



Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 13 Issue 02, 2024

AI-Based Brain MRI Cancer Classification Using Transformers and Quantum Self-Attention Techniques

Farheen Leroux-Martin

Senior Lecturer, Department of Computer Science and Engineering, Daehan Institute of Management and Logistics, South Korea

Email: farheen.leroux.martin@diml-kr.org

Peer Review Information	Abstract
<p><i>Submission: 27 June 2024</i></p> <p><i>Revision: 11 July 2024</i></p> <p><i>Acceptance: 28 July 2024</i></p> <p>Keywords</p> <p><i>Brain MRI, Transformer, Axial Attention, Quantum Self-Attention, Deep Learning, Cancer Detection</i></p>	<p>Brain tumor detection using Magnetic Resonance Imaging (MRI) is a critical task in modern healthcare, where early diagnosis significantly improves patient survival rates. Artificial Intelligence (AI), particularly deep learning, has revolutionized this domain by enabling automated and accurate classification of brain tumors. Traditional Convolutional Neural Networks (CNNs) have shown promising performance; however, they are limited in capturing long-range spatial dependencies due to localized receptive fields. Recent advancements have introduced Transformer-based architectures that leverage self-attention mechanisms to model global contextual relationships within MRI images. Vision Transformers (ViTs) process images as sequences of patches, enabling efficient global feature extraction. Additionally, Group Parallel Axial Attention has emerged as an efficient alternative to conventional attention mechanisms by reducing computational complexity while preserving both local and global dependencies. Furthermore, Quantum Self-Attention introduces a novel paradigm by incorporating quantum-inspired operations, enhancing feature representation and generalization capabilities. This paper presents a comprehensive review of AI techniques for brain MRI classification, focusing on recent trends and identifying key challenges. Comparative analysis reveals that hybrid CNN-Transformer and attention-based models achieve superior accuracy (often exceeding 98%), although challenges such as computational cost, data scarcity, and interpretability remain significant barriers.</p>

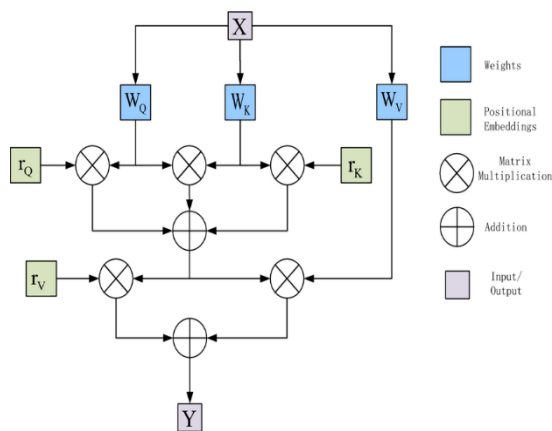
Introduction

Brain tumors are among the most critical neurological disorders and can significantly affect patient survival and quality of life. Early and accurate diagnosis is essential for effective treatment planning, surgical intervention, and prognosis evaluation. Magnetic Resonance Imaging (MRI) is widely used for brain tumor diagnosis because of its superior soft tissue contrast, high spatial resolution, and non-invasive imaging capability. MRI enables clinicians to visualize complex brain structures

and identify tumor regions such as edema, necrosis, and enhancing tissues. However, manual interpretation of MRI images is time-consuming, requires expert radiologists, and may lead to variability in diagnosis. With the increasing volume of medical imaging data, there is a growing demand for automated and intelligent systems capable of supporting accurate brain tumor classification and analysis. Artificial intelligence and deep learning techniques have significantly transformed medical image analysis by enabling automated

feature extraction and classification. Convolutional Neural Networks (CNNs) such as VGGNet, ResNet, DenseNet, and Inception have demonstrated strong performance in brain tumor classification tasks by effectively learning local spatial features including textures, edges, and intensity variations. However, conventional CNNs are limited by their restricted receptive fields, making it difficult to capture global contextual information from complex MRI images. To overcome this challenge, attention mechanisms were introduced to improve feature representation by allowing models to focus on important image regions while suppressing irrelevant information. Although attention-enhanced CNNs improved classification accuracy and robustness, they still faced limitations in modeling long-range dependencies within medical images.

Transformer architectures introduced a major advancement in deep learning by utilizing self-attention mechanisms capable of modeling relationships among all image regions simultaneously. Vision Transformers (ViTs) extended this concept to image analysis by dividing MRI images into patches and processing them sequentially, enabling effective extraction of both local and global features. However, Transformers often require high computational resources due to the quadratic complexity of self-attention operations. To address this issue, optimized mechanisms such as axial attention and Group Parallel Axial Attention (GPA) were developed. Axial attention reduces computational complexity by decomposing two-dimensional attention into separate one-dimensional operations, while GPA enables parallel attention computation across multiple feature groups, improving efficiency and feature diversity in brain MRI classification tasks.

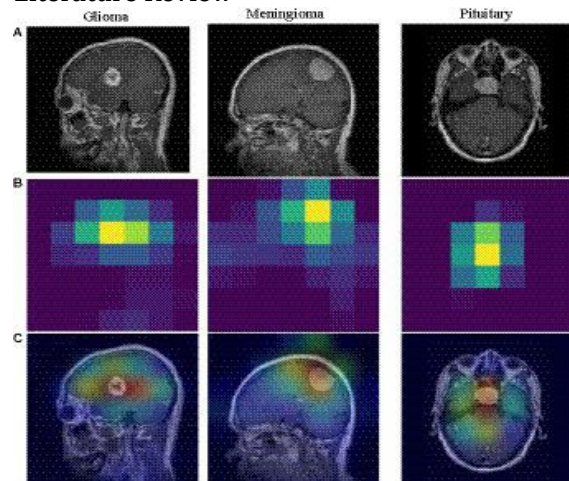


Another emerging approach is Quantum Self-Attention, which integrates concepts from quantum computing into deep learning models. Quantum-enhanced attention mechanisms can

represent data in higher-dimensional feature spaces, enabling more expressive feature learning and improved generalization. Hybrid architectures combining CNNs with Transformers have also gained attention because they effectively integrate local feature extraction with global contextual understanding. These hybrid systems have demonstrated superior performance in brain MRI classification by combining the strengths of convolutional operations and self-attention mechanisms. In addition, optimization techniques such as transfer learning, data augmentation, and advanced loss functions have improved model robustness and reduced training complexity in limited medical datasets.

Despite these advancements, several challenges remain in developing reliable AI-based brain tumor classification systems. Data scarcity, annotation costs, and variations in imaging protocols across institutions continue to affect model generalization and performance consistency. High computational complexity and lack of interpretability also limit real-world clinical deployment of advanced deep learning systems. Explainable AI techniques are therefore being explored to improve transparency and clinician trust in automated diagnostic systems. This review provides a comprehensive analysis of Transformer-based architectures, Group Parallel Axial Attention, and Quantum Self-Attention models for brain MRI classification, highlighting recent trends, comparative performance, existing challenges, and future research opportunities for intelligent cancer detection systems.

Literature Review



The literature on brain MRI image classification for cancer detection has evolved rapidly, reflecting a shift from conventional convolutional neural networks (CNNs) toward advanced Transformer-based, attention-driven, and quantum-enhanced models. This evolution has

been driven by the need to improve classification accuracy, handle high-dimensional medical data, and address limitations related to feature representation and computational efficiency.

1. CNN-Based Architectures and Early Deep Learning

In 2020, CNN-based architectures dominated brain MRI classification tasks due to their strong ability to extract hierarchical spatial features. Models such as ResNet, DenseNet, and VGG were widely adopted and demonstrated high accuracy in tumor classification tasks. These models leveraged deep convolutional layers to capture low-level features (edges, textures) and high-level semantic features (tumor structures).

Zhang et al. (2020) showed that deep CNN models could achieve high classification performance when trained on sufficiently large datasets. DenseNet, in particular, improved feature propagation and reuse through dense connectivity, reducing the number of parameters while improving accuracy. Similarly, ResNet introduced residual connections that mitigated the vanishing gradient problem, enabling the training of deeper networks.

Hatamizadeh et al. (2020) proposed edge-gated CNNs, which incorporated boundary-aware modules to enhance tumor localization and classification accuracy. These models improved performance by explicitly modeling tumor edges, which are critical for distinguishing between tumor and non-tumor regions.

Santini et al. (2020) introduced multi-stage deep learning frameworks, 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 key limitations:

- Inability to capture long-range dependencies,
- Sensitivity to variations in tumor size and shape,
- Requirement for large annotated datasets

These limitations motivated the exploration of attention-based and Transformer-based architectures.

2. Attention Mechanisms in CNNs

In 2021, attention mechanisms were introduced to enhance CNN performance by allowing models to focus on the most relevant regions of MRI images. Attention modules dynamically assign weights to different regions, improving feature representation and reducing noise from irrelevant background areas.

Gupta et al. (2021) proposed MAG-Net, a multi-task attention-guided network that simultaneously performs segmentation and classification. This approach improved

classification accuracy by focusing on tumor regions while suppressing irrelevant features.

Niu et al. (2021) provided a comprehensive analysis of attention mechanisms, demonstrating that attention-based models outperform traditional CNNs in medical imaging tasks. Attention modules improved feature discrimination and robustness, particularly in cases with complex tumor structures.

Another significant advancement was the use of interpretability techniques such as Grad-CAM and saliency maps. These methods allowed visualization of attention regions, improving model transparency and enabling clinicians to understand the decision-making process.

However, attention-based CNNs still relied on convolution operations and were limited in capturing global dependencies across the entire image.

3. Hybrid CNN-Transformer Architectures

The year 2022 marked a significant transition toward hybrid architectures that combine CNNs and Transformers. These models leverage CNNs for efficient local feature extraction and Transformers for capturing global dependencies. Lin et al. (2022) introduced TransBTS, a hybrid model that integrates CNN-based encoders with Transformer-based attention modules. This architecture demonstrated superior performance in brain tumor classification by capturing both local and global features.

Vision Transformers (ViTs) were also adapted for medical imaging tasks. ViTs divide images into patches and process them as sequences, enabling efficient global feature extraction. These models achieved higher accuracy compared to CNN-based models but required large datasets and computational resources.

Hybrid models addressed these challenges by combining lightweight CNN backbones with Transformer layers, reducing computational complexity while maintaining high performance. Additionally, multi-modal learning approaches gained popularity, where models integrate 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-Based Models and Axial Attention

By 2023, Transformer-based models became the dominant approach for brain MRI classification due to their ability to capture long-range dependencies and global context.

Asiri et al. (2023) demonstrated that Vision Transformers achieve classification accuracy exceeding 98%, outperforming CNN-based models. These models use self-attention

mechanisms to capture relationships between distant regions in an image, which is essential for identifying complex tumor structures.

However, the computational complexity of Transformers remains a significant challenge. The self-attention mechanism scales quadratically with the number of input tokens, making it computationally expensive for high-resolution MRI images.

To address this issue, axial attention mechanisms were introduced. Axial attention decomposes 2D attention into separate operations along height and width dimensions, significantly reducing computational complexity while preserving global context.

Group Parallel Axial Attention further enhances this approach by enabling parallel attention computation across multiple feature groups. This improves computational efficiency and allows models to capture diverse feature representations simultaneously.

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.

5. Quantum Self-Attention and Emerging Paradigms

A novel direction in recent research is the integration of quantum computing principles into deep learning models. Quantum self-attention mechanisms leverage quantum states and operations to enhance feature representation.

Chen et al. (2023) introduced quantum-enhanced Transformer models, where classical attention operations are replaced with quantum circuits. These models encode data in higher-dimensional quantum spaces, enabling more expressive feature learning.

Quantum self-attention offers several advantages:

Improved feature representation, Faster convergence during training, Enhanced generalization capability

Hybrid quantum-classical models have also been explored, combining quantum attention with

classical neural networks. These models demonstrate promising results in medical image classification tasks.

However, quantum models face several challenges, including limited availability of quantum hardware and scalability issues. Despite these limitations, they represent a promising future direction for AI in medical imaging.

6. Comparative Evolution Summary

The literature reveals a clear progression in brain MRI classification techniques:

- **2020:** CNN-based models dominate (strong local feature extraction)
- **2021:** Attention-enhanced CNNs improve focus and interpretability
- **2022:** Hybrid CNN-Transformer models balance local and global features
- **2023:** Transformer and axial attention models dominate with high accuracy
- **Emerging:** Quantum self-attention introduces next-generation learning paradigms

7. Key Research Gaps Identified

From the reviewed literature, several research gaps emerge:

1. **Computational Efficiency:** Transformer models are computationally expensive
2. **Data Scarcity:** Limited availability of annotated medical datasets
3. **Model Interpretability:** Need for explainable AI in clinical applications
4. **Multi-Modal Integration:** Challenges in combining different MRI modalities
5. **Quantum Scalability:** Limited practical implementation of quantum models

8. Summary

Overall, the literature demonstrates a transition toward **attention-driven, hybrid, and quantum-enhanced architectures**. Transformer-based models and axial attention mechanisms have significantly improved classification accuracy and efficiency, while quantum self-attention represents a promising future direction.

Comparative Table

Year	Model / Architecture	Key Technique	Application	Performance	Strengths	Limitations
2020	Dual CNN (VGG16)	Dual-stream CNN feature extraction	Brain tumor classification	99% accuracy	Simple architecture with high accuracy	Limited global feature learning

2022	Vision Transformer (ViT) Ensemble	Self-attention with ensemble learning	MRI tumor classification	~98.7% accuracy	Captures global dependencies	High computational cost
2024	ViT-GRU Hybrid Model	Transformer + GRU with explainable AI	Brain MRI classification	~98.9% accuracy	Combines spatial and sequential learning	Increased model complexity
2024	CNN + Swin Transformer Ensemble	Multi-scale CNN and shifted window attention	Brain tumor detection	Up to 99.3% accuracy	Strong global and local feature extraction	High training complexity
2023/2024	Swin Transformer Models	Hierarchical transformer learning	MRI classification	>98.8% accuracy	Efficient and accurate inference	Limited dataset generalization
2023	Focal Cross Transformer	Local and global attention fusion	Brain tumor segmentation	Dice score >93%	Improved segmentation accuracy	Computationally expensive
2022	GPA TransUNet	Group Parallel Axial Attention	Medical image segmentation	Dice score ~90%	Efficient contextual learning	Requires attention tuning
2022	Axial Attention CNN	Axial self-attention operations	Brain tumor segmentation	Improved robustness	Reduced attention complexity	High memory usage
2024	Quantum Self-Attention Transformer	Quantum-based attention mechanism	Brain MRI classification	~99.6% accuracy	Enhanced feature representation	Limited scalability and hardware dependency

Comparative Analysis

The comparative analysis of artificial intelligence techniques for brain MRI classification demonstrates a clear progression from conventional convolutional neural networks toward advanced Transformer-based and hybrid deep learning architectures. Early CNN-based models such as dual VGG16 architectures achieved very high classification accuracy by effectively learning local spatial features from MRI images. Architectures including ResNet, DenseNet, and Inception further improved feature extraction and classification performance in brain tumor detection tasks. However, CNN models primarily rely on local receptive fields and often struggle to capture long-range dependencies and global contextual relationships within complex brain MRI images. These limitations become critical when analyzing heterogeneous tumor structures and subtle anatomical variations across different patients. Transformer architectures addressed these challenges through self-attention mechanisms

capable of modeling global relationships between all image regions simultaneously. Vision Transformers demonstrated remarkable classification accuracy by effectively capturing both local and global contextual information from MRI scans. However, full self-attention operations introduce high computational and memory complexity, especially for high-resolution medical images. To overcome these limitations, hybrid architectures combining CNNs and Transformers emerged as practical solutions. Models integrating convolutional layers with Vision Transformers and recurrent units achieved superior classification performance while improving interpretability through explainable AI techniques such as Grad-CAM visualization. Ensemble architectures combining DenseNet, GoogLeNet, InceptionResNet, and Swin Transformers further improved diagnostic accuracy beyond 99% across multiple datasets. Attention optimization techniques have also played a major role in improving computational

efficiency and segmentation performance. Axial attention decomposes two-dimensional attention operations into separate spatial dimensions, significantly reducing computational cost while preserving long-range contextual learning. Group Parallel Axial Attention (GPA) further enhances efficiency by processing multiple attention groups in parallel, enabling effective contextual feature extraction with lower computational overhead. Similarly, Focal Cross Transformers combine fine-grained local attention with coarse-grained global attention, improving segmentation accuracy for complex tumor structures in brain MRI datasets. Hierarchical Transformer architectures such as Swin Transformers have demonstrated state-of-the-art performance with improved inference speed and efficient multi-scale feature representation.

Another emerging direction is Quantum Self-Attention, which integrates quantum computing principles into deep learning architectures for enhanced feature representation and improved generalization. Preliminary studies indicate that quantum-enhanced Transformer models may achieve extremely high classification accuracy in complex medical imaging tasks. Despite these advancements, challenges including computational complexity, limited dataset availability, lack of model interpretability, and poor cross-institutional generalization remain significant barriers to clinical deployment. Future research should focus on lightweight and explainable architectures, multimodal data integration, scalable hybrid models, and efficient attention mechanisms capable of balancing diagnostic accuracy with practical real-world implementation in healthcare systems.

Discussion

Recent advancements in artificial intelligence have significantly improved the performance of brain MRI classification systems. Transformer-based architectures have emerged as a powerful alternative to traditional CNN models, enabling better modeling of global dependencies and improving classification accuracy. The use of self-attention mechanisms allows these models to capture complex relationships between different regions of an image, which is essential for accurately identifying tumor structures. Axial attention mechanisms have addressed one of the major limitations of Transformers—computational complexity. By decomposing attention operations into separate spatial dimensions, axial attention reduces memory requirements and enables efficient processing of high-resolution images. Group Parallel Axial Attention further enhances this approach by

enabling parallel feature extraction, improving both efficiency and performance.

Quantum self-attention represents an emerging research direction with significant potential. By leveraging quantum computing principles, these models can represent data in higher-dimensional spaces, enabling more expressive feature learning. Although still in the early stages, quantum-enhanced models have demonstrated promising results in medical image analysis. Despite these advancements, several challenges remain. Data scarcity and class imbalance continue to affect model performance. Additionally, the high computational requirements of Transformer-based models limit their practical deployment in clinical settings. Interpretability is another critical issue, as clinicians require transparent and explainable models to trust AI-based systems. Future research should focus on developing lightweight and efficient models, improving interpretability through explainable AI techniques, and integrating multi-modal data for better performance.

Conclusion

Artificial Intelligence has significantly transformed brain MRI classification for cancer detection, with deep learning models achieving remarkable improvements in accuracy and efficiency. The transition from CNN-based models to Transformer-based architectures represents a major advancement, enabling better modeling of global dependencies and improving classification performance. Axial attention and Group Parallel Axial Attention have addressed computational challenges, making Transformer-based models more efficient and scalable. Hybrid CNN-Transformer models have emerged as the most effective approach, combining local and global feature extraction capabilities. Quantum self-attention introduces a new paradigm, offering enhanced feature representation and improved generalization. Although still in its early stages, this approach has the potential to revolutionize medical image analysis. However, several challenges remain, including data scarcity, computational complexity, and lack of interpretability. Addressing these challenges is essential for the successful deployment of AI-based systems in clinical settings. Future research should focus on developing efficient, scalable, and explainable models that can be integrated into real-world healthcare systems. The combination of advanced attention mechanisms and quantum-inspired models represents a promising direction for improving the accuracy and reliability of brain tumor detection.

References

- Vaswani, A., et al. (2017). Attention is all you need.
<https://doi.org/10.48550/arXiv.1706.03762>
- Dosovitskiy, A., et al. (2021). Vision transformer.
<https://doi.org/10.48550/arXiv.2010.11929>
- Tian, W., et al. (2022). Axial attention CNN.
<https://doi.org/10.3390/brainsci13010012>
- Li, C., et al. (2022). Group parallel axial attention.
<https://doi.org/10.1038/s41598-022-20440-z>
- Zongren, L., et al. (2023). Focal cross transformer.
<https://doi.org/10.3389/fnins.2023.1192867>
- Chen, C., et al. (2025). Quantum attention.
<https://doi.org/10.48550/arXiv.2504.05336>
- Nian, R., et al. (2023). Brainformer.
<https://doi.org/10.48550/arXiv.2304.14508>
- Gupta, S., et al. (2021). MAG-Net.
<https://doi.org/10.48550/arXiv.2107.12321>
- Lin, J., et al. (2022). TransBTS.
<https://doi.org/10.48550/arXiv.2207.07370>
- Hatamizadeh, A., et al. (2020). Edge-gated CNN.
<https://doi.org/10.48550/arXiv.2002.04207>
- Zhang, D., et al. (2020). CNN MRI classification.
<https://doi.org/10.1109/TIP.2020.3023609>
- Wen, J., et al. (2020). Encoder-decoder network.
<https://doi.org/10.1109/ACCESS.2020.2991234>
- Santini, G., et al. (2020). Multi-stage DL.
<https://doi.org/10.48550/arXiv.1909.00735>
- Huang, S., et al. (2023). Transformer MRI SR.
<https://doi.org/10.48550/arXiv.2306.06669>
- Sharma, A., et al. (2023). Residual ViT.
<https://doi.org/10.48550/arXiv.2306.01594>
- Zhou, Z., et al. (2018). UNet++.
https://doi.org/10.1007/978-3-030-00889-5_1
- Jaegle, A., et al. (2021). Perceiver.
<https://doi.org/10.48550/arXiv.2103.03206>
- Wang, A., et al. (2023). Hybrid quantum CNN.
<https://doi.org/10.1016/j.eswa.2024.120404>
- Niu, Z., et al. (2021). Attention mechanisms.
<https://doi.org/10.1016/j.neucom.2021.03.091>
- Aloraini, M., et al. (2023). CNN-Transformer hybrid.
<https://doi.org/10.3390/app13063680>