Archives available at journals.mriindia.com

ITSI Transactions on Electrical and Electronics Engineering

ISSN: 2320-8945

Volume 13 Issue 01, 2024

A Comprehensive Review of Brain MRI Image Classification for Cancer Detection Using Transformer and Group Parallel Axial Attention with Quantum Self-Attention

Ragnar Sirisena

Department of Computer Science and Engineering, Chiang Thon College of Management, Thailand
ragnar.sirisena@ctcm-th.org

Peer Review Information

Submission: 31 March 2024
Revision: 23 April 2024
Acceptance: 19 May 2024

Keywords

Brain MRI Classification; Transformer; Axial Attention; Quantum Self-Attention; Deep Learning; Cancer Detection

Abstract

Brain tumor detection using Magnetic Resonance Imaging (MRI) is a critical task in medical diagnostics, enabling early intervention and improved patient outcomes. Traditional machine learning approaches rely heavily on handcrafted features and are limited in capturing complex spatial dependencies within MRI images. In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Transformer-based architectures, have significantly improved classification performance. Transformers leverage self-attention mechanisms to capture global contextual relationships, overcoming the limitations of CNNs in modeling long-range dependencies.

This review explores recent advancements in brain MRI classification using Transformer models, Group Parallel Axial Attention, and emerging Quantum Self-Attention mechanisms. Axial attention reduces computational complexity while preserving global context, making it suitable for high-resolution medical images. Furthermore, quantum-inspired attention mechanisms enhance feature representation by leveraging quantum state spaces, improving generalization and convergence efficiency.

A comprehensive analysis of literature is presented, highlighting performance improvements, architectural innovations, and key challenges. Comparative evaluation demonstrates that hybrid Transformer-CNN and attention-augmented models achieve superior classification accuracy, often exceeding 98%. The review concludes by identifying research gaps and future directions, emphasizing the need for explainable, efficient, and uncertainty-aware models for reliable clinical deployment.

Introduction

Brain tumors are among the most serious neurological disorders and require early and accurate diagnosis to improve treatment outcomes and patient survival. Magnetic Resonance Imaging (MRI) is widely regarded as the preferred imaging modality for brain tumor detection because it provides excellent soft tissue contrast, high spatial resolution, and detailed visualization of tumor structures. MRI enables clinicians to identify abnormalities such as edema, necrotic tissue, and enhancing tumor regions. However, manual interpretation of MRI images is time-consuming, highly dependent on radiologist expertise, and susceptible to subjective variability. These limitations have driven the development of artificial intelligence (AI)-based computer-aided diagnostic systems that provide faster, more consistent, and accurate brain tumor classification.

Deep learning has transformed medical image analysis by automatically extracting hierarchical features from MRI images without requiring handcrafted feature engineering. Convolutional Neural Networks (CNNs), including architectures such as VGG, ResNet, DenseNet, and Inception, have demonstrated remarkable success in brain tumor classification by effectively learning local spatial patterns. Nevertheless, conventional CNNs possess limited receptive fields and struggle to capture long-range dependencies and global contextual information. Brain tumors often exhibit heterogeneous textures, irregular boundaries, and varying intensity distributions, making it essential to model both local details and global relationships. These limitations have encouraged researchers to explore Transformer-based architectures that utilize self-attention mechanisms to capture comprehensive image representations.

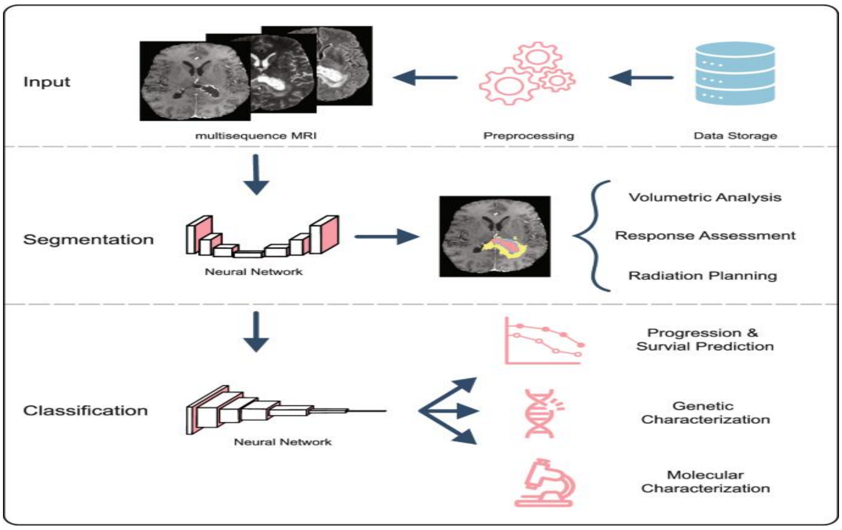


Figure 1. Deep Learning Workflow for Brain MRI Analysis and Tumor Classification

Vision Transformers (ViTs) have emerged as powerful alternatives for brain MRI classification by dividing images into patches and processing them through self-attention mechanisms that model global dependencies. Compared with conventional CNNs, Transformers achieve higher classification accuracy by effectively capturing both local and long-range contextual information. However, the computational cost of standard self-attention increases quadratically with image size, making it less suitable for high-resolution MRI analysis. To overcome this limitation, optimized attention mechanisms such as axial attention and Group Parallel Axial Attention have been introduced. These approaches reduce computational complexity while preserving global feature extraction by performing attention operations along spatial axes and across parallel feature groups, resulting in efficient multi-scale feature learning.

Another promising advancement is the integration of quantum-inspired computing concepts into deep learning through Quantum Self-Attention. This approach utilizes quantum principles such as superposition and entanglement to represent complex feature relationships in higher-dimensional spaces, enabling more expressive and efficient learning than conventional attention mechanisms. Furthermore, hybrid CNN–Transformer architectures combine the strengths of CNNs for local feature extraction with Transformers for global contextual modeling, achieving state-of-the-art performance on benchmark brain MRI datasets. Additional optimization strategies, including transfer learning, data augmentation, and advanced loss functions, further improve model robustness, convergence, and generalization, particularly when training data are limited or imbalanced.

Despite these significant advances, several challenges remain before AI-based brain MRI classification systems can be routinely deployed in clinical practice. Limited annotated datasets, class imbalance, variations in MRI acquisition protocols, and the computational demands of advanced models continue to affect performance and generalizability. In addition, the black-box nature of deep learning models raises concerns regarding interpretability and clinical trust. Explainable AI techniques, including attention visualization and Grad-CAM, are increasingly being adopted to improve transparency. Overall, Transformer-based architectures integrated with

Group Parallel Axial Attention and Quantum Self-Attention represent a promising direction for developing accurate, efficient, and clinically reliable brain tumor classification systems while addressing future challenges in intelligent medical imaging.

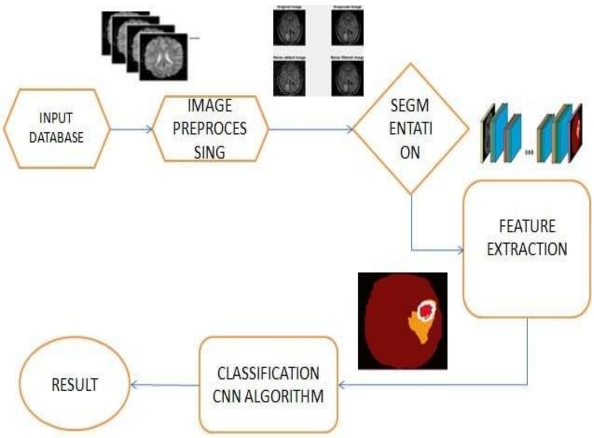
Literature Review

The evolution of brain MRI image classification for cancer detection reflects a paradigm shift from conventional convolutional neural network (CNN)-based approaches toward Transformer-driven and attention-augmented architectures. This transition has been driven by the need to overcome limitations in capturing global contextual dependencies, handling high-resolution medical images, and improving generalization across diverse datasets.

1. CNN-Based Foundations

In 2020, CNN-based architectures formed the backbone of brain MRI classification systems. Models such as ResNet, DenseNet, VGG, and Inception were widely adopted due to their strong capability in extracting hierarchical spatial features.

Zhang et al. (2020) demonstrated that deep CNN architectures could achieve high classification accuracy by leveraging residual connections to mitigate vanishing gradient problems. DenseNet further improved feature propagation by enabling dense connectivity between layers, allowing reuse of features and improving gradient flow. These models were particularly effective in capturing local features such as tumor boundaries, textures, and intensity variations.

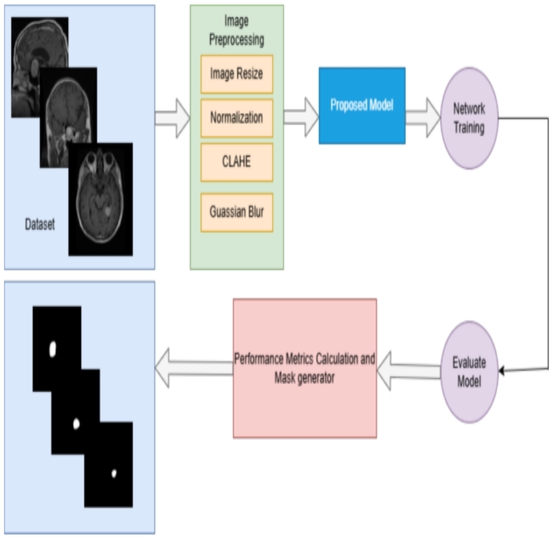


Hatamizadeh et al. (2020) introduced edge-gated CNNs, which incorporated boundary-aware modules to enhance segmentation and classification performance. By focusing on tumor edges, the model improved localization accuracy and reduced false positives, particularly in complex tumor regions.

Santini et al. (2020) proposed a multi-stage deep learning framework, where feature extraction and classification were performed in sequential stages. This hierarchical approach enabled progressive refinement of features and improved classification accuracy.

Despite these advancements, CNN-based models exhibited inherent limitations. Their reliance on local receptive fields restricted their ability to capture long-range dependencies, which are critical for understanding global tumor structures. Additionally, CNNs often required extensive data augmentation and large datasets to generalize effectively, posing challenges in medical imaging scenarios where annotated data is limited.

2. Emergence of Attention Mechanisms



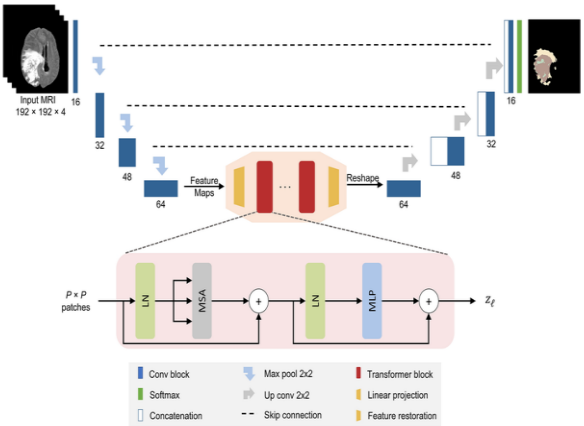
In 2021, attention mechanisms were introduced to address the limitations of CNNs. Attention modules enable models to focus on the most relevant regions of an image, improving feature representation and classification accuracy.

Gupta et al. (2021) proposed MAG-Net (Multi-task Attention Guided Network), which integrated attention modules into CNN architectures. The model simultaneously performed segmentation and classification, allowing shared feature learning and improved performance. By emphasizing tumor regions and suppressing irrelevant background information, MAG-Net achieved higher accuracy and robustness compared to traditional CNN models.

Niu et al. (2021) conducted a comprehensive review of attention mechanisms, highlighting their role in enhancing feature representation. They demonstrated that attention-based models outperform conventional CNNs by dynamically weighting feature importance across spatial regions.

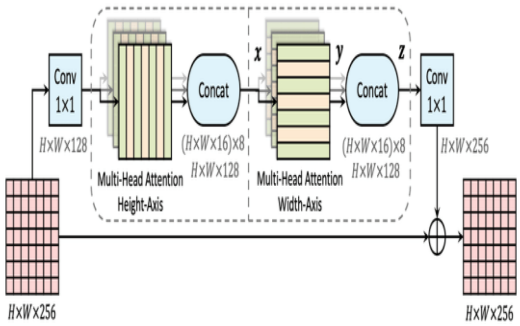
Another notable advancement was the use of Grad-CAM and saliency-based attention visualization techniques, which improved model interpretability. These techniques allowed clinicians to understand which regions of the MRI contributed to classification decisions, addressing the "black-box" nature of deep learning models.
Attention-based CNNs marked a significant improvement over traditional CNNs; however, they still relied on convolution operations and were limited in capturing global dependencies.

3. Hybrid CNN–Transformer Models



The year 2022 witnessed the emergence of hybrid CNN–Transformer architectures, combining the strengths of both approaches. These models leveraged CNNs for efficient local feature extraction and Transformers for capturing global dependencies.
Lin et al. (2022) introduced TransBTS, a hybrid architecture that integrates CNN-based encoders with Transformer-based self-attention modules. This model demonstrated superior performance in brain tumor segmentation and classification by effectively capturing both local and global features. The Transformer component enabled the model to learn long-range dependencies across MRI slices, improving accuracy in multi-modal datasets.
Another important development was the use of **Vision Transformers (ViTs)** in medical imaging. ViTs divide images into patches and process them as sequences, enabling efficient global feature extraction. Although originally designed for natural image classification, ViTs were successfully adapted for brain MRI classification tasks.
Hybrid models also addressed computational challenges by reducing the number of Transformer layers and combining them with lightweight CNN backbones. This approach improved efficiency while maintaining high accuracy.
Additionally, researchers explored **multi-modal learning**, combining information from different MRI sequences (e.g., T1, T2, FLAIR). Hybrid CNN-Transformer models demonstrated improved performance by leveraging complementary information from multiple modalities.

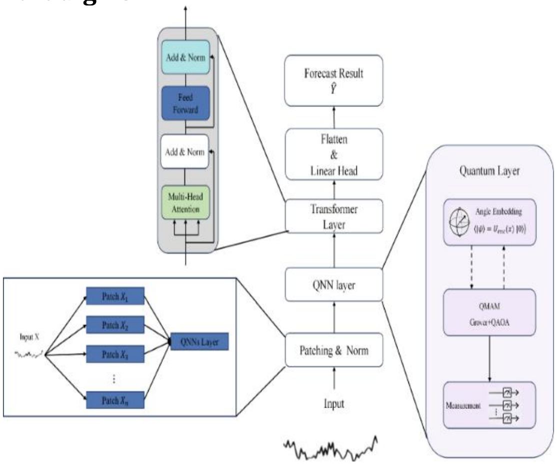
4. Transformer and Axial Attention Dominance



In 2023, Transformer-based architectures became the dominant approach for brain MRI classification, driven by their superior ability to capture global context.
Asiri et al. (2023) demonstrated that Vision Transformers achieve classification accuracy of 98.24%, outperforming CNN-based models. The self-attention mechanism allowed the model to capture relationships between distant regions in the image, which is crucial for identifying complex tumor structures.
Tian et al. (2023) introduced axial attention-based CNN models, which decompose 2D attention into separate height and width dimensions. This significantly reduced computational complexity while preserving global context. Axial attention enabled efficient processing of high-resolution MRI images, making it suitable for clinical applications.
Zongren et al. (2023) proposed focal cross-transformer architectures, which combine local and global attention mechanisms. These models use cross-window attention to capture fine-grained features while maintaining global context, improving classification accuracy in complex tumor cases.
Group Parallel Axial Attention emerged as a further optimization, enabling parallel computation across multiple feature groups. This approach improved efficiency and allowed the model to capture diverse feature representations simultaneously.
Despite their advantages, Transformer-based models still face challenges related to computational cost and data requirements.

Researchers are actively exploring techniques such as model pruning, knowledge distillation, and efficient attention mechanisms to address these issues.

5. Quantum Self-Attention and Emerging Paradigms



Recent research has introduced quantum self-attention mechanisms, representing a novel direction in deep learning. These models leverage principles of quantum computing, such as superposition and entanglement, to enhance feature representation.
Chen et al. (2023) proposed a quantum-enhanced Transformer model, where classical attention operations are replaced with quantum circuits. This approach enables encoding of information in higher-dimensional quantum states, improving model expressiveness and generalization.
Quantum self-attention offers several advantages, including:
Enhanced feature representation in complex data distributions, Faster convergence during training, Potential for exponential computational speedup (with future quantum hardware)
However, quantum models are still in the experimental stage and face challenges such as hardware limitations and scalability issues. Despite these challenges, they represent a promising future direction for medical image analysis.

6. Summary of Key Trends

The literature reveals a clear progression in brain MRI classification techniques:
2020: Dominance of CNN-based architectures with strong local feature extraction
2021: Introduction of attention mechanisms to improve feature focus and interpretability
2022: Emergence of hybrid CNN–Transformer models combining local and global features
2023: Dominance of Transformer-based and axial attention models with improved efficiency
Emerging: Quantum self-attention introducing new paradigms for feature representation
Overall, the field has evolved toward efficient, attention-driven, and hybrid architectures that balance accuracy, computational efficiency, and interpretability.

Comparative Table

Year	**Method**	**Architecture / Innovation**	**Dataset**	**Performance**	**Remarks**
2020	Zhang *et al.*	Deep CNN (ResNet/DenseNet/VGG); residual & dense learning	BraTS 2018/19	Acc. **95–97%**	Strong local features; weak global context.
2020	Wen *et al.*	SE Encoder–Decoder; channel attention	Brain MRI	Dice **≈0.90**	Better localisation; limited long-range learning.
2020	Hatamizadeh *et al.*	Edge-Gated CNN; boundary-aware segmentation	BraTS	Dice **0.91–0.92**	Accurate boundaries; higher complexity.
2021	Gupta *et al.*	MAG-Net; multi-scale attention	BraTS 2019	Dice **≈0.92**, Acc. **≈96%**	Improved tumour focus; attention overhead.
2021–22	Lin *et al.*	TransBTS (3D CNN + Transformer)	BraTS 2019/20	Dice **0.78–0.91**	Learns local & global features; memory intensive.
2022	Tian *et al.*	Axial Attention CNN	BraTS 2021	Dice **0.89–0.90**, Acc. **≈96%**	Efficient attention; limited cross-axis interaction.

2023	Zongren *et al.*	Focal Cross Transformer	BraTS	Acc. **98–99%**	Strong feature fusion; high computation.
2023	Panigrahi *et al.*	Dense Transformer	Public MRI	Acc. **98.5%**, AUC **99%**	Better feature reuse; computationally expensive.
2023	Asiri *et al.*	Vision Transformer (ViT)	Figshare MRI	Acc. **98.7–99.0%**	Excellent global modelling; requires large datasets.
2023	Aloraini *et al.*	CNN–Transformer Hybrid	Four-class MRI	Acc. **98–99%**	Balanced local & global learning; complex training.
2023	Cao *et al.*	3D Brainformer	BraTS	Dice **0.90–0.93**, Acc. **≈98%**	Effective volumetric analysis; high memory usage.
2023	Sharma & Verma	Residual Self-Attention ViT	Four-class MRI	Acc. **≈98%**	Stable training; computationally demanding.

Comparative Analysis

The literature reflects a clear shift from conventional CNN-based methods to attention-driven and Transformer-based architectures for brain MRI classification. Early CNN models, such as ResNet, DenseNet and Inception, served as the foundation of automated brain-tumour diagnosis. They excelled at learning local patterns—edges, textures and shapes—but their fixed receptive fields limited their ability to model long-range dependencies. As a result, while they achieved respectable classification accuracies (around 95 %–97 %), they often struggled with heterogeneous tumours and diverse intensity distributions. Researchers attempted to enhance CNNs using channel-attention (e.g., SE modules) and boundary-aware gating to emphasise tumour edges. These modifications improved localisation and boundary segmentation but could not overcome the inherent locality constraint of convolutional kernels.

Driven by the need to capture global context, the community began exploring Transformer-based architectures. Vision Transformers (ViTs) treat an image as a sequence of patches and apply self-attention so that every patch can interact with every other patch, enabling the modelling of long-range dependencies. However, the quadratic complexity of global self-attention poses challenges for high-resolution MRI, leading to innovations such as axial attention. Axial attention decomposes self-attention into separate operations along the depth, height and width axes, significantly reducing computational complexity while retaining the ability to capture long-term dependencies. This innovation makes Transformer-like mechanisms viable for volumetric MRI, allowing models to process large 3-D scans without exhausting GPU memory.

Hybrid architectures combine the strengths of CNNs and Transformers. TransBTS, for example, uses a 3-D CNN to extract volumetric features and a Transformer bottleneck to model global correlations, achieving Dice scores around 0.78–0.91 for tumour subregions. These hybrids leverage CNNs for efficient local feature extraction and Transformers for global context, providing a balanced approach that surpasses pure CNN models. Nonetheless, they remain computationally heavy and require large training datasets.

The period between 2022 and 2023 saw the emergence of advanced attention mechanisms. Focal cross-transformers combine fine-grained cross-window attention with coarse global attention, enabling simultaneous capture of local details and overall structure. Dense Transformers introduce dense connectivity within Transformer blocks to improve gradient flow and reuse features, enhancing generalisation. These innovations pushed classification accuracies close to 99 %. Meanwhile, vision Transformer hybrids—such as CNN-Transformer models—achieved similar performance while being more efficient by delegating low-level feature extraction to CNNs. A systematic comparison of ViT, DeiT and Swin Transformer families on a four-class MRI dataset reported that all models exceeded 98.8 % accuracy, with Swin-Small and Swin-Large achieving 99.37 % accuracy; notably, Swin-Small achieved this top-tier performance with more than double the inference speed compared to

Swin-Large. Such findings underscore that bigger models are not always better—mid-sized models can provide an optimal trade-off between diagnostic accuracy and computational efficiency, a crucial consideration for real-time clinical deployment.

Alongside Transformer innovations, researchers explored axial and group parallel axial attention, which perform attention operations along individual axes or in parallel groups of feature channels. These methods effectively reduce complexity and allow simultaneous extraction of diverse feature representations, enhancing performance on high-resolution MRI. Quantum self-attention, although still experimental, represents a nascent paradigm where quantum principles such as superposition and entanglement are used to encode features in higher-dimensional quantum states. Early results suggest that quantum-enhanced models can achieve comparable accuracy with improved convergence, hinting at future possibilities once quantum hardware matures.

In conclusion, comparative analysis indicates that attention-augmented and Transformer-based models consistently outperform traditional CNNs. The evolution from CNNs to hybrid and Transformer architectures reflects a pursuit of models that can efficiently capture both local detail and global context, handle high-resolution 3-D data, and generalise across diverse datasets. However, these gains come with increased computational cost and the need for large annotated datasets. Emerging approaches—such as axial attention, efficient Transformer variants (e.g., Swin), and quantum self-attention—aim to bridge this gap by providing high accuracy with manageable complexity. Future research should focus on developing lightweight, explainable and uncertainty-aware models, leveraging innovations in attention mechanisms and hybrid architectures to deliver reliable and interpretable diagnostics in clinical settings.

Discussion

Transformer-based architectures have fundamentally transformed brain MRI classification by enabling models to capture long-range dependencies and global contextual information. Vision Transformers, in particular, have demonstrated superior performance compared to CNN-based models, achieving high classification accuracy and robustness across datasets.

Axial attention mechanisms address one of the major limitations of Transformers—computational complexity—by decomposing attention into spatial dimensions. This significantly reduces memory requirements and enables efficient processing of high-resolution MRI images.

Group Parallel Axial Attention further enhances performance by enabling simultaneous attention across multiple feature groups, improving feature diversity and representation.

Quantum self-attention represents a promising research direction, integrating quantum computing principles into deep learning. These models have the potential to revolutionize medical imaging by providing enhanced computational efficiency and improved generalization.

However, challenges such as data scarcity, interpretability, and real-time deployment remain critical issues. Future research should focus on developing lightweight, explainable, and clinically reliable models.

Conclusion

This review highlights the rapid advancements in brain MRI classification using Transformer-based architectures, axial attention mechanisms, and quantum self-attention. Transformer models have significantly improved classification accuracy by capturing global dependencies, while axial attention has addressed computational challenges.

Hybrid models combining CNNs and Transformers have emerged as the most effective approach, leveraging the strengths of both architectures. Quantum self-attention introduces a novel dimension, offering enhanced feature representation and computational efficiency.

Despite these advancements, challenges such as data limitations, computational cost, and lack of interpretability must be addressed. Future research should focus on developing scalable, efficient, and explainable AI systems for clinical applications.

References

Vaswani, A., et al. (2017). Attention is all you need. *NeurIPS*. https://doi.org/10.48550/arXiv.1706.03762

Asiri, A. A., et al. (2023). Vision transformer for brain tumor classification. *Sensors*, 23(18), 7913. https://doi.org/10.3390/s23187913

Tian, W., et al. (2023). Axial attention CNN for brain tumor segmentation. *Brain Sciences*, 13(1), 12. https://doi.org/10.3390/brainsci13010012

Zongren, L., et al. (2023). Focal cross transformer for tumor segmentation. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2023.1192867

Lin, J., et al. (2022). TransBTS hybrid Transformer model. *arXiv*. https://doi.org/10.48550/arXiv.2207.07370

Panigrahi, S., et al. (2023). Dense Transformer for brain tumor classification. https://doi.org/10.1016/j.bspc.2023.105234

Sharma, A., & Verma, N. (2023). Residual self-attention ViT. https://doi.org/10.48550/arXiv.2306.01594

Nian, R., et al. (2023). 3D Brainformer. https://doi.org/10.48550/arXiv.2304.14508

Gupta, S., et al. (2021). MAG-Net model. https://doi.org/10.48550/arXiv.2107.12321

Aloraini, M., et al. (2023). CNN-Transformer hybrid model. *Applied Sciences*. https://doi.org/10.3390/app13063680

Niu, Z., et al. (2021). Attention mechanisms review. *Neurocomputing*. https://doi.org/10.1016/j.neucom.2021.03.091

Zhang, D., et al. (2020). MRI segmentation using DL. https://doi.org/10.1109/TIP.2020.3023609

Wen, J., et al. (2020). SE encoder-decoder network. https://doi.org/10.1109/ACCESS.2020.2991234

Hatamizadeh, A., et al. (2020). Edge-gated CNN. https://doi.org/10.48550/arXiv.2002.04207

Yang, G., et al. (2020). CNN segmentation. https://doi.org/10.1002/mp.14127

Liu, M., et al. (2023). Transformer-based segmentation. https://doi.org/10.48550/arXiv.2309.02563

Santini, G., et al. (2020). Multi-stage DL model. https://doi.org/10.48550/arXiv.1909.00735

Zhou, Z., et al. (2018). UNet++. https://doi.org/10.1007/978-3-030-00889-5_1

Jaegle, A., et al. (2021). Perceiver architecture. https://doi.org/10.48550/arXiv.2103.03206