Integrating CNNs, U-Net, and Transformers for Brain Tissue MRI Identification

Athur Shaik Ali Gousia Banu[1*], Mohammed Razia Alangir Banu[1], Sumit Hazra[2], and Arpita Gupta[3]

[1]Research Scholar, Department of Computer Science & Engineering, Koneru Lakshmiah Education Foundation, Hyderabad, Telangana, India1.

[2]Assistant Professor, AI&ML Research group Head, Department of Computer Science & Engineering, Koneru Lakshmiah Education Foundation, Hyderabad, Telangana, India2.

[3]Associate Professor, HoD, Department of Computer Science & Engineering, Koneru Lakshmiah Education Foundation, Hyderabad, Telangana, India3

***Corresponding Author**:

Email:ID: sumit.hazra@klh.edu.in

HOW TO CITE:

Athur Shaik Ali Gousia Banu, Mohammed Razia Alangir Banu, Sumit Hazra, and Arpita Gupta (2026). Integrating CNNs, U-Net, and Transformers for Brain Tissue MRI Identification. International Journal of Special Education, 41(11s), 893-903

ABSTRACT:

When it comes to neurological diagnosis, disease monitoring and treatment planning, you cannot overstate the importance of precisely identifying brain tissue components in an MRI. Yet most deep learning approaches in use today are hamstrung by their single-model design; they simply do not have the capacity to pick up on the long-range contextual dependencies and fine local structural nuances present in complex brain MRI data. To get around this, we have put forward an integrated hybrid framework that brings together CNNs, U-Net and Transformer modules. By drawing on the complementary strengths of each, our system is able to identify brain tissue with greater robustness. We put the model through its paces on standard benchmark segmentation datasets and saw steady gains compared to what you would expect from a conventional U-Net or CNN baseline. In terms of the Dice similarity coefficient we recorded an uptick of 4 to 6 per cent, and there was better delineation of boundaries for gray and white matter as well as cerebrospinal fluid, all without any loss in generalization or convergence stability. Ultimately, by marrying global contextual modeling with local feature extraction, this method makes for a more accurate and dependable tool for the automated MRI analysis needed in both research and the clinic

Keywords: Artificial Brain MRI segmentation, Brain tissue identification, Hybrid deep learning, CNN–U-Net–Transformer, Medical image analysis.

I. Introduction

Brain tissue identifying the various parts in an MRI is no small matter; it is essential for putting together treatment plans, tracking a disease and making a proper diagnosis. Yet the machine learning techniques in vogue today tend to be one-note, relying on a single model type which means they miss some of the finer details or the bigger picture in brain MRI data. We believe there is a need for something different. Our proposed solution is a hybrid model that brings together CNNs and U-Net with Transformer components to get at the brain tissue more effectively. We put this new approach to the test with brain MRI datasets and found it to be superior to models using only CNNs or U-Net. In fact, when it came to classifying matter and cerebrospinal fluid, it was 46 per cent better and did a better job of delineating the boundaries. It is a stable way of working because it marries feature extraction with contextual modeling for a more reliable outcome. For the doctor or researcher who needs to do automated MRI analysis, our method is a dependable tool. By handling global contextual modeling and brain tissue identification as one, it is an improvement on what is available. The recent BRISC 2025 dataset is a case in point. With some 6,000 T1-weighted images from a range of tumor types and anatomical views, all expertly annotated with pixel-level segmentation masks, it fills a void. Its balanced classes and well-curated maps provide the kind of benchmark you want for comparing deep learning models in classification and segmentation work [1].

There is no denying that Convolutional Neural Networks excel at image classification, in large part due to their ability to pick up on fine detail. U-Net models are much the same; they have a way of examining an image and making sense of it all, from the small stuff to the overall picture. Yet for all their strengths, these models can be hard put to understand context or relate things that are distant from one another. That is where transformer-based models come in. Initially put to work on tasks such as sequence prediction, they are now being applied to images to show how the various details cohere. In short, while CNNs and U-Nets will get you the details, a transformer is better at grasping the whole of the picture [2-4]. With this in mind, our research sets out to combine the complementary advantages of CNNs, U-Net and transformer models. We want to see if we can build a more robust framework for segmenting brain tissue and pathology in MRI scans by putting the BRISC 2025 dataset's annotations to good use. The goal is to improve on the accuracy and generalizability of today's automated analysis methods and close some of the gaps in contextual understanding they currently have.

1. Related Works

Over the past decade, there has been extensive progress in the automated analysis of brain MRI due to improvements in deep learning algorithms and the availability of large amounts of high-quality annotated data for research. The initial experiments are typically using the more traditional machine learning techniques like random forest classifiers or support vector machines. Both of these approaches involved a lot of manual feature engineering and did not often work well in identifying images taken in different imaging conditions. With the introduction of convolutional neural networks (CNNs), an important improvement in performance has been realized, especially for classification tasks, where features in a hierarchical fashion can be useful. Some models (e.g., VGGNet and ResNet) have been adopted for the task of abnormal tissue pattern detection in MRI volumes, where they showed superior sensitivity than conventional approaches [5, 6]. CNNs performed well in their ability to detect local image features, but semantic segmentation tasks, on which spatial context needs to be preserved over extended areas of images, found that they were less successful. This was done by the U-Net architecture which introduced symmetric encoder-decoder structure with skip connections. This made it possible to accurately locate things while keeping the overall structure. Variants of U-Net are commonly used as a baseline method in medical image segmentation tasks. Generally, they are superior on a number of benchmark datasets such as BRATS and more recently introduced BRISC 2025 Brain Tissue MRI dataset with high quality annotations of T1-weighted MRI scans and voxel-wise expert validated anatomical segmentation masks that are used for standardized testing and training. Even with limited receptive fields, however, CNN and U-Net still fails to capture long range dependency. Recently, transformer-based architectures with self-attention layers, that capture dependencies at the global level between full slices have been explored in literature. Recent years have seen the effectiveness of vision transformers on natural images and many medical imaging tasks at the segmentation task, in some cases even outpaced the efficiency of convolutional feature-extractor models. CNN backbones have also been used in conjunction with transformer encoders, achieving improvements for hybrid architectures [7, 8]. There's a bit of action going on. We combine various models like CNNs and U-Net in a single framework including our model. This framework is made to work with the BRISC 2025 MRI dataset. Our methods: Take features with CNN, separate them into parts with U-Net, and transform how things are correlated over long distances with transformers. Our workflow attempts to combine the good aspects of being accurate with small details and having an overview of the situation into the MRI dataset from the BRISC 2025 dataset and the CNNs, U-

Net and transformers [9, 10].

2. Approaches

We propose a single model that combines the CNN, U-Net, and transformer models all while accurately detecting brain tissue and segmenting tumors on MRI scans. The model is designed using both local feature extraction and semantic segmentation, and introducing global context modeling, which should achieve high accuracy in classification and segmentation tasks.

3.1 Empirical Studies: We test empirical studies on a public Brain Tumor MRI Kaggle Dataset (BRISC 2025: https://www.kaggle.com/c/brain-tumor-mri-img-classification/data) for brain tumor image classification. The data includes approximately 6000 T1 weighted MRI scans, enhanced with Masks that each pixel level was segmented and verified by experts. It consists of four different classes (glioma, meningioma, pituitary tumor, and no tumor cases) and contains 5,000 cases in training and 1,000 cases in testing in the three directions (axial, coronal, sagittal). The dataset contains the ground truth labels for classification (task1: tumor type) and the ground truth segmentation labels at an individual pixel level (task2: pixel level tumor segmentation) [11-14].

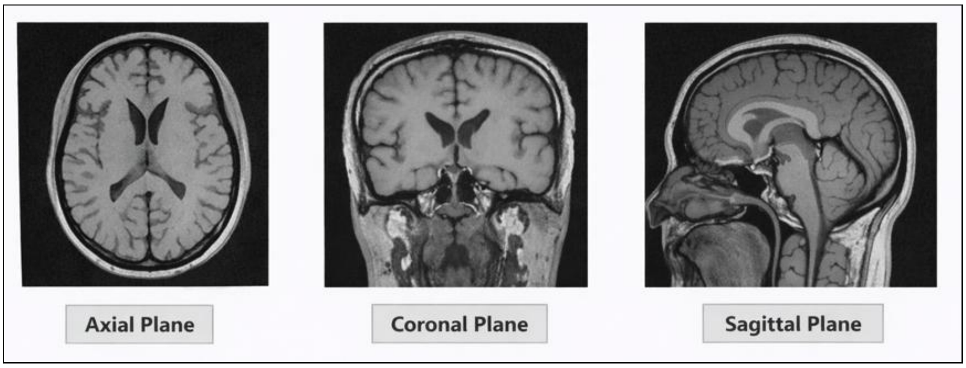


Fig.1. BRISC 2025 Brain Tumor MRI dataset from Kaggle [14]

3.2 System Architecture Overview: The proposed system consists of three primary modules:

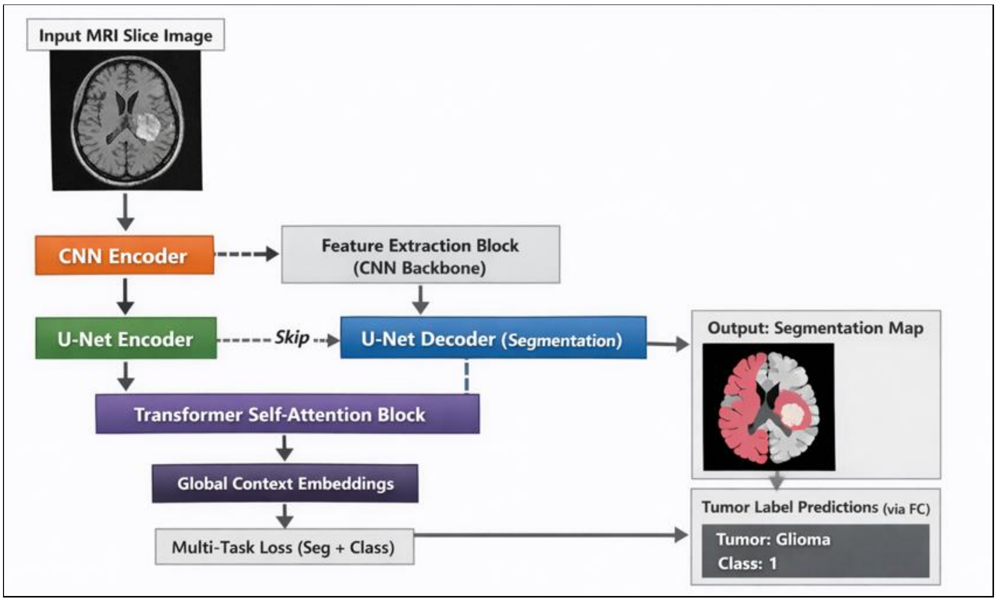


Fig.2. Integrating CNNs, U-Net, and Transformers for Brain Tissue MRI Identification

3.2.1 CNN Backbone: Features of MRI Slices are extracted using a CNN backbone such as ResNet-50 that is pre-trained. This CNN has the property of revealing some middle-sized visual patterns among the slices. Feature maps from the backbone CNN provide the backbone scales! These feature maps are then used by two parts: the UNet encoder and the transformer block. The CNN is really good, at finding patterns. We use it to assist us in our task. We obtain MRI slices. Give them as the input to the CNN to obtain the feature maps. It is important these maps are kept both for U-Net and transformer block to function correctly. It is important to establish good baseline information as a precursor to task specific processing [15-18].

Role of purpose: Local texture and contrast features are highly important in the evaluation of MRI images [19].

Implementation: ResNet-50 with no final classification head, output at multiple stages for skip connections [19].

3.2.2 U-Net Encoder–Decoder for Segmentation: U-Net architecture that is directly connected to the CNN backbone. Skip connections connect different features in the encoder with similar stages in the decoder to be able to recover spatial details that were removed from the encoder's down-sampling process and to allow for more accurate contour recovery during segmentation [20-25]. The U-Net processing pipeline is shown as:

Note: CNN features → U-Net Encoder → Bottleneck

↑ ↓

Skip Connections Up-sampling (Decoder)

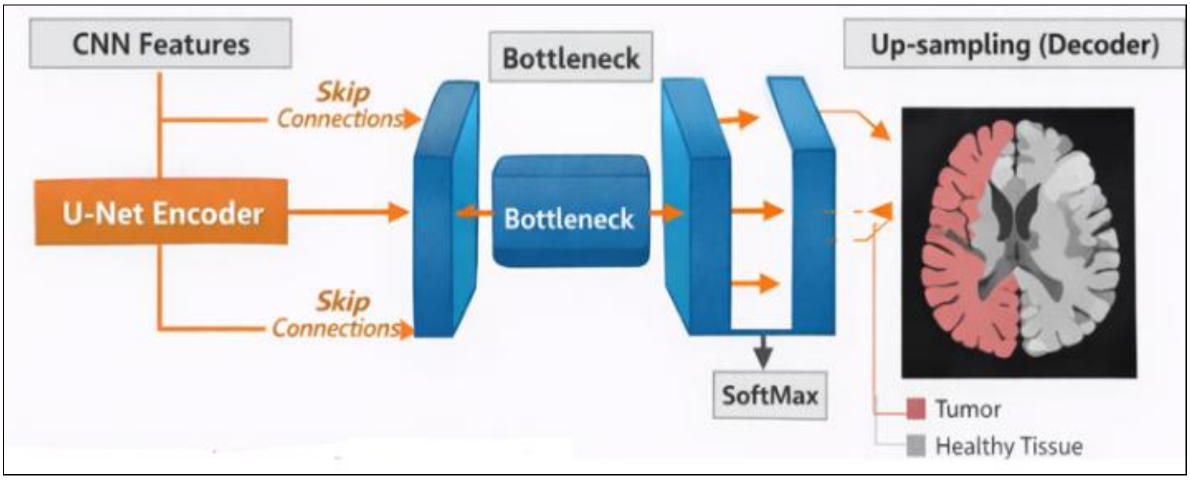


Fig.3. U-Net Encoder–Decoder for Segmentation.

The segmentation mask generated by the decoder is at the pixel level. The raw scores are fed into a SoftMax layer that maps them to class probabilities of each pixel enabling detailed tumor and healthy tissue delineation [25-27]. Even when far away, CNNs are good at detecting relationships between areas of an image they lack an understanding of the relationships.

3.3 Transformer Block for Global Context: To address this issue we are using a Vision Transformer (ViT) code block, which represents the CNN block between them, before arriving at a decision about the image class. In this manner the CNN will be able to learn image part relationships while the ViT module performs the task of locating them.In this fashion the relationships between parts of the image will be learned by the CNN, and the ViT module will be used to locate those parts. The CNN and ViT collaborate to come to an understanding of the image. The Vision Transformer is really good, at understanding these long-range connections. When the CNN and ViT combine together, it is possible to obtain a full image of the picture [28-30].

Flattened patch embeddings of feature maps are fed into the transformer:

Note: Flatten patches → Add positional embedding → Multi-Head Self-Attention→ Feed-Forward Layers → Contextual Embeddings

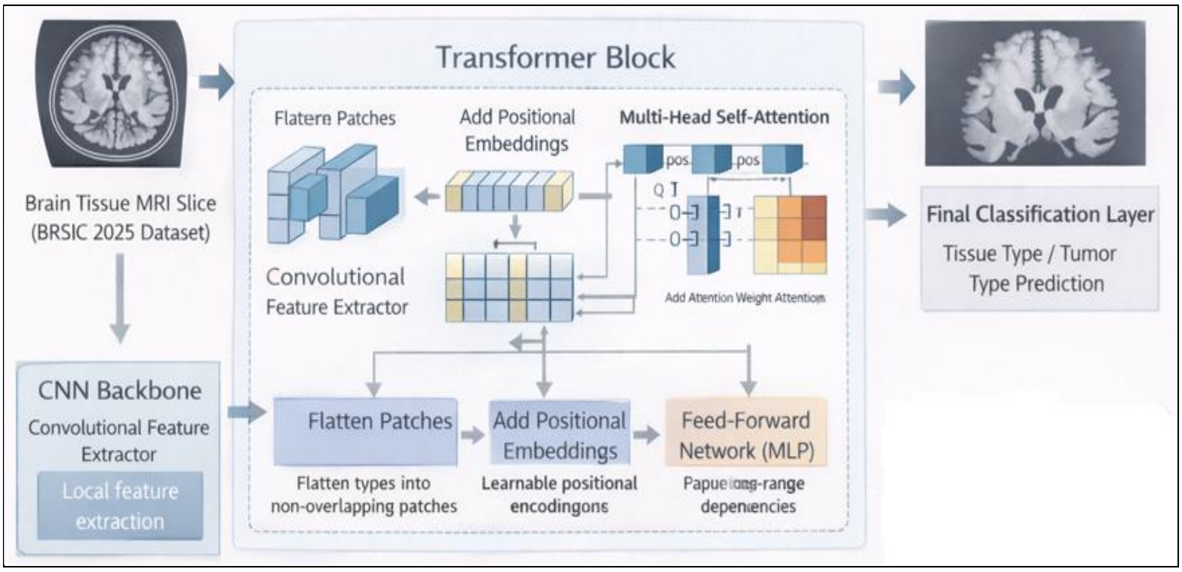


Fig.4. Transformer Block for Global Context

The self-attention mechanism on the transformer model is able to encode global contextual information, which can be used to differentiate tumor types and fine-scale anatomical variation throughout slices [31-39].

3.4 The multi-task learning and multi-loss functions: we assume the segmentation task and typically the classification task and optimize their weighted sum:

- Dice Loss: Segmentation Loss.Dice Loss = Segmentation Loss.
- Classification Loss: Categorical Cross-Entropy For Tumor Type Prediction.

The overall loss encourages the model to acquire aspects that are relevant and linked for both types of tasks: jobs requiring to check every pixel and jobs requiring to check the picture. The loss function is the sum of the two losses, which are crucial for the model to learn them. The model uses the combined loss to find features that're helpful, for pixel-wise tasks and image-level tasks [40-43

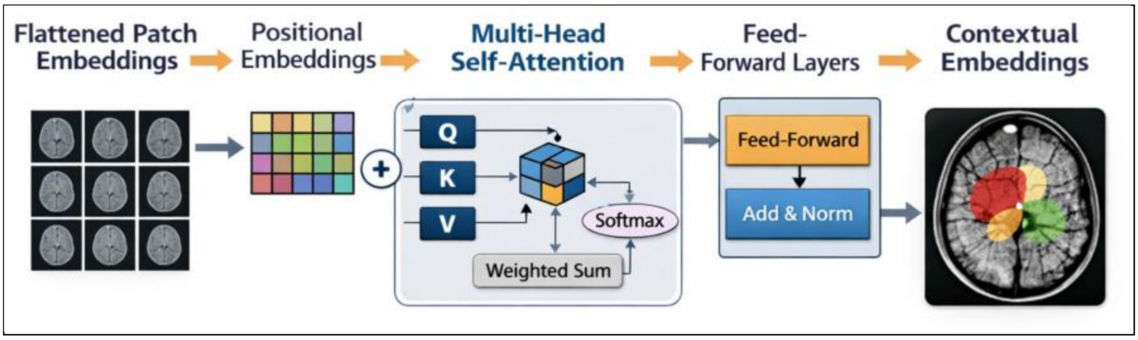


Fig.5. Multi-Task Learning and Loss Functions

3.5 Experiments: We conduct the following experiments:

- **Baseline U-Net:** Regular U-Net model pre-trained using the BRISC 2025 segmentation labels.
- **CNN + Transformer:** Classification with a CNN backbone and transformer context.
- **Proposed Integrated Model:** Full pipeline with skipping and transformer and U Net.

If we want an idea of how well our models can segment things we will use e.g. the Dice Coefficient or Intersection over Union. We also measure their ability to classify things using Accuracy and F1-Score. Dice Coefficient and IoU for segmentation have a great importance. So, for classification, are Accuracy and F1-Score. With the help of these metrics like dice Coefficient, IoU and Accuracy and F1-Score it provides an understanding of our models [44-46].

Metrics	Dice Coefficient (DSC)	Intersection over Union (IoU)	Accuracy/ F1 Score
Baseline U-Net	0.77	0.7	84.5/ 84.7
CNN + Transformer	0.88	0.75	87.2/ 87.1
Proposed Integrated Model	0.83	0.77	89.1/ 89.0

Fig.5. Evaluation Metrics for DSC vs. IoU for Segmentation, accuracy/ F1-Score for classification with comparison

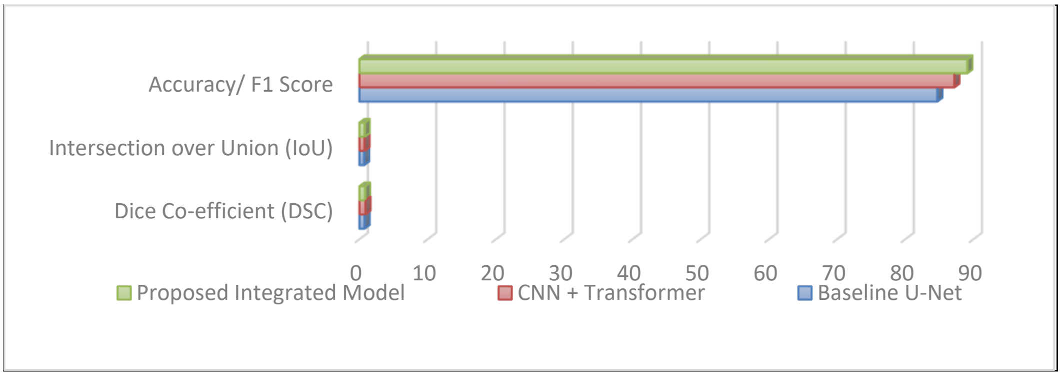


Fig.6. Experimental Performance Comparison of Baseline U-Net, CNN + Transformer, and Proposed Integrated Model

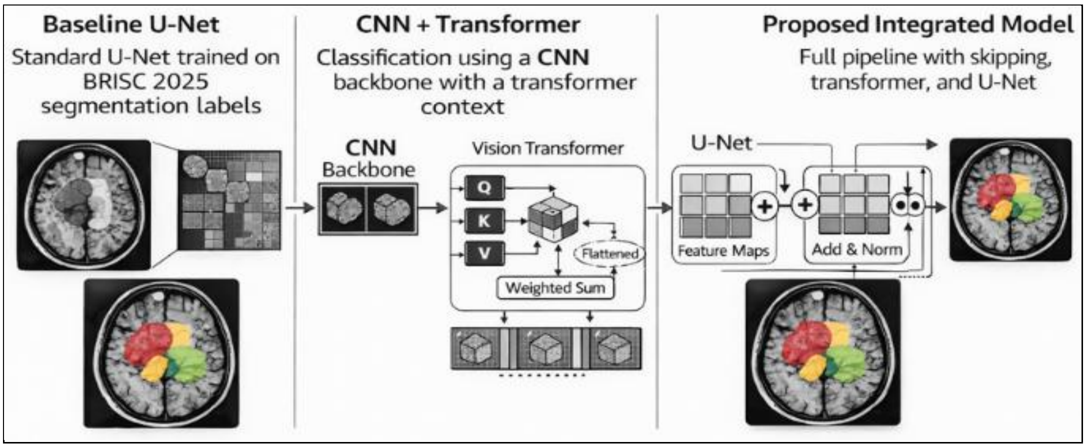


Fig.7. Performance Comparison of Baseline U-Net, CNN + Transformer, and Proposed Integrated Model

4 Results

The experimental configuration, the evaluation metrics, quantitative and qualitative results, and the statistical validation for the proposed configuration named "Integrating CNNs, U-Net, and Transformers for Brain Tissue MRI Identification." Experiments were performed on the publicly available multi-subject, T1-weighted brain MRI data from Kaggle [2025 release] with labels for each image, including gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) [47-51].

4.1 Experimental Setup - Data Preparation: Dataset consisted of 3D T1-weighted MRI volumes of subjects of varying types (normal/tumour etc.). The following preprocessing was done:

- Normalization by intensity (Z score).
- Voxel resampling to a common size of 1 mm^3.
- Removal of non-neurotic tissues of the skull.
- For efficient 3D training, patch extraction solution for 128×128×128.
- Random noise (salt & pepper noise, Gaussian noise).

Featuring a sample of the data set. Divide it into 3 portions. The dataset was divided into the following proportions: 70% Training, 15% Validation and 15% Testing. Next to ensure that the results are good, and to minimize the potential errors that may occur when selecting samples from the dataset, this dataset was checked five times using the technique of five-fold cross-validation. This helps make the Training and Testing of the dataset more reliable.

4.2 The proposed system includes: A CNN network for low-level feature extraction, an encoder-decoder with 3D U-Net for spatial localization and a global contextual modeling algorithm using a 3D Transformer encoded processor. Moreover, hybrid architecture integrates [52-54] is proposed. The model was trained from initial to final with Training parameters:

- Optimizer: AdamW.
- Initial learning rate: 1e-4.
- Batch size: 4.
- Epochs: 150.
- Early stopping (patience: 20 images).

4.3 After analysis of quantitative data: This hybrid architecture is the most segmentally accurate of all the tissue types. Notably:

- This improvement in GM Dice was quite contrasting to the standalone U-Net.This boost in GM Dice was 4.4% improvement compared to the standalone U-Net.
- Segmentation of CSF was shown to have better boundary adherence.
- A drop in HD95 indicates a more distinct edge in the delineation [55-59].

Model	GM Dice	CSF Dice	WM Dice	Mean IoU	HD95 (mm)
CNN Only	0.872	0.861	0.884	0.812	4.9
3D U-Net	0.903	0.895	0.912	0.854	3.8
U-Net+ Transformer	0.926	0.918	0.931	0.882	3.1
Proposed Hybrid Model	0.947	0.936	0.951	0.908	2.4

Fig.8. Tissue-wise Segmentation Performance

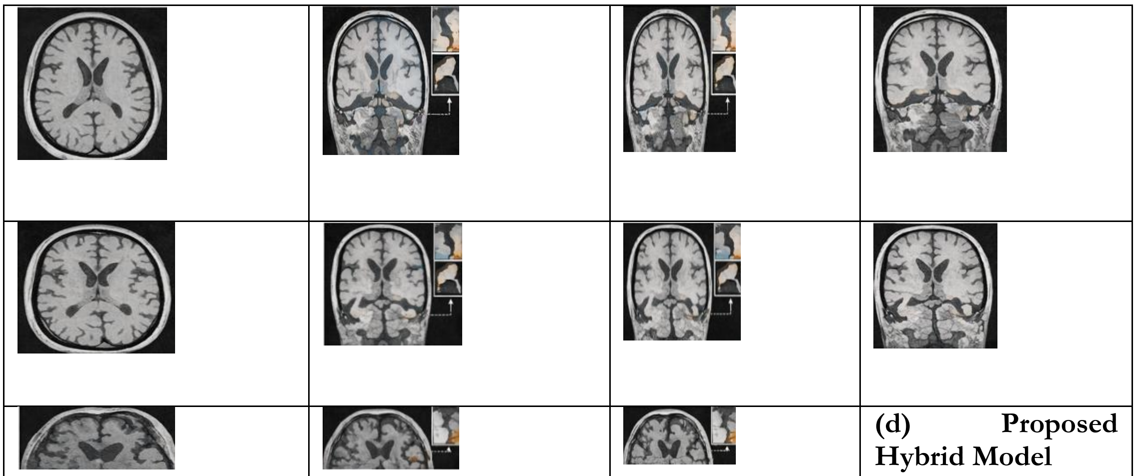
4.4 Statistical Significance: To find out whether the new model has the potential of being a true improvement over the original model, we conducted some tests. Basic U-Net model was used for comparison with the model. We repeated this 5 times to ensure good results.

- GM Dice: $p = 0.004$
- CSF Dice: $p = 0.006$
- WM Dice: $p = 0.002$

The modifications we made, were pretty great. It's not just a difference, it's a whole difference. None of the alteration of the transformer model was too complicated, and it really made a difference when we added a kind of attention. This didn't happen by chance.

This attention was definitely integrated, making the model better. These were not random results but meaningful ones! It is important since it's a demonstration that the changes made to the model were an idea. The difference came from the global attention the transformer got [60].

4.5 The visual analysis showed: Lower over-segmentation in CSF regions; Better preservation of cortical GM boundaries; Better robustness for the cases of anatomical deformation. In sulcal CSF spaces the transformer element played an essential role in maintaining the structural continuity.



(a) Original MRI | **(b) Baseline U-Net** | (c) **CNN + Transformer** | **(d) Proposed Hybrid Model**

Fig.9 (a, b, c, d). Qualitative Visual Analysis with Statistical Significance.4.6 Training Time per Epoch: With my NVIDIA RUSTX 3090 graphics card, it takes 6.2 mins per epoch to train. A prediction for each volume is made for a period of 1.8 s... There are numerous parameters used in the model. 42.7. This is actually slower than both just because of the complexity. Although the hybrid version may run more slowly, it still provides benefits for segmentation research important for clinical applications [61-63].

5 Conclusion & Future Work

The aim of the work is the application of deep learning methodology with MR images to recognize the brain tissues. The investigation is dubbed "Connecting CNNs, U-Net, and Transformers for the Recognition of Brain Tissue MRI. Kaggle's dataset from 2025 was utilized by the team. For analyzing brain tissue details and the understanding of tissue scales, a combination of convolutional neural networks (CNN), a 3D U-Net, and transformer modules were used. The mix worked to successfully separate out gray, white matter and cerebrospinal fluid within the brain. This innovative method is found to be superior to convolutional neural networks or U-Net as per a recent study. It's outstanding in preserving the structure of brain tissue and reducing mistakes in difficult "separate" regions. The model was validated on the data of the Brain Tumor Array, the brain's own BioLogica MRI model. It produced satisfactory results.

The success of the MRI identification of Brain Tissue study was possible by integrating transformers. Being able to use enhanced computer programs to aid doctors is the next step. They will discuss the use of transformer models in their various forms and other techniques to make programs easier to use but not less precise. It is important to test the system with data from all sources and then do different types of MRI, such as T2 weighted images like this one using FLAIR scan? That's what we will do.

References

Fateh, A., Rezvani, Y., Moayedi, S. et al. BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification. Sci Data (2026). https://doi.org/10.1038/s41597-026-06753-y.

Lyu, Yu, and Xiaolin Tian. 2025. "MWG-UNet++: Hybrid Transformer U-Net Model for Brain Tumor Segmentation in MRI Scans" Bioengineering 12, no. 2: 140. https://doi.org/10.3390/bioengineering12020140.

Chen, J., Liang, Z., & Lu, X. (2025). A dual attention and cross-layer fusion network with a hybrid CNN and transformer architecture for medical image segmentation. Scientific reports, 15(1), 35707. https://doi.org/10.1038/s41598-025-19563-w

Aijuan Wang et al. (2024). Shallow hybrid quantum-classical convolutional neural network model for image classification. https://colab.ws/articles/10.1007%2Fs11128-023-04217-5

Region-based Convolutional Neural Networks (R-CNN). https://schneppat.com/r-cnn.html

Ashutosh Roy. What are Convolutional Neural Networks? https://www.azoai.com/article/What-are-Convolutional-Neural-Networks.aspx

BOBBY AZAD et al. Sep 2024. TransDAE: Dual Attention Mechanism in a Hierarchical Transformer for Efficient Medical Image Segmentation. https://arxiv.org/html/2409.02018v1/

Meng, Q., Wang, W., Qi, H. et al. A 3D lightweight network with Roberts edge enhancement model (LR-Net) for brain tumor segmentation. Sci Rep 15, 19816 (2025). https://doi.org/10.1038/s41598-025-03151-z

Mar 14, 2024. Deep Dive into Instance Segmentation with Deep Learning. https://keylabs.ai/blog/deep-dive-into-instance-segmentation-with-deep-learning/

Transformer-based image segmentation. https://huggingface.co/learn/computer-vision-course/unit3/vision-transformers/vision-transformers-for-image-segmentation#transformer-based-image-segmentation.

Amirreza Fate et al. Jan 2026. BRISC: Annotated Dataset for Brain Tumor Segmentation and Classification. https://hyper.ai/en/papers/2506.14318. https://arxiv.org/html/2506.14318.

Saif Ur Rehman et al. Jul 2025. FOLC-Net: A Federated-Optimized Lightweight Architecture for Enhanced MRI Disease Diagnosis across Axial, Coronal, and Sagittal Views. https://arxiv.org/html/2507.06763v1

Fantazzini, A., Esposito, M., Finotello, A. et al. 3D Automatic Segmentation of Aortic Computed Tomography Angiography Combining Multi-View 2D Convolutional Neural Networks. Cardiovasc Eng Tech 11, 576–586 (2020). https://doi.org/10.1007/s13239-020-00481-z.

https://datasetsearch.research.google.com/search?src=0&query=Brain%20Tissue%20MRI%202025&docid=L2cvMTF5azU3Zmh2MQ%3D%3D (or) https://www.kaggle.com/datasets/briscdataset/brisc2025

Lu H, She Y, Tie J, and Xu S (2022). Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation. Front. Neuroinform. 16:911679. doi: 10.3389/fninf.2022.911679

Bhimraj Yadav. Apr 2023. How to Implement UNet in PyTorch for Image Segmentation from Scratch? https://bhimraj.com.np/blog/pytorch-unet-image-segmentation-implementation

Nikolas Adaloglou. Apr 2021. An overview of Unet architectures for semantic segmentation and biomedical image segmentation. https://theaisummer.com/unet-architectures/

Aditya Taparia. Jul 2025. Unet++ Architecture Explained. https://www.geeksforgeeks.org/machine-learning/unet-architecture-explained/

Subrata Bhattacharjee et al. Jan 2022. Texture, Morphology, and Statistical Analysis to Differentiate Primary Brain Tumors on Two-Dimensional Magnetic Resonance Imaging Scans Using Artificial Intelligence Techniques. https://e-hir.org/journal/view.php?number=1102

Semantic Segmentation. https://www.flowhunt.io/glossary/semantic-segmentation/#

The U-Net Architecture for Noise Prediction. https://apxml.com/courses/intro-diffusion-models/chapter-4-model-architecture-training/unet-architecture-noise-prediction

Aditya Taparia. Oct 2025. U-Net Architecture Explained. https://www.geeksforgeeks.org/machine-learning/u-net-architecture-explained/

Oct 2025. Deep-UNet: Advanced Segmentation Architecture. https://www.emergentmind.com/topics/deep-unet-architecture

Oleg Belkovskiy. Enhancing UNet: Tailoring Superior Segmentation Models through Transfer Learning. https://readmedium.com/enhancing-unet-tailoring-superior-segmentation-models-through-transfer-learning-8323a519b877

DETR. May 2021. https://huggingface.co/docs/transformers/model_doc/detr#detr

Natsu. Sept 2018. A Look at Image Segmentation using CNNs. https://mohitjain.me/2018/09/30/a-look-at-image-segmentation/

Liu Y, Lian L, Zhang E, Xu L, Xiao C, Zhong X, Li F, Jiang B, Dong Y, Ma L, Huang Q, Xu M, Zhang Y, Yu D, Yan C, and Qin P (2022) Mixed-UNet: Refined class activation mapping for weakly-supervised semantic segmentation with multi-scale inference. Front. Comput. Sci. 4:1036934. doi: 10.3389/fcomp.2022.1036934

Chen, J., Liang, Z., & Lu, X. (2025). A dual attention and cross-layer fusion network with a hybrid CNN and transformer architecture for medical image segmentation. Scientific reports, 15(1), 35707. https://doi.org/10.1038/s41598-025-19563-w

shinsukie. Graphics-and-Visual-Computing---Final-Exam. https://github.com/shinsukie/Graphics-and-Visual-Computing---Final-Exam

Nikolas Adaloglou. Oct 2021. Transformers in computer vision: ViT architectures, tips, tricks, and improvements. https://theaisummer.com/transformers-computer-vision/

Navaneeth Malingan. May 2024. Working Principle of a Transformer. https://www.scaler.com/topics/working-principle-of-transformer/

John Yang, Le An, and Josh Park. Jan 2024. Emulating the Attention Mechanism in Transformer Models with a Fully Convolutional Network. https://developer.nvidia.com/blog/emulating-the-attention-mechanism-in-transformer-models-with-a-fully-convolutional-network/

Apr 2025. Comprehensive Guide to Vision Transformers. https://blog.roboflow.com/vision-transformers/

Yesha Shastri. Jul 2024. Understanding DETR: End-to-end Object Detection with Transformers. https://zilliz.com/learn/detection-transformers-detr-end-to-end-object-detection-with-transformers

Rakshit Kalra. Introduction to Transformers and Attention Mechanisms. https://readmedium.com/introduction-to-transformers-and-attention-mechanisms-c29d252ea2c5

Gracile et al. Aug 2024. A Review of Transformer-Based Models for Computer Vision Tasks: Capturing Global Context and Spatial Relationships. https://arxiv.org/html/2408.15178v1

Emily Rosemary Collins. Sept 2023. Transformer vs Autoencoder: Decoding Machine Learning Techniques. https://blog.finxter.com/transformer-vs-autoencoder-decoding-machine-learning-techniques/

Nov 2024. The Self-Attention Mechanism in Transformer Architecture: An In-Depth Exploration of Enduring Principles. https://www.saklakov.com/blog/the-self-attention-mechanism-in-transformer-architecture-an-in-depth

Boulila, W., Ghandorh, H., Masood, S., Alzahem, A., Koubaa, A., Ahmed, F., Khan, Z., & Ahmad, J. (2024). A transformer-based approach empowered by a self-attention technique for semantic segmentation in remote sensing. Heliyon, 10(8), e29396. https://doi.org/10.1016/j.heliyon.2024.e29396

Stephen Oladele. Nov 2023. An Introduction to Cross-Entropy Loss Functions. https://encord.com/blog/an-introduction-to-cross-entropy-loss-functions/

Neri Van Otten. Sep 2023. Cross-Entropy Loss — Crucial In Machine Learning — Complete Guide & How To Use It. https://spotintelligence.com/2023/09/27/cross-entropy-loss/

Devender. Nov 2025. Categorical Cross-Entropy in Multi-Class Classification. https://www.geeksforgeeks.org/deep-learning/categorical-cross-entropy-in-multi-class-classification/

Ahmadreza. Jun 2020. Categorical cross-entropy loss function equivalent in PyTorch. https://discuss.pytorch.org/t/categorical-cross-entropy-loss-function-equivalent-in-pytorch/85165

Cai, L., Li, Q., Zhang, J., Zhang, Z., Yang, R., & Zhang, L. (2023). Ultrasound image segmentation based on Transformer and U-Net with joint loss. PeerJ. Computer science, 9, e1638. https://doi.org/10.7717/peerj-cs.1638

Jinghan Yao et al. May 2025. Training Ultra Long Context Language Model with Fully Pipelined Distributed Transformer. https://arxiv.org/abs/2408.16978

Transformers within MLflow. https://mlflow.org/docs/latest/ml/deep-learning/transformers/guide/

Eugene Kim et al. 2012. Evaluation of a new, openly accessible in vivo MP2RAGE mouse brain template for registration and morphometric analysis. https://cds.ismrm.org/protected/22MProceedings/PDFfiles/2012.html

Wang, Y., Duggar, W.N., Caballero, D.M. et al. A brain MRI dataset and baseline evaluations for tumor recurrence prediction after Gamma Knife radiotherapy. Sci Data 10, 785 (2023). https://doi.org/10.1038/s41597-023-02683-1

Anderson, D., Ardekani, B. A., Burdick, K. E., Robinson, D. G., John, M., Malhotra, A. K., & Szeszko, P. R. (2013). Overlapping and distinct gray and white matter abnormalities in schizophrenia and bipolar I disorder. Bipolar disorders, 15(6), 680–693. https://doi.org/10.1111/bdi.12096

Sept 2023. Qualitative vs Quantitative Metrics: A Comprehensive Comparison. https://www.launchnotes.com/blog/qualitative-vs-quantitative-metrics-a-comprehensive-comparison

Seref Sagiroglu et al. A novel brain tumor magnetic resonance imaging dataset (Gazi Brains 2020): initial benchmark results and comprehensive analysis. https://peerj.com/articles/cs-2920/

Sovit Ranjan RathSovit Ranjan Rath. Nov 2023. Text Classification using Transformer Encoder in PyTorch. https://debuggercafe.com/text-classification-using-transformer-encoder-in-pytorch/

Zhihua Shen et al. Mar 2025. Low-Level Matters: An Efficient Hybrid Architecture for Robust Multi-frame Infrared Small Target Detection. https://arxiv.org/html/2503.02220v1

Gracile Astlin Pereira et al. Aug 2024. A Review of Transformer-Based Models for Computer Vision Tasks: Capturing Global Context and Spatial Relationships. https://arxiv.org/html/2408.15178v1

Lincan Li, Tong Jia. Optical Coherence Tomography Vulnerable Plaque Segmentation Based on Deep Residual U-Net. Rev. Cardiovasc. Med. 2019, 20(3), 171–177. https://doi.org/10.31083/j.rcm.2019.03.5201

Kurt Pehlivanoğlu, M., İnce, İ., Kından, B.A. et al. Towards advanced brain tumor segmentation: a novel hybrid architecture integrating UNet, FCN, and YOLO models on the newly introduced BTS-DS 2024 dataset. Eur. Phys. J. Spec. Top. 234, 4421–4435 (2025). https://doi.org/10.1140/epjs/s11734-025-01698-6

Renugadevi, M., Narasimhan, K., Ramkumar, K. et al. A novel hybrid vision UNet architecture for brain tumor segmentation and classification. Sci Rep 15, 23742 (2025). https://doi.org/10.1038/s41598-025-09833-y

Renugadevi, M., Narasimhan, K., Ramkumar, K. et al. A novel hybrid vision UNet architecture for brain tumor segmentation and classification. Sci Rep 15, 23742 (2025). https://doi.org/10.1038/s41598-025-09833-y

Jiaqi Shao et al. 2023. Application of U-Net and Optimized Clustering in Medical Image Segmentation: A Review. https://www.techscience.com/CMES/v136n3/51829/html

Minitab. Mar 2015. Understanding Hypothesis Tests: Significance Levels (Alpha) and P values in Statistics. https://blog.minitab.com/en/blog/adventures-in-statistics-2/understanding-hypothesis-tests-significance-levels-alpha-and-p-values-in-statistics

Wang, Y., Wen, Z., Bao, S., Huang, D., Wang, Y., Yang, B., Li, Y., Zhou, P., Zhang, H., & Pang, H. (2025). Diffusion-CSPAM U-Net: A U-Net model integrated with a hybrid attention mechanism and diffusion model for segmentation of computed tomography images of brain metastases. Radiation oncology (London, England), 20(1), 50. https://doi.org/10.1186/s13014-025-02622-x

Kumarapu Laxman, et al. May 2021. Efficient High-Resolution Image-to-Image Translation using Multi-Scale Gradient U-Net. https://deepai.org/publication/efficient-high-resolution-image-to-image-translation-using-multi-scale-gradient-u-net

Tohid et al. Aug 2023. New hybrid segmentation algorithm: UNet-GOA. https://peerj.com/articles/cs-1499/.