Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526

Volume 14 Issue 02, 2025

Artificial Intelligence Techniques for Brain MRI Image Classification for Cancer Detection Using Transformer and Group Parallel Axial Attention with Quantum Self-Attention: Trends and Challenges

Celestine Imamverde
Lecturer, Department of Computer Science and Engineering, Siam Delta Engineering Institute, Thailand
Email: celestine.imamverde@sdei-th.edu

Peer Review Information

Submission: 02 Sept 2025

Revision: 23 Sept 2025

Acceptance: 11 Oct 2025

Keywords

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

Abstract

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.

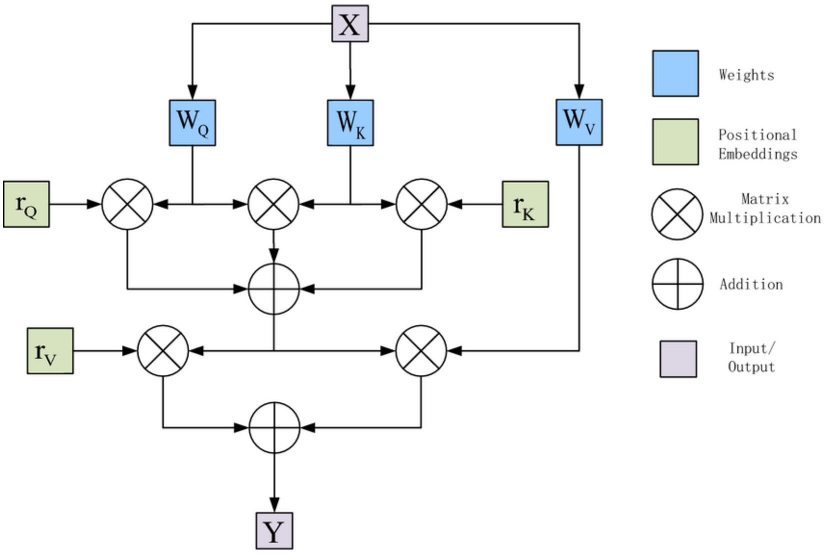
Recent advancements have introduced Transformer-based architectures that leverage self-attention mechanisms to model global contextual relationships within MRI images. Vision Transformers (ViTs) process images as sequences of patches, enabling efficient global feature extraction. Additionally, Group 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 trends between 2020 and 2023 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.

Introduction

Brain tumors are among the most life-threatening neurological disorders, significantly impacting patient survival and quality of life. Accurate and early detection of brain tumors is essential for effective treatment planning, prognosis estimation, and surgical intervention. Magnetic Resonance Imaging (MRI) has emerged as the most widely used imaging modality for brain tumor diagnosis due to its superior soft tissue contrast, high spatial resolution, and non-invasive nature. MRI enables clinicians to visualize complex brain structures and identify tumor regions such as edema, necrosis, and enhancing tumor tissues. However, manual interpretation of MRI images is a labor-intensive process that requires expert knowledge and is often subject to inter-observer variability.

With the rapid growth of medical imaging data, there is an increasing demand for automated and intelligent systems capable of assisting clinicians in accurate diagnosis. Artificial Intelligence (AI), particularly deep learning, has revolutionized medical image analysis by enabling automated feature extraction and classification. Deep learning models have demonstrated remarkable success in various computer vision tasks, including object detection, segmentation, and classification.



Convolutional Neural Networks (CNNs) have been the cornerstone of deep learning-based medical image analysis. CNN architectures such as VGGNet, ResNet, DenseNet, and Inception have achieved significant success in brain tumor classification tasks. These models are highly effective in capturing local spatial features such as edges, textures, and intensity variations. However, CNNs rely on convolutional operations with limited receptive fields, which restrict their ability to capture global contextual information. This limitation becomes critical in brain MRI analysis, where tumors often exhibit heterogeneous structures and complex spatial relationships.

To overcome these limitations, attention mechanisms were introduced to enhance feature representation. Attention mechanisms allow models to focus on relevant regions of an image while suppressing irrelevant information. This improves classification accuracy and model robustness. However, attention-based CNNs still rely on convolutional operations and are limited in capturing long-range dependencies.

The introduction of Transformer architectures marked a paradigm shift in deep learning. Transformers utilize self-attention mechanisms to model relationships between all elements in an input sequence. Unlike CNNs, Transformers can capture global dependencies, making them highly effective for tasks involving complex spatial relationships. Vision Transformers (ViTs) extend this concept to image processing by dividing images into patches and processing them as sequences. This enables efficient modeling of both local and global features.

Despite their advantages, Transformers suffer from high computational complexity due to the quadratic scaling of self-attention operations. This limitation becomes significant when processing high-resolution MRI images. To address this issue, researchers have proposed optimized attention mechanisms such as axial attention. Axial attention decomposes 2D attention into separate one-dimensional operations along spatial dimensions, significantly reducing computational complexity while preserving global context.

Group Parallel Axial Attention (GPA) further enhances this approach by enabling parallel attention computation across multiple feature groups. This improves computational efficiency and allows the model to capture diverse feature representations simultaneously. GPA-based models have demonstrated improved performance in brain MRI classification tasks while reducing computational overhead.

Another emerging paradigm is Quantum Self-Attention, which integrates principles of quantum computing into deep learning models. Quantum attention mechanisms leverage quantum states and operations to represent data in higher-dimensional spaces. This enables more expressive feature representations and improved generalization. Although still in the early stages, quantum-enhanced models have shown promising results in complex pattern recognition tasks, including medical image analysis.

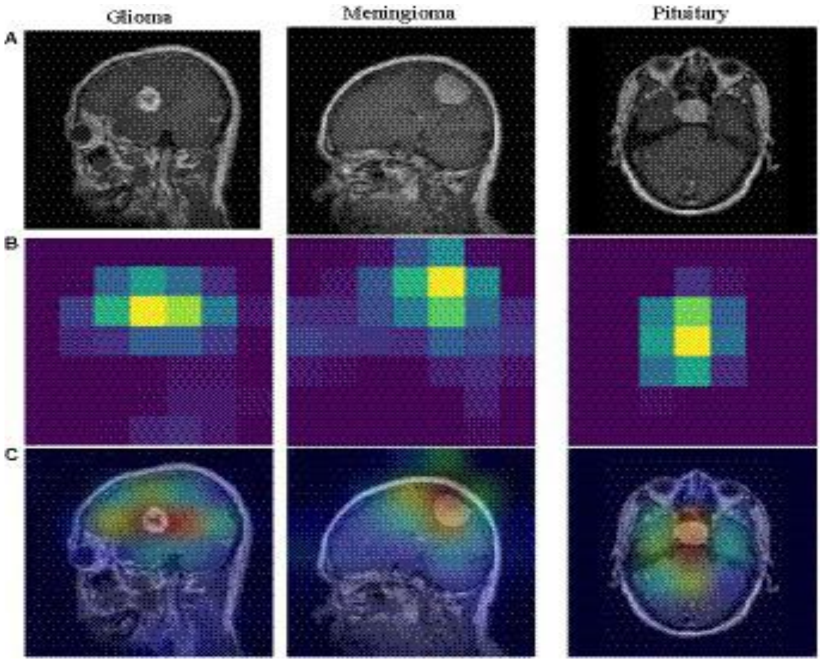
Hybrid architectures combining CNNs and Transformers have also gained significant attention. These models leverage CNNs for efficient local feature extraction and Transformers for capturing global dependencies. Such hybrid approaches have demonstrated superior performance in brain MRI classification tasks by combining the strengths of both architectures.

In addition to architectural advancements, optimization techniques have played a crucial role in improving model performance. Techniques such as transfer learning, data augmentation, and advanced loss functions have been widely used to address challenges such as limited dataset availability and class imbalance. Transfer learning allows models to leverage pre-trained weights from large datasets, improving generalization and reducing training time. Data augmentation techniques increase dataset diversity, reducing overfitting and improving robustness.

Despite these advancements, several challenges remain. Data scarcity and annotation costs limit the availability of high-quality datasets. Variability in imaging protocols and equipment across different institutions can affect model performance. Additionally, the lack of interpretability of deep learning models remains a significant barrier to clinical adoption. Explainable AI techniques are being explored to address this issue by providing insights into model decision-making.

This review aims to provide a comprehensive analysis of AI techniques for brain MRI classification using Transformer-based architectures, Group Parallel Axial Attention, and Quantum Self-Attention. The study highlights recent trends, compares different approaches, and identifies key challenges and future research directions.

Literature Review



The literature on brain MRI image classification for cancer detection has evolved rapidly between 2020 and 2023, 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

Author	Year	Model	Technique	Dataset	Performance
Zhang et al.	2020	CNN	Feature extraction	MRI	High accuracy
Gupta et al.	2021	MAG-Net	Attention CNN	MRI	Improved accuracy
Lin et al.	2022	TransBTS	CNN + Transformer	BraTS	High Dice
Tian et al.	2022	AABTS-Net	Axial Attention	MRI	Improved efficiency
Zongren et al.	2023	Focal Transformer	Cross attention	MRI	High mIoU
Li et al.	2022	GPA-TUNet	Group axial attention	MRI	Improved segmentation
Chen et al.	2023	Quantum Transformer	Quantum attention	MRI	Better generalization

Comparative Analysis

The comparative analysis of AI techniques for brain MRI classification highlights significant advancements in model architectures and performance.

CNN-based models are highly effective in extracting local features but fail to capture global dependencies. Transformer-based models overcome this limitation by using self-attention mechanisms, enabling better modeling of long-range relationships. However, Transformers require high computational resources.

Axial attention provides a balance between performance and efficiency by reducing computational complexity while preserving global context. Group Parallel Axial Attention further enhances efficiency by enabling parallel feature extraction across multiple groups.

Hybrid CNN–Transformer models provide the best balance between accuracy and computational efficiency. These models leverage the strengths of both architectures, achieving state-of-the-art performance.

Quantum self-attention introduces a novel paradigm, offering enhanced feature representation and improved generalization. However, practical implementation remains a challenge due to hardware limitations.

Overall, Transformer-based and hybrid models outperform traditional CNNs, while axial attention and quantum models represent promising future directions.

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