f_c = \frac{1}{H'W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} F_{c,i,j} \qquad (2)$$

Sequential Modeling with Transformer Encoder: The pooled feature vector $f$is reshaped into a single-token sequence as $T = R^{1 * C_{out}(1280)}$. this sequence is transfer to encoder, which consists of 3 layers, each layer will have multi head self attention method and positional level feed forward method. The multi head attention method captured Query (Q), Key (K) and Value (V) from each vector with equation (3), (4) and (5). Where W is weights as the input dimension. The dot product of attention method is computed with equation (6).

$$Q_h = TW_h^Q (3)$$
$$K_h = TW_h^k (4)$$
$$V_h = TW_h^v (5)$$
$$Att(Q_h, K_h, V_h) = softmax\left(\frac{Q_h H_h^T}{\sqrt{d_k}}\right) V_h (6)$$

The output of all attention methods is concatenated linearly, and then it will provide final attention output.

Position-wise feed forward Network (FFN): In this each token will be considered into a 2 layer feed forward transformation, by equation (7) and positional level embedding with equation (8). In this W and b variables are updated parameters.

$$FFN(z) = \sigma(zW_1 * b_1)W_2 + b_2 \qquad (7)$$
$$PE_{pos,2i} = sin\left(\frac{pos}{10000^{d\frac{2i}{model}}}\right) \qquad (8)$$

The output is transformed through three layers of transformers. For the contextual embedding layer, the first token is passed to a fully connected layer for classification using equation (9) with these spatial and temporal features combined to give the final output.

$$yp = \emptyset_{FC}(z) (9).$$

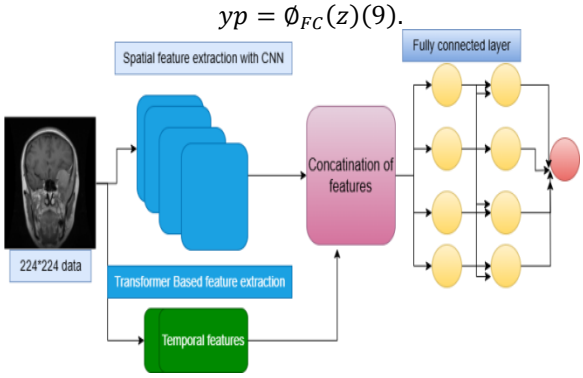

Figure 1: Proposed fusion models for brain tumor detection

3.1 Data set

The proposed model was trained on a Kaggle BRATS data set, which combines four classes: glioma, meningioma, no tumor, and pituitary. This dataset comprises 7023 brain images. All the samples are preprocessed into a 224*224 size. All the samples are then separated into training and testing sets in an 80:20 ratio. The samples of brain MRI are shown in Figure 2. All the samples are normalized to 0.465, 0.446, 0.416, with a standard deviation of 0.229, 0.224, 0.225, respectively. This ensures that no sample will dominate the other low-resolution samples.

3.2 hardware used for training

The proposed model was implemented using Python with TensorFlow and Keras libraries. All experiments were conducted on the Kaggle platform using a Tesla T4 GPU (16 GB VRAM) environment. The training was conducted for 10 epochs with a batch size of 32, using the Adam optimizer with an initial learning rate of 0.0001. A dropout rate of 0.2 was applied to reduce overfitting.

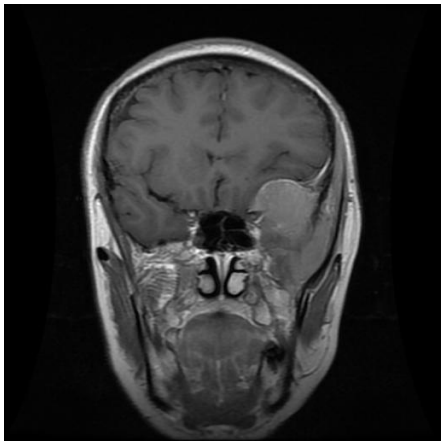

Figure 2: Sample brain MRI image.

4 Result analysis

The proposed fusion approach is iterated for 10 epochs, with a batch size of 16, and a learning rate of 0.0001, as shown in Table 1. The model achieved an accuracy of 74.72% with a training loss of 0.6639, while the test accuracy reached 86.92%, accompanied by a test loss of 0.4340. This indicates a strong baseline performance, likely attributed to the combination of MobileNetV2's efficient feature extraction and the Transformer's contextual understanding. Over successive epochs, the training accuracy improved steadily, reaching 91.88% by the final epoch, with the training loss decreasing to 0.2163. Similarly, the test accuracy increased to 91.76%, while the test loss reduced significantly to 0.1891, showcasing the model's enhanced capability to classify tumor categories accurately. From Figure 3, a marked improvement in test accuracy was observed between Epochs 8 and 10, where the model transitioned from 89.99% to 91.76%, with a corresponding reduction in test loss from 0.2319 to 0.1891.

Table 1: Parameters used for training the model

Parameter	Value
No. of Attention Heads	8
Hidden Size (FFN)	512
Dropout Rate	0.2
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32

The model achieves strong performance in the "Notumor" and "Pituitary" categories, with particularly high predictive reliability, evidenced by near-perfect metrics. The performance for "Glioma" and "Meningioma" shows slightly lower but still competitive results. These variations may stem from potential similarities in visual patterns between these tumor types, challenging the model's discriminative power. Nevertheless, the consistent improvement observed across all categories highlights the model's capacity to learn complex representations and adapt to varying class-specific patterns.

The overall classification observed from Table 2, with an accuracy of 91.8% across 841 test samples, underscores the model's generalization ability. Additionally, both the macro and weighted averages indicate a balanced performance across classes, ensuring that no individual category dominates or suffers from significant misclassification. Class-wise accuracy is illustrated in Figures 4 and 5.

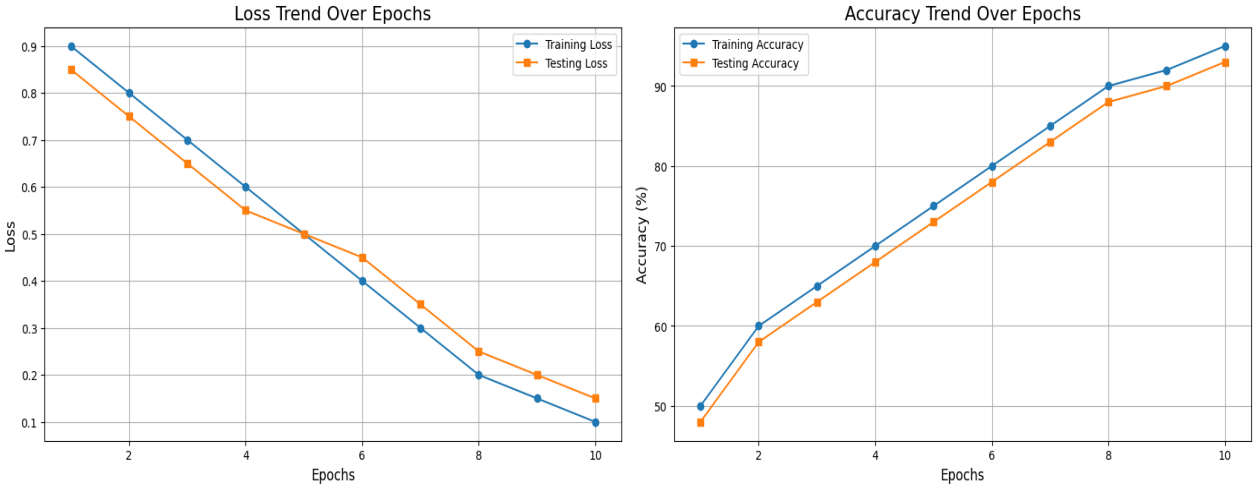

Figure 3: learning curves of the fusion model

Table 2: Performance of the proposed model

	P(%)	R(%)	F1(%)	Support

Glioma	93	89	91	190
Meningioma	91	85	87	186
Notumor	91	99	96	285
Pituitary	92	99	96	180
ACC			91.8	841
M-avg	92	91.5	91.8	841
W-avg	92	91.5	91.8	841

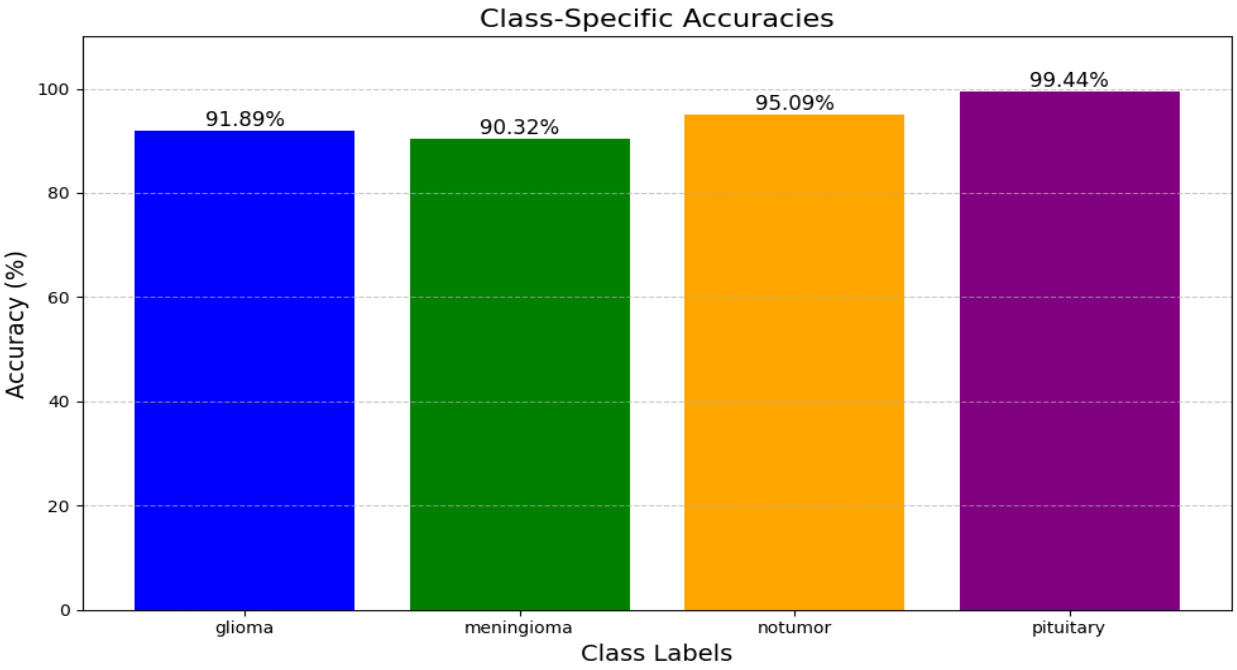

Figure 4: Class-wise performance of the proposed model

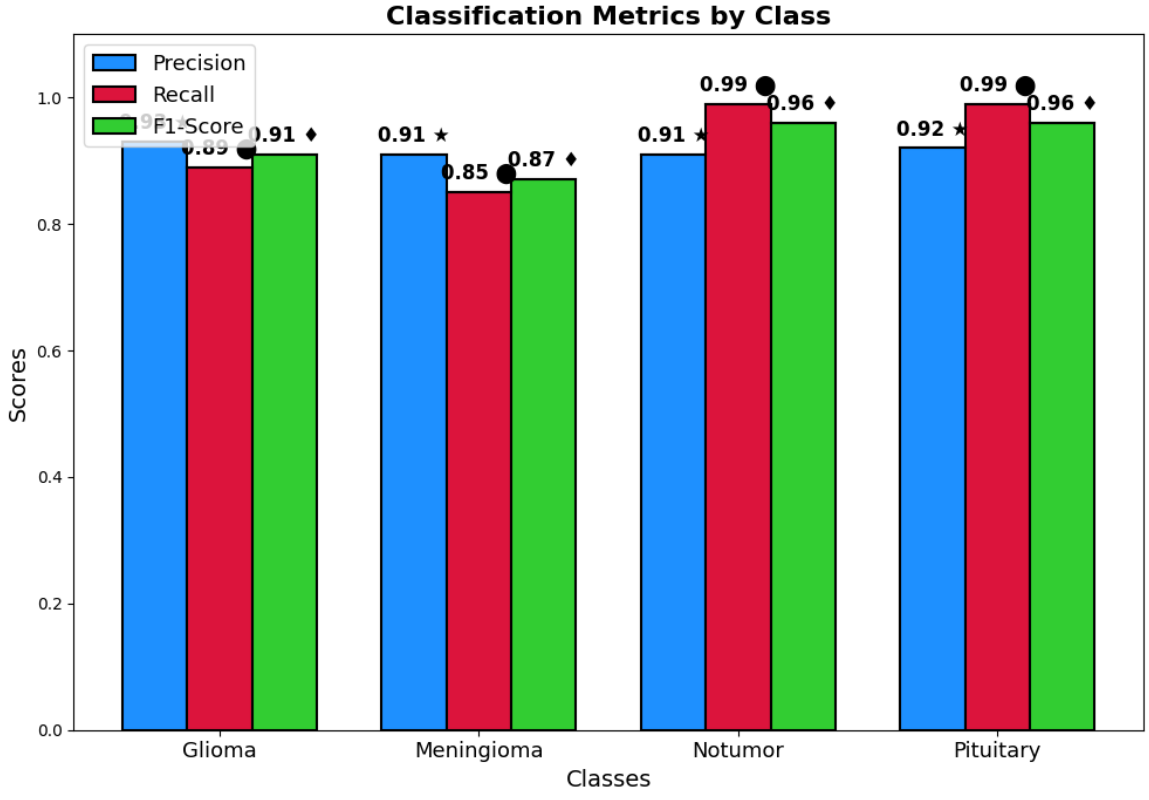

Figure 5: Class-wise performance of the fusion model in terms of precision, recall, and F1-score

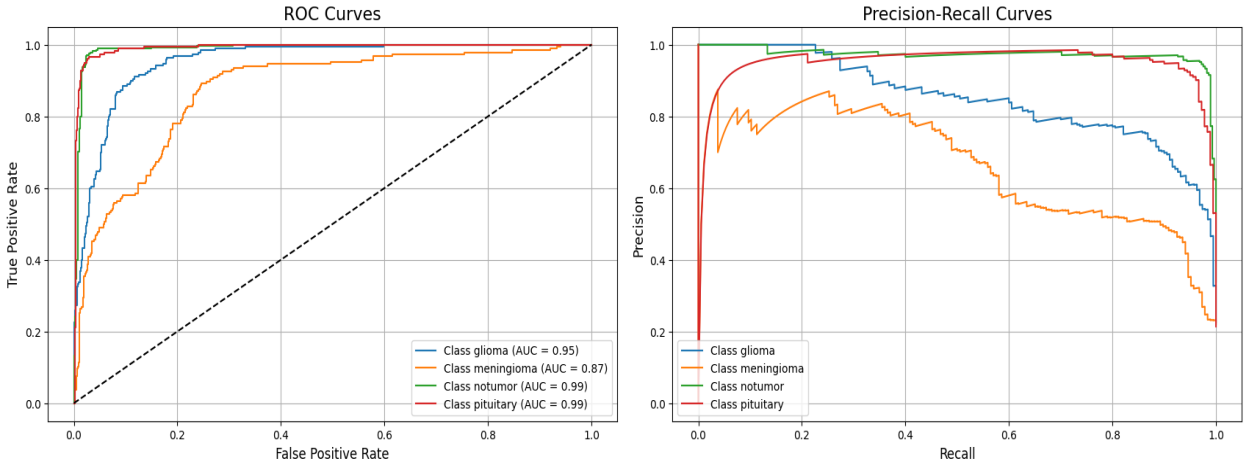

Figure 6: ROC and PR curve of proposed models

From Figure 6, the area under the ROC curve (AUC) highlights the model's effectiveness, with Glioma, Notumor, and Pituitary classes achieving high AUC values, indicating strong discrimination capabilities. However, the Meningioma class demonstrates slightly lower AUC, reflecting challenges in accurately distinguishing this class. Similarly, precision-recall curves reveal the relationship between positive prediction precision and sensitivity across different thresholds. Classes such as Notumor and Pituitary exhibit high performance, showcasing the robustness of the model in these cases. In contrast, the performance for the Meningioma class is comparatively modest, emphasizing areas for potential refinement.

Figures 7 and 8 illustrate feature maps extracted by the convolutional layers of the model for a sample input image. These maps provide a visual representation of the learned features at different layers, highlighting areas of importance and attention within the image. The feature maps capture various patterns, ranging from simple edges and textures in initial layers to more abstract and class-specific features in deeper layers. Bright regions within the maps indicate areas with strong activations.

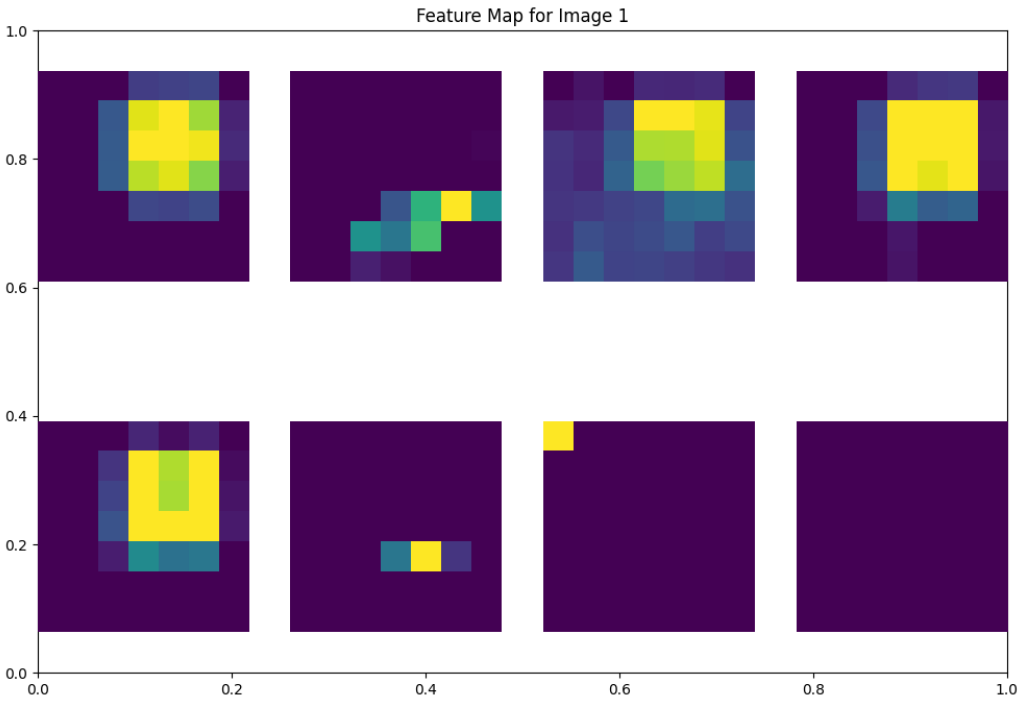

Figure 7 Feature extraction map of sample image-1

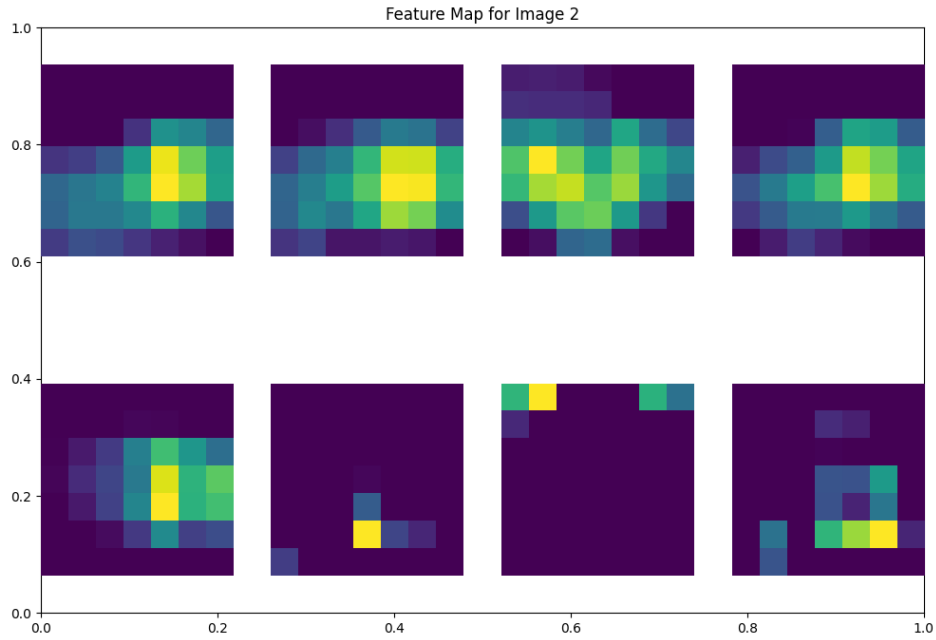

Figure 8 Feature extraction map of sample image-2

Figure 9 illustrates successful predictions by the model, where both the actual and predicted labels are identified as "glioma." These results indicate that the model effectively captured key features associated with gliomas, allowing for accurate classification. From Figure 10, where the actual label is "glioma," but the model incorrectly predicted "pituitary." Such an error highlights the overlap or similarity in visual features between glioma and pituitary cases, which may have led to confusion in the model's classification process.

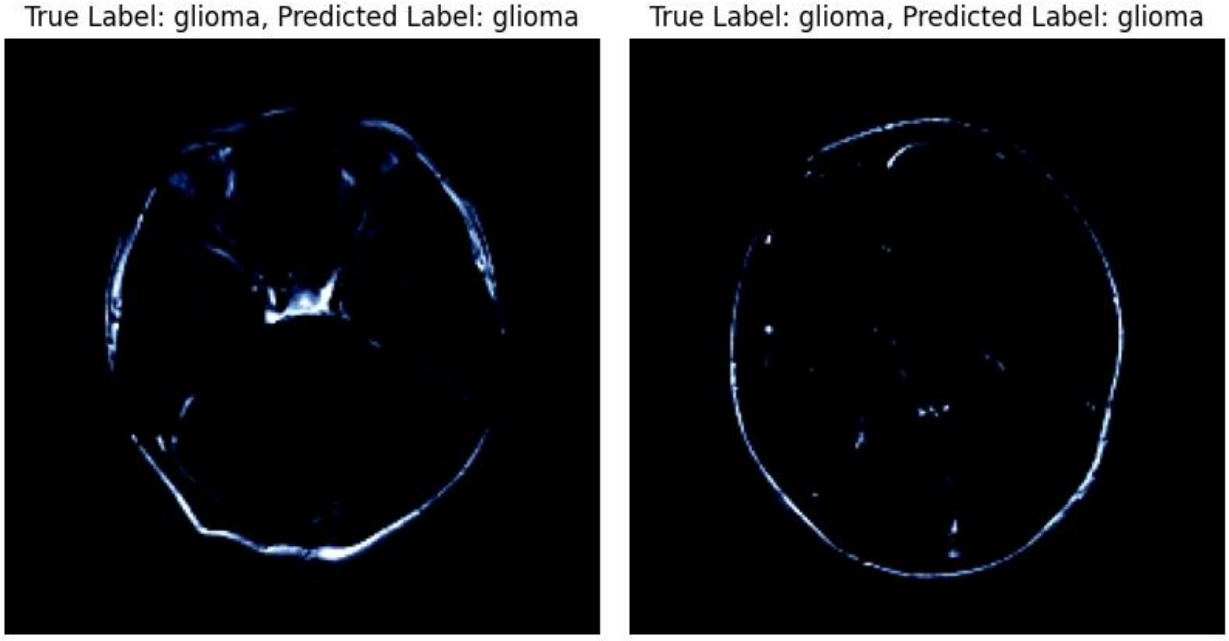

Figure 9: Actual and predicted labels of the proposed model after training

Figure 10: Misclassified sample by the proposed models

Table 3: Comparison of the proposed model with the prescribed models

Citation No.	Methodology	Dataset Used	Accuracy(%)
[22]	Dual Vision Transformer-DSUNET for brain tumor segmentation	MRI Brain Tumor	90.00
[26]	Gated residual recurrent neural networks	BraTS, ISBI	89.1
[27]	deep learning	BRATS	86.2
[28]	UTNet	BRATS	87.8
Proposed model	CNN-Transformer Fusion model	MRI Brain Tumor	91.8

Table 3 presents the performance of various methodologies for brain tumor segmentation and classification tasks using different datasets. The Dual Vision Transformer-DSUNET model, as reported in [22], achieves an accuracy of 90% on the same dataset. Similarly, the Gated Residual Recurrent Neural Networks employed in [26] show an accuracy of 89.1% when evaluated on the BraTS and ISBI datasets, reflecting their capability in processing temporal and spatial information. A deep learning-based approach utilized in [27] achieved an accuracy of 86.2% on the BRATS dataset, indicating its utility, albeit with slightly lower performance. The UTNet model, proposed in [28], reported an accuracy of 87.8% on the same BRATS dataset, leveraging its unique architectural enhancements for tumor segmentation. In comparison, the proposed CNN-Transformer Fusion model achieves an accuracy of 91.8% on the MRI Brain Tumor dataset, showcasing its superior ability to integrate the strengths of convolutional neural networks and transformers, resulting in improved feature representation and classification performance.

5 Conclusion

In this study, a hybrid CNN-Transformer Fusion Model was implemented for enhanced brain tumor classification. The model effectively combines the localized features, which are extracted with CNNs, with the global contextual understanding provided by Transformers. Comprehensive evaluations on a diverse dataset reveal the model's robust performance, achieving an overall accuracy of 91.8%, surpassing several existing state-of-the-art methods. The integration of CNN and a multi-layer Transformer Encoder enables the approach to learn complex spatial and temporal features, improving its performance to classify tumor types with high

consistency. At the same time, the model demonstrates remarkable performance in distinguishing "No Tumor" and "Pituitary" classes, minor challenges in classifying "Glioma" and "Meningioma" highlight opportunities for further optimization. Future work will focus on augmenting the dataset with additional samples and exploring advanced Transformer architectures to enhance discriminative capabilities.

References

[1] Hekmat, A., Zhang, Z., Khan, S. U. R., Shad, I., & Bilal, O. (2025). An attention-fused architecture for brain tumor diagnosis. *Biomedical Signal Processing and Control*, *101*,10.1016/j.bspc.2024.107221

[2] Benzorgat, N., Xia, K., &Benzorgat, M. N. E. (2024). Enhancing brain tumor MRI classification with an ensemble of deep learning models and transformer integration. *PeerJ Computer Science*, *10*, e2425. https://doi.org/10.7717/peerj-cs.2425

[3] Nobel, S. N., Swapno, S. M. R., Islam, M. B., Azad, A. K. M., Alyami, S. A., Alamin, M., ... & Moni, M. A. (2024). A Novel Mixed Convolution Transformer Model for the Fast and Accurate Diagnosis of Glioma Subtypes. *Advanced Intelligent Systems*, 2400566. https://doi.org/10.1002/aisy.202400566

[4] Katran, L. F., AlShemmary, E. N., & Al-Jawher, W. A. (2024). A Review of Transformer Networks in MRI Image Classification. *Al-Furat Journal of Innovations in Electronics and Computer Engineering*, 148-162. DOI:10.46649/fjiece.v3.2.12a.21.5.2024

[5] Mzoughi, H., Njeh, I., BenSlima, M., Farhat, N., &Mhiri, C. (2024). Vision transformers (ViT) and deep convolutional neural network (D-CNN)-based models for MRI brain primary tumors images multi-classification supported by explainable artificial intelligence (XAI). *The Visual Computer*, 1-20. DOI:10.1007/s00371-024-03524-x

[6] Alzahrani, S. M., & Qahtani, A. M. (2024). Knowledge distillation in transformers with tripartite attention: Multiclass brain tumor detection in highly augmented MRIs. *Journal of King Saud University-Computer and Information Sciences*, *36*(1), 101907. https://doi.org/10.1016/j.jksuci.2023.101907

[7] Nguyen-Tat, T. B., Nguyen, T. Q. T., Nguyen, H. N., & Ngo, V. M. (2024). Enhancing brain tumor segmentation in MRI images: A hybrid approach using UNet, attention mechanisms, and transformers. *Egyptian Informatics Journal*, *27*, 100528. DOI:10.13140/RG.2.2.18164.36485

[8] Gasmi, K., Ben Aoun, N., Alsalem, K., Ltaifa, I. B., Alrashdi, I., Ammar, L. B., ... & Shehab, A. (2024). Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique. *Frontiers in Neuroinformatics*, *18*, 1444650. https://doi.org/10.3389/fninf.2024.1444650

[9] Rasheed, Z., Ma, Y. K., Ullah, I., Al-Khasawneh, M., Almutairi, S. S., &Abohashrh, M. (2024). Integrating Convolutional Neural Networks with Attention Mechanisms for Magnetic Resonance Imaging-Based Classification of Brain Tumors. *Bioengineering*, *11*(7), 701. https://doi.org/10.3390/bioengineering11070701

[10] Pacal, I. (2024). A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *International Journal of Machine Learning and Cybernetics*, 1-19. https://doi.org/10.1007/s13042-024-02110-w

[11] Srinivas, B., Anilkumar, B., devi, N., & Aruna, V. B. K. L. (2024). A fine-tuned transformer model for brain tumor detection and classification. *Multimedia Tools and Applications*, 1-25. DOI:10.1007/s11042-024-19652-4

[12] Kang, M., Ting, F. F., Phan, R. C. W., Ge, Z., & Ting, C. M. (2024). A Multimodal Feature Distillation with CNN-Transformer Network for Brain Tumor Segmentation with Incomplete Modalities. *arXiv preprint arXiv:2404.14019*. https://doi.org/10.48550/arXiv.2404.14019

[13] Asiri, A. A., Shaf, A., Ali, T., Pasha, M. A., Khan, A., Irfan, M., & Alamri, S. (2024). Advancing brain tumor detection: harnessing the Swin Transformer's power for accurate classification and performance analysis. *PeerJ Computer Science*, *10*, e1867. https://doi.org/10.7717/peerj-cs.1867

[14] Tampu, I. E., Bianchessi, T., Blystad, I., Lundberg, P., Nyman, P., Eklund, A., & Haj-Hosseini, N. (2024). Pediatric brain tumor classification using deep learning on MR-images with age fusion. *Neuro-Oncology Advances*, vdae205. https://doi.org/10.1093/noajnl/vdae205

[15] Tabatabaei, S., Rezaee, K., & Zhu, M. (2023). Attention transformer mechanism and fusion-based deep learning architecture for MRI brain tumor classification system. *Biomedical Signal Processing and Control*, *86*, 105119.

[16] Aloraini, M., Khan, A., Aladhadh, S., Habib, S., Alsharekh, M. F., & Islam, M. (2023). Combining the transformer and convolution for effective brain tumor classification using MRI images. *Applied Sciences*, *13*(6), 3680.DOI:10.3390/app13063680

[17] Odusami, M., Damasevicius, R., Milieskaite-Belousoviene, E., &Maskeliunas, R. (2024). Multimodal Neuroimaging Fusion for Alzheimer's Disease: An Image Colorization Approach With Mobile Vision Transformer. *International Journal of Imaging Systems and Technology*, *34*(5), e23158. https://doi.org/10.1002/ima.23158

[18] Sun, X., Bhatti, U. A., Huang, M., & Zhang, Y. (2024). EF-UV: Feature Enhanced fusion of U-Net and VIT Transformer for Brain Tumor MRI Image Segmentation. DOI:10.21203/rs.3.rs-5329372/v1

[19] Chen, C., Wang, H., Chen, Y., Yin, Z., Yang, X., Ning, H., ... & Zhao, J. (2023). Understanding the brain with attention: A survey of transformers in brain sciences. *Brain-X*, *1*(3), e29. https://doi.org/10.1002/brx2.29

[20] Saleh, A. H., Atila, Ü., & Menemencioğlu, O. (2024). Multimodal Fusion for Enhanced Semantic Segmentation in Brain Tumor Imaging: Integrating Deep Learning and Guided Filtering Via Advanced 3D Semantic Segmentation Architectures. *International Journal of Imaging Systems and Technology*, *34*(5), e23152. http://dx.doi.org/10.1002/ima.23152

[21] Zebari, N. A., Mohammed, C. N., Zebari, D. A., Mohammed, M. A., Zeebaree, D. Q., Marhoon, H. A., ... & Martinek, R. (2024). A deep learning fusion model for accurate classification of brain tumours in Magnetic Resonance images. *CAAI Transactions on Intelligence Technology*. http://dx.doi.org/10.1049/cit2.12276

[22] Zakariah, M., Al-Razgan, M., &Alfakih, T. (2024). Dual Vision Transformer-DSUNET With Feature Fusion for Brain Tumor Segmentation. https://doi.org/10.1016/j.heliyon.2024.e37804

[23] Nazir, K., Madni, T. M., Janjua, U. I., Javed, U., Khan, M. A., Tariq, U., & Cha, J. H. (2023). 3D Kronecker Convolutional Feature Pyramid for Brain Tumor Semantic Segmentation in MR Imaging. *Computers, Materials & Continua*, *76*(3). DOI:10.32604/cmc.2023.039181

[24] Ramamoorthy, H., Ramasundaram, M., Raj, R. S. P., &Randive, K. (2023). TransAttU-Net Deep Neural Network for Brain Tumor Segmentation in Magnetic Resonance Imaging Réseau neuronal profondTransAttU-Net pour la segmentation des tumeurscérébrales avec l'imagerie par résonancemagnétique. *IEEE Canadian Journal of Electrical and Computer Engineering*. DOI:10.1109/ICJECE.2023.3289609

[25] Ramakrishnan, A. B., Sridevi, M., Vasudevan, S. K., Manikandan, R., &Gandomi, A. H. (2024). Optimizing brain tumor classification with hybrid CNN architecture: Balancing accuracy and efficiency through oneAPI optimization. *Informatics in Medicine Unlocked*, *44*, 101436. DOI:10.1016/j.imu.2023.101436

[26] Chen, J., Li, Y., Jin, Y., Luo, X., & Lu, G. (2019). Gated residual recurrent neural networks for multi-modal medical image segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 354-362. https://doi.org/10.1007/978-3-030-32248-9_40

[27] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA)*, 240-248. https://doi.org/10.1007/978-3-319-67558-9_28

[28] Gao, Y., Zhou, M., Metaxas, D. N., & Li, K. (2018). UTNet: a hybrid transformer architecture for medical image segmentation. *IEEE Transactions on Medical Imaging*, *37*(6), 1413-1423. https://doi.org/10.1109/TMI.2018.2806960