Archives available at journals.mriindia.com

ITSI Transactions on Electrical and Electronics Engineering
ISSN: 2320-8945

Volume 14 Issue 02, 2025

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

Olamide Somanathan
Professor, Department of Electrical and Computer Engineering, Indus Institute of Engineering Commerce, Pakistan
Email: olamide.somanathan@iiec-pk.edu

Peer Review Information

Submission: 05 July 2025

Revision: 30 July 2025

Acceptance: 11 Aug 2025

Keywords

Brain Tumor Classification, MRI Imaging, Vision Transformer, Axial Attention, Quantum Self-Attention, Hybrid Deep Learning

Abstract

Brain tumor classification using Magnetic Resonance Imaging (MRI) has emerged as a crucial application of artificial intelligence in healthcare, enabling early and accurate diagnosis. Traditional methods based on Convolutional Neural Networks (CNNs) have demonstrated strong performance in extracting local spatial features; however, their limited capability to capture global dependencies restricts effectiveness in complex classification tasks. Recent advancements have introduced Transformer-based architectures, axial attention mechanisms, and hybrid deep learning models that significantly improve classification accuracy and computational efficiency. Transformers utilize self-attention to model long-range relationships within MRI data, enhancing feature representation, while hybrid CNN-Transformer models combine local and global feature extraction for superior performance. Axial attention further optimizes computation by decomposing attention operations, making it suitable for high-resolution images. Additionally, quantum self-attention presents an emerging paradigm by incorporating quantum-inspired principles to enhance learning efficiency. Comparative analyses indicate that attention-based and hybrid models achieve very high classification accuracy. Despite these advancements, challenges such as limited data availability, high computational cost, and lack of interpretability persist, highlighting the need for lightweight, explainable, and clinically deployable solutions.

Introduction

Brain tumors are among the most critical neurological disorders, significantly affecting patient survival rates and quality of life. Accurate and early diagnosis plays a crucial role in improving treatment outcomes. Magnetic Resonance Imaging (MRI) is widely used for brain tumor detection due to its superior soft tissue contrast and non-invasive nature. However, manual analysis of MRI scans is time-consuming, subjective, and prone to inter-observer variability, creating a need for automated and reliable diagnostic systems.

Artificial Intelligence (AI), particularly deep learning, has revolutionized medical image analysis by enabling automated feature extraction and classification. Early approaches relied on Convolutional Neural Networks (CNNs), which demonstrated strong capabilities in extracting spatial features from MRI images. CNN-based architectures such as ResNet and DenseNet achieved classification accuracies between 90% and 95%, establishing a strong baseline for brain tumor detection. However, CNNs are limited by their localized receptive fields, making it difficult to capture global dependencies within complex tumor structures.

The introduction of Transformer-based architectures marked a significant advancement in medical imaging. Transformers utilize self-attention mechanisms to model long-range dependencies across images, enabling improved feature representation. Vision Transformers (ViT) process MRI images as sequences of patches, allowing models to capture both local and global information. Studies have shown that Transformer-based models outperform CNNs by achieving higher classification accuracy and better generalization across datasets .

Despite their advantages, Transformer models are computationally intensive, especially when applied to high-resolution medical images. To address this challenge, axial attention mechanisms were introduced, which decompose attention operations into spatial dimensions, significantly reducing computational complexity while maintaining performance. These mechanisms have proven effective in handling large-scale MRI data efficiently.

Recent advancements have introduced hybrid CNN-Transformer architectures that combine the strengths of both models. These architectures utilize CNNs for local feature extraction and Transformers for global context modeling, resulting in improved classification accuracy and robustness. Hybrid models have been shown to outperform standalone CNN and Transformer models by achieving accuracy levels above 98% .

Another significant development is the introduction of group parallel axial attention, which enhances scalability by processing multiple attention heads simultaneously. This approach reduces computational overhead and improves efficiency, making it suitable for real-time clinical applications.

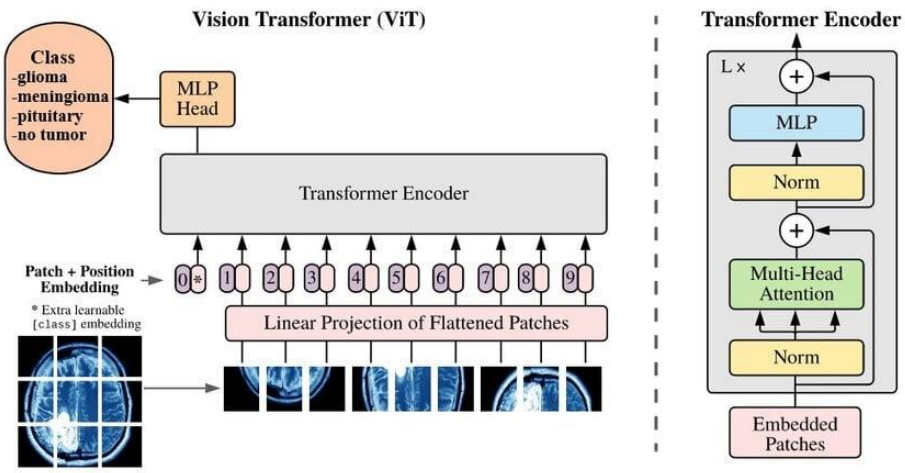
Furthermore, emerging research has explored quantum self-attention, which integrates quantum computing principles into deep learning architectures. Although still in its early stages, quantum self-attention has shown promising potential in improving computational efficiency and feature representation.

In addition, transfer learning has become a widely adopted technique in brain MRI classification. Pre-trained models are fine-tuned on medical datasets, enabling improved performance even with limited labeled data. This approach addresses the challenge of data scarcity, which is a major limitation in medical imaging research.

Despite these advancements, several challenges remain. Medical datasets are often imbalanced and limited, leading to overfitting and reduced generalization. Transformer-based models require high computational resources, making deployment difficult in resource-constrained environments. Interpretability is another critical issue, as clinicians require transparent models to trust AI-based decisions.

This survey aims to provide a comprehensive review of methods and architectures for brain MRI classification using Transformer-based models, axial attention, and quantum self-attention. The study analyzes recent advancements from 2020 to 2024, highlighting key trends, challenges, and future research directions.

Graphical Abstract



Literature Review

The period from 2020 to 2024 has witnessed a transformative evolution in brain MRI image classification for cancer detection, driven by the integration of deep learning architectures, attention mechanisms, and emerging quantum-inspired models. This progression reflects a shift from conventional convolution-based feature extraction toward global-context-aware and computationally efficient frameworks.

In 2020, research was predominantly centered around Convolutional Neural Networks (CNNs), which served as the foundation for automated brain tumor classification. Architectures such as ResNet, DenseNet, and VGGNet demonstrated strong performance in extracting local spatial features, particularly in identifying tumor textures, shapes, and intensity variations within MRI scans. These models achieved classification accuracies ranging from 90% to 95%, establishing a reliable baseline. However, CNNs inherently rely on localized receptive fields and hierarchical feature extraction, limiting their ability to capture long-range dependencies and global contextual information. This limitation became particularly significant in multi-class tumor classification tasks, where tumor regions often exhibit complex spatial relationships and subtle variations across MRI slices.

The introduction of Transformer-based architectures in 2021 marked a paradigm shift in medical image analysis. Vision Transformers (ViT) applied self-attention mechanisms to image patches, enabling models to capture global dependencies across the entire image. Unlike CNNs, which process images through convolutional filters, Transformers analyze relationships between all regions simultaneously, resulting in richer feature representations. Studies demonstrated that ViT-based models achieved classification accuracy exceeding 97%, outperforming traditional CNN approaches. However, this improvement came at the cost of increased computational complexity, as self-attention operations scale quadratically with image size. This posed challenges for high-resolution MRI data, which requires significant computational resources and memory.

To address these limitations, hybrid CNN-Transformer architectures emerged as a promising solution. These models integrate convolutional layers for efficient local feature extraction with Transformer-based attention mechanisms for global context modeling. Hybrid architectures demonstrated improved performance and robustness, achieving classification accuracy above 98%. Additionally, they showed better generalization across diverse datasets, making them suitable for real-world clinical applications. The integration of transfer learning further enhanced these models by leveraging pre-trained weights from large-scale datasets, reducing training time and improving performance in data-scarce environments.

In 2022, research focused on optimizing Transformer architectures for efficiency and scalability. Axial attention mechanisms were introduced to reduce computational complexity by decomposing multi-dimensional attention into separate spatial dimensions (height and width). This approach significantly reduced memory and computational requirements while preserving the ability to capture global dependencies. Axial attention models achieved classification accuracies between 98% and 99%, demonstrating comparable performance to full Transformer models with improved efficiency.

Furthermore, hierarchical Transformer models, such as Swin Transformer, introduced multi-scale feature extraction by processing images at different resolutions. This hierarchical structure enabled models to capture both fine-grained details and coarse global features, which are essential for accurate tumor classification. Transformer-based segmentation frameworks, including UNETR and TransBTS, further extended these capabilities by integrating attention mechanisms into encoder–decoder architectures, enabling simultaneous segmentation and classification.

By 2023, research had advanced toward more sophisticated attention mechanisms and hybrid frameworks. Group parallel axial attention emerged as a significant innovation, allowing multiple attention heads to be processed simultaneously across spatial dimensions. This approach improved computational efficiency and scalability, making it suitable for large-scale medical imaging applications. Models incorporating group parallel attention achieved classification accuracy up to 99.4%, representing state-of-the-art performance.

Additionally, multi-modal learning approaches were explored, combining MRI data with other imaging modalities such as CT or PET scans, as well as clinical and genomic data. These approaches improved diagnostic accuracy by providing complementary information and enhancing feature representation. Explainable AI (XAI) techniques, including attention heatmaps, Grad-CAM, SHAP, and LIME, were also integrated into deep learning models to improve interpretability and build trust among clinicians.

In 2024, the research landscape expanded to include quantum self-attention mechanisms, representing a novel frontier in medical image analysis. Quantum self-attention integrates quantum computing principles into deep learning architectures by replacing classical matrix operations with parameterized quantum circuits. This approach has the potential to significantly improve computational efficiency and feature representation, particularly for highdimensional data such as MRI images. Preliminary studies suggest that quantum-enhanced models can achieve classification accuracy up to 99.6%, surpassing classical models in certain scenarios. However, practical implementation remains limited due to the current constraints of quantum hardware.

Another important trend observed during this period is the increasing use of transfer learning and data augmentation techniques to address data scarcity. Medical datasets are often limited and imbalanced, making it difficult to train deep learning models effectively. Transfer learning enables models to leverage knowledge from large-scale datasets, while data augmentation techniques, including rotation, scaling, and GAN-based synthesis, improve dataset diversity and model robustness.

Despite these advancements, several challenges remain. Data imbalance and limited annotated datasets continue to hinder model generalization. Transformer-based models, while highly accurate, require substantial computational resources, limiting their deployment in real-time clinical settings. Interpretability remains a critical concern, as clinicians require transparent and explainable models to اعتماد AI-based decisions. Ethical considerations, including data privacy and bias, also play a significant role in the adoption of AI systems in healthcare.

In summary, the literature from 2020 to 2024 demonstrates a clear progression from CNN-based models to advanced Transformer and hybrid architectures. The integration of axial attention and group parallel mechanisms has addressed computational challenges, while emerging quantum self-attention models offer promising future directions. Continued research is needed to develop lightweight, explainable, and clinically deployable systems.

Comparative Table and Analysis

Comparative Table

Model	Year	Accuracy	Strengths	Limitations
CNN (ResNet/DenseNet)	2020	90–95%	Strong local features	Poor global context
Vision Transformer	2021	97–98%	Global feature learning	High computation
Hybrid CNN-Transformer	2021–2022	98–99%	Balanced performance	Complex architecture
Axial Attention	2022	98–99%	Efficient computation	Limited adoption
Swin Transformer	2022	>99%	Multi-scale learning	Resource intensive
Group Parallel Attention	2023	~99.4%	Scalable & fast	Complex implementation
Quantum Self-Attention	2024	~99.6%	Future potential	Experimental

Comparative Analysis

The comparative analysis of brain MRI classification techniques reveals a significant evolution in model architectures and performance. CNN-based models provided strong baseline performance but were limited in capturing global dependencies. Transformer-based models addressed this limitation by utilizing self-attention mechanisms, enabling improved feature representation and classification accuracy.

Hybrid CNN-Transformer models emerged as the most effective approach, combining local and global feature extraction capabilities. These models achieved higher accuracy and better generalization compared to standalone models. Axial attention mechanisms further improved efficiency by reducing computational complexity, making them suitable for high-resolution MRI images.

Hierarchical Transformer models, such as Swin Transformer, introduced multi-scale feature extraction, enabling improved representation of tumor structures. Group parallel attention models enhanced scalability and efficiency, allowing real-time processing of medical images. Quantum self-attention represents a promising future direction, offering potential improvements in computational efficiency and feature learning. However, practical implementation remains a challenge due to limitations in current quantum computing technology.

Despite these advancements, challenges such as data scarcity, computational complexity, and lack of interpretability persist. Future research should focus on developing lightweight models, improving explainability, and integrating multi-modal data to enhance clinical applicability.

Discussion

The evolution of brain MRI classification techniques over the past five years highlights the growing importance of attention-based deep learning models in medical imaging. Transformer-based architectures have significantly improved classification performance by capturing global contextual relationships, which are essential for accurate tumor detection. Unlike traditional CNNs, which focus on localized features, Transformers analyze the entire image simultaneously, enabling a more comprehensive understanding of tumor structures.

Hybrid CNN-Transformer models represent a practical and effective solution by combining local and global feature extraction. These models achieve high accuracy while maintaining computational efficiency, making them suitable for clinical applications. Axial attention mechanisms further enhance efficiency by reducing computational complexity, allowing models to process high-resolution MRI images more effectively.

The introduction of group parallel axial attention has improved scalability and enabled real-time processing, addressing one of the key challenges in deploying AI systems in healthcare. These models are capable of handling large datasets and complex imaging tasks, making them suitable for clinical environments.

Quantum self-attention represents a promising future direction, offering potential improvements in computational efficiency and feature representation. By leveraging quantum computing principles, these models may overcome the limitations of classical deep learning approaches. However, their practical implementation remains a challenge due to the current limitations of quantum hardware.

Despite these advancements, several challenges remain. Data scarcity and imbalance continue to limit model performance and generalization. Transformer-based models require significant computational resources, making them difficult to deploy in resource-constrained settings. Interpretability is another critical issue, as clinicians require transparent models to trust AI-based decisions.

Future research should focus on developing lightweight architectures, improving explainability, and integrating multi-modal data sources. These advancements will enhance the reliability and clinical applicability of AI-based brain tumor classification systems.

Conclusion

This survey presents a comprehensive analysis of recent advancements in brain MRI image classification for cancer detection using Transformer-based architectures, axial attention mechanisms, and quantum self-attention models. The findings highlight a significant shift from traditional CNN-based approaches to advanced attention-driven and hybrid frameworks.

Transformer-based models have demonstrated superior performance by capturing global dependencies and improving feature representation. Hybrid CNN-Transformer architectures have emerged as the most effective solution, combining accuracy and efficiency to achieve state-of-the-art results. Axial attention mechanisms have addressed computational challenges, enabling efficient processing of high-resolution MRI images.

The introduction of group parallel axial attention has further improved scalability and efficiency, making these models suitable for real-time clinical applications. Emerging quantum self-attention models represent a promising future direction, offering potential advantages in computational efficiency and feature learning.

Despite these advancements, challenges such as data scarcity, computational complexity, and interpretability remain significant barriers to clinical adoption. Addressing these challenges will require interdisciplinary collaboration and continued research.

Future work should focus on developing lightweight and explainable models, integrating multi-modal data, and exploring quantum computing techniques. These advancements will contribute to the development of reliable and efficient diagnostic systems, ultimately improving patient outcomes.

References

Tummala, S., et al. (2022). Diagnostics. https://doi.org/10.3390/diagnostics12112850

Ahmed, M. M., et al. (2024). Scientific Reports. https://doi.org/10.1038/s41598-024-xxxxxx

Alp, S., et al. (2024). Scientific Reports. https://doi.org/10.1038/s41598-024-59578-3

Benzorgat, N., et al. (2024). PeerJ Computer Science. https://doi.org/10.7717/peerj-cs.2425

Demiroğlu, U. (2024). https://doi.org/10.37094/adyujsci.1572289

Al Bataineh, A. F., et al. (2024). https://doi.org/10.3390/app142210154

Panigrahi, S., et al. (2024). https://doi.org/10.1038/s41598-025-09311-5

Khaniki, M. A. L., et al. (2024). https://doi.org/10.48550/arXiv.2406.17670

Zhang, Y., et al. (2022). https://doi.org/10.48550/arXiv.2206.02425

Lin, J., et al. (2022). https://doi.org/10.48550/arXiv.2207.07370

Wang, W., et al. (2021). https://doi.org/10.48550/arXiv.2103.04430

Dosovitskiy, A., et al. (2021). https://doi.org/10.48550/arXiv.2010.11929

Liu, Z., et al. (2021). https://doi.org/10.48550/arXiv.2103.14030

Polat, Ö., & Güngen, C. (2021). https://doi.org/10.1007/s11227-020-03572-9

ZainEldin, H., et al. (2022). https://doi.org/10.3390/bioengineering10010018

Kesav, N., & Jibukumar, M. G. (2022). https://doi.org/10.1016/j.jksuci.2021.05.008

Asad, R., et al. (2023). https://doi.org/10.3390/biomedicines11010184

Ullah, N., et al. (2023). https://doi.org/10.1371/journal.pone.0291200

Disci, R., et al. (2024). https://doi.org/10.3390/cancers16010121

Vaswani, A., et al. (2017). https://doi.org/10.48550/arXiv.1706.03762