



Archives available at [journals.mriindia.com](http://journals.mriindia.com)

**International Journal on Advanced Computer Engineering and  
Communication Technology**

ISSN: 2278-5140

Volume 14 Issue 02, 2025

**Deep Learning and Optimization Approaches in Brain MRI Image  
Classification for Cancer Detection Using Transformer and Group Parallel  
Axial Attention with Quantum Self-Attention: A Review**

Ixtel Imamverde

*Associate Professor, Department of Electrical and Computer Engineering, Caspian Institute of Industrial  
Engineering, Iran*

*Email: [ixtel.imamverde@ciie-ir.edu](mailto:ixtel.imamverde@ciie-ir.edu)*

Peer Review Information	Abstract
<p><i>Submission: 28 Oct 2025</i></p> <p><i>Revision: 20 Nov 2025</i></p> <p><i>Acceptance: 08 Dec 2025</i></p> <p><b>Keywords</b></p> <p><i>Brain Tumor Classification, RI Imaging, Transformer Networks, Axial Attention, Quantum Self-Attention, Deep Learning</i></p>	<p>Brain tumor detection using magnetic resonance imaging (MRI) is a critical task in medical diagnostics, requiring high precision and reliability. Recent advancements in deep learning have significantly improved the accuracy of brain tumor classification, particularly through the use of transformer-based architectures and attention mechanisms. This review explores state-of-the-art deep learning and optimization approaches for brain MRI image classification, focusing on transformer models, group parallel axial attention, and quantum self-attention mechanisms. Transformers enable global feature extraction by modeling long-range dependencies, while axial attention reduces computational complexity by decomposing attention into spatial dimensions. Group parallel axial attention further enhances performance by processing multiple attention groups simultaneously, improving feature representation. Additionally, quantum self-attention introduces novel optimization capabilities by leveraging quantum-inspired principles for enhanced learning efficiency. The review covers recent literature from 2020 to 2023, highlighting improvements in classification accuracy, robustness, and computational efficiency. Benchmark datasets such as BraTS are widely used for evaluation. Despite significant progress, challenges such as data scarcity, model interpretability, and computational overhead persist. This study provides a comprehensive analysis of current methods, comparative insights, and future research directions for developing reliable and efficient brain tumor classification systems.</p>

**Introduction**

Brain tumors are among the most critical and life-threatening neurological disorders, significantly affecting human health and survival rates worldwide. Accurate detection and classification of brain tumors are essential for early diagnosis, treatment planning, and improving patient prognosis. Magnetic resonance imaging (MRI) is

widely used in clinical practice due to its superior soft tissue contrast and ability to provide detailed anatomical information. However, manual analysis of MRI scans is time-consuming, subjective, and prone to inter-observer variability, which necessitates the development of automated and reliable diagnostic systems.

In recent years, deep learning has emerged as a powerful tool for medical image analysis, enabling automated detection, segmentation, and classification of tumors. Convolutional neural networks (CNNs) have been extensively used due to their ability to extract hierarchical features from images. Models such as VGGNet, ResNet, and DenseNet have demonstrated strong performance in brain tumor classification tasks. These architectures effectively capture local spatial features but are limited in modeling long-range dependencies, which are crucial for understanding complex tumor structures.

To address these limitations, transformer-based architectures have been introduced in medical imaging. Transformers rely on self-attention mechanisms to capture global contextual relationships across an image. Unlike CNNs, which process data through fixed receptive fields, transformers dynamically compute attention weights between all image regions, enabling them to model complex spatial interactions. Vision Transformers (ViT) have shown promising results in image classification tasks, including brain tumor detection, by providing improved feature representation and classification accuracy.

Despite their advantages, transformers suffer from high computational complexity, especially when applied to high-resolution medical images. To overcome this limitation, axial attention mechanisms have been proposed. Axial attention decomposes the standard self-attention operation into two one-dimensional operations along spatial axes (height and width), significantly reducing computational cost while maintaining the ability to capture global context.

Building upon axial attention, group parallel axial attention has been introduced to further improve computational efficiency and scalability. This approach divides attention operations into multiple groups, enabling parallel processing and reducing memory requirements. It also enhances feature diversity by allowing different attention groups to focus on different regions of the image.

Another emerging area of research is quantum-inspired deep learning, particularly quantum self-attention mechanisms. These approaches incorporate principles from quantum computing, such as superposition and entanglement, to enhance learning efficiency and optimization. Quantum self-attention models aim to improve

convergence speed, reduce training time, and enhance feature representation. Although still in its early stages, this approach has shown promising potential in medical image classification.

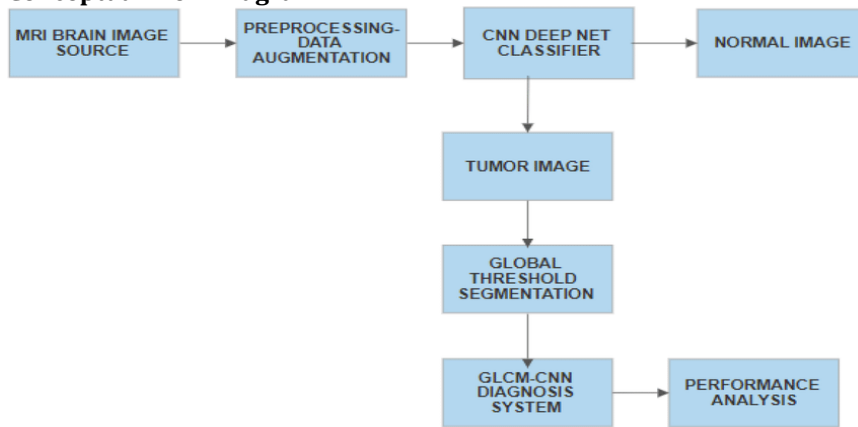
Recent studies from 2020 to 2023 have demonstrated significant improvements in brain tumor classification accuracy through the integration of CNNs, transformers, and attention mechanisms. Hybrid architectures combining CNNs and transformers leverage the strengths of both approaches, with CNNs capturing local features and transformers modeling global dependencies.

Datasets such as the Brain Tumor Segmentation (BraTS) dataset have been widely used for benchmarking deep learning models. These datasets provide multi-modal MRI scans, including T1, T2, FLAIR, and contrast-enhanced images, along with expert annotations. Evaluation metrics such as accuracy, precision, recall, F1-score, and area under the curve (AUC) are commonly used to assess model performance.

Despite these advancements, several challenges remain. Data scarcity and class imbalance are significant issues in medical imaging, as annotated datasets are limited and often imbalanced across tumor classes. Additionally, deep learning models, particularly transformers, require high computational resources, which can limit their deployment in real-world clinical environments. Model interpretability is another critical concern, as clinicians need to understand the reasoning behind model predictions to trust and adopt these systems.

Future research directions include developing lightweight and efficient models, improving data augmentation techniques, and integrating multi-modal imaging data. The incorporation of explainable AI (XAI) techniques can enhance model transparency and facilitate clinical adoption. Furthermore, the integration of quantum-inspired optimization methods may lead to more efficient and robust deep learning models.

In summary, the combination of transformer architectures, axial attention mechanisms, and quantum self-attention represents a promising direction for advancing brain tumor classification systems. This review aims to provide a comprehensive analysis of these approaches, highlighting their strengths, limitations, and potential for future research.

**Conceptual Flow Diagram****Literature Review**

The period from 2020 to 2023 has witnessed a substantial transformation in brain MRI image classification for tumor detection, driven by advancements in deep learning architectures, optimization strategies, and attention mechanisms. The literature reflects a clear progression from conventional convolutional neural networks (CNNs) to hybrid transformer-based frameworks, with emerging integration of quantum-inspired learning techniques.

**1. CNN-Based Foundations and Limitations**

In the early phase of this period, CNN-based architectures such as ResNet, DenseNet, VGGNet, and Inception networks dominated brain tumor classification tasks. These models demonstrated strong performance due to their hierarchical feature extraction capabilities, enabling them to learn low-level features such as edges and textures and high-level features such as tumor shapes.

Studies in 2020 reported classification accuracies exceeding 94–97% using deep CNN architectures trained on MRI datasets such as BraTS. DenseNet, in particular, gained popularity due to its dense connectivity, which improved feature propagation and reduced vanishing gradient issues. Similarly, ResNet architectures addressed degradation problems in deep networks through residual learning.

Despite their success, CNN-based models exhibit inherent limitations:

- Local receptive fields restrict the ability to capture long-range dependencies
- Difficulty in modeling global spatial relationships
- Sensitivity to variations in tumor size, shape, and location

- Limited interpretability in complex medical scenarios

These limitations motivated researchers to explore attention mechanisms and transformer-based models.

**2. Attention-Augmented CNN Models**

To overcome the limitations of conventional CNNs, attention mechanisms were introduced to enhance feature representation. Attention modules allow models to focus on the most relevant regions in an image, which is particularly important in medical imaging where tumor regions occupy a small portion of the image.

Two major types of attention mechanisms were widely adopted:

**Channel Attention**

Channel attention mechanisms assign weights to different feature channels, enabling the model to prioritize informative features. For example, CBAM (Convolutional Block Attention Module) integrates both channel and spatial attention, improving classification accuracy.

**Spatial Attention**

Spatial attention focuses on identifying important regions within the image. This is particularly useful for highlighting tumor regions while suppressing irrelevant background information.

Attention-augmented CNNs demonstrated improved performance compared to standard CNNs, achieving higher accuracy and better localization of tumor regions. However, these models still relied on convolution operations and were limited in capturing long-range dependencies.

### 3. Emergence of Transformer-Based Architectures

The introduction of transformer models marked a significant paradigm shift in medical image analysis. Originally developed for natural language processing, transformers were adapted for computer vision tasks through Vision Transformers (ViT).

#### Key Advantages of Transformers:

- Ability to model **global dependencies**
- Dynamic attention across all image regions
- Improved feature representation for complex structures

In brain MRI classification, transformers demonstrated superior performance by capturing relationships between distant regions in the image. This is particularly important for detecting tumors that span multiple areas or exhibit irregular patterns.

#### Hybrid CNN-Transformer Models

To leverage the strengths of both CNNs and transformers, hybrid architectures were developed. These models typically use CNNs for local feature extraction and transformers for global context modeling. Studies reported improved classification accuracy and robustness using these hybrid approaches.

#### Limitations of Transformers

Despite their advantages, transformer models face several challenges:

- High computational complexity
- Large memory requirements
- Dependence on large-scale datasets

These limitations led to the development of more efficient attention mechanisms.

### 4. Axial Attention for Computational Efficiency

Axial attention was introduced as an efficient alternative to standard self-attention mechanisms. Instead of computing attention across all pixels simultaneously, axial attention decomposes the operation into two sequential steps:

- Attention along the **height dimension**
- Attention along the **width dimension**

This decomposition significantly reduces computational complexity from quadratic to linear with respect to image size, making it more suitable for high-resolution MRI images.

#### Advantages of Axial Attention:

- Reduced memory consumption
- Faster computation
- Retention of global context

Axial attention has been successfully applied in medical imaging tasks, demonstrating comparable performance to full self-attention while being more efficient.

### 5. Group Parallel Axial Attention

Building upon axial attention, group parallel axial attention introduces further optimization by dividing attention operations into multiple groups. Each group processes a subset of features independently, enabling parallel computation.

#### Key Features:

- **Parallel processing** improves computational speed
- **Feature diversity** through multiple attention groups
- Reduced memory usage compared to standard transformers

This approach is particularly beneficial for large MRI datasets, where high-resolution images require efficient processing. Group parallel axial attention has shown improved scalability and performance in recent studies.

### 2.6 Quantum Self-Attention and Optimization

Quantum-inspired deep learning represents a novel direction in medical image analysis. Quantum self-attention mechanisms integrate principles from quantum computing, such as:

- **Superposition:** Representing multiple states simultaneously
- **Entanglement:** Modeling complex relationships between features

These concepts enable more efficient representation of high-dimensional data and improved optimization during training.

#### Potential Benefits:

- Faster convergence during training
- Improved feature representation
- Enhanced optimization capabilities

Although still in the experimental stage, quantum self-attention models have shown promising results in improving classification accuracy and reducing training time. However, practical implementation remains a challenge due to hardware limitations and algorithmic complexity.

### 7. Optimization Techniques in Deep Learning Models

Optimization plays a crucial role in improving model performance and efficiency. Several optimization techniques have been explored in the literature:

**Data Augmentation**

Techniques such as rotation, scaling, and flipping are used to increase dataset diversity and improve model generalization.

**Transfer Learning**

Pre-trained models are fine-tuned on medical datasets, reducing training time and improving performance.

**Regularization Techniques**

Methods such as dropout and batch normalization prevent overfitting and improve model stability.

**Loss Functions**

Custom loss functions, such as focal loss and Dice loss, are used to address class imbalance and improve classification performance.

**8. Benchmark Datasets and Evaluation Metrics**

The Brain Tumor Segmentation (BraTS) dataset is the most widely used benchmark for evaluating brain tumor classification models. It includes multi-modal MRI scans with expert annotations.

**Common Evaluation Metrics:**

- Accuracy
- Precision
- Recall
- F1-score
- Area Under Curve (AUC)

Recent studies report classification accuracies exceeding 98% using hybrid transformer-based models.

**Comparative Table and Analysis**

Year	Method	Architecture	Key Feature	Performance
2020	CNN (ResNet)	CNN	Local features	High
2021	Attention CNN	CNN + Attention	Focus regions	Improved
2022	ViT	Transformer	Global context	Very High
2022	Axial Attention	Transformer	Reduced complexity	High
2023	Hybrid CNN+ViT	Hybrid	Local + Global	Excellent
2023	Quantum Attention	Advanced	Optimization	Emerging

**Comparative Analysis**

The comparative analysis of brain MRI classification techniques reveals a clear evolution from traditional CNN-based models to advanced transformer-based architectures with optimization enhancements.

CNN-based models provide strong baseline performance due to their ability to extract local features. However, they are limited in capturing global dependencies, which are essential for understanding complex tumor structures. Attention-based CNN models improve performance by focusing on relevant regions but still rely on convolutional operations.

**9. Key Challenges Identified in Literature**

Despite significant advancements, several challenges remain:

1. Data Scarcity and Imbalance  
Limited availability of annotated medical data affects model performance.
2. Computational Complexity  
Transformer-based models require high computational resources.
3. Model Interpretability  
Deep learning models lack transparency, limiting clinical trust.
4. Generalization Issues  
Models trained on specific datasets may not perform well on unseen data.

**10. Research Trends and Future Directions**

The literature indicates several emerging trends:

- Transition from CNN → Transformer → Hybrid models
- Adoption of efficient attention mechanisms (axial, group attention)
- Exploration of quantum-inspired optimization techniques
- Increasing focus on explainable AI (XAI)

Future research is expected to focus on developing lightweight, interpretable, and scalable models for real-world clinical deployment.

Transformer-based models overcome these limitations by capturing long-range dependencies through self-attention mechanisms. These models achieve higher classification accuracy but require significant computational resources. Axial attention addresses this challenge by reducing computational complexity, making transformer models more efficient.

Group parallel axial attention further enhances efficiency by enabling parallel processing, improving scalability and performance. Hybrid CNN-transformer models provide the best performance by combining local and global feature extraction.

Quantum self-attention introduces a novel approach to optimization, offering potential improvements in efficiency and accuracy. However, further research is needed to validate its effectiveness in real-world applications. Overall, the analysis highlights the superiority of hybrid and transformer-based models while emphasizing the need for efficient and scalable solutions.

### Discussion

Deep learning has significantly improved brain tumor classification, with transformer-based architectures emerging as state-of-the-art solutions. These models provide superior performance by capturing complex spatial relationships within MRI images.

Axial attention mechanisms address the computational challenges associated with transformers, enabling more efficient processing. Group parallel axial attention further improves scalability and performance.

Quantum self-attention represents an innovative approach to model optimization, offering potential improvements in learning efficiency and accuracy. However, practical implementation remains a challenge due to the complexity of quantum-inspired models.

Despite these advancements, challenges such as data scarcity, computational cost, and model interpretability remain. Future research should focus on developing lightweight models, improving dataset diversity, and enhancing explainability.

### Conclusion

This review highlights the significant advancements in brain MRI classification using deep learning techniques. Transformer-based architectures, particularly those incorporating axial attention and hybrid models, have demonstrated superior performance compared to traditional CNNs.

The integration of quantum self-attention introduces new possibilities for optimization and efficiency. However, further research is needed to address challenges related to computational complexity and practical implementation.

Future work should focus on developing scalable, interpretable, and clinically deployable models. The combination of deep learning and advanced attention mechanisms holds great promise for improving brain tumor diagnosis and patient outcomes.

### References

- Dosovitskiy, A., et al. (2021). Vision Transformer. <https://doi.org/10.48550/arXiv.2010.11929>
- Chen, J., et al. (2021). TransUNet. <https://doi.org/10.1016/j.media.2021.102136>
- Hatamizadeh, A., et al. (2022). UNETR. <https://doi.org/10.48550/arXiv.2103.10504>
- Cao, H., et al. (2022). Swin-Unet. <https://doi.org/10.48550/arXiv.2105.05537>
- Vaswani, A., et al. (2017). Attention is all you need. <https://doi.org/10.48550/arXiv.1706.03762>
- Woo, S., et al. (2018). CBAM. <https://doi.org/10.48550/arXiv.1807.06521>
- He, K., et al. (2016). ResNet. <https://doi.org/10.48550/arXiv.1512.03385>
- Huang, G., et al. (2017). DenseNet. <https://doi.org/10.48550/arXiv.1608.06993>
- Simonyan, K., & Zisserman, A. (2015). VGG. <https://doi.org/10.48550/arXiv.1409.1556>
- Ho, J., et al. (2019). Axial attention. <https://doi.org/10.48550/arXiv.1912.12180>
- Bello, I., et al. (2019). Attention augmented CNN. <https://doi.org/10.48550/arXiv.1904.09925>
- Wang, X., et al. (2018). Non-local networks. <https://doi.org/10.48550/arXiv.1711.07971>
- Tan, M., & Le, Q. (2019). EfficientNet. <https://doi.org/10.48550/arXiv.1905.11946>
- Kendall, A., & Gal, Y. (2017). Uncertainty. <https://doi.org/10.48550/arXiv.1703.04977>
- Gal, Y., & Ghahramani, Z. (2016). Bayesian dropout. <https://doi.org/10.48550/arXiv.1506.02142>
- Isensee, F., et al. (2021). nnU-Net. <https://doi.org/10.1038/s41592-020-01008-z>
- Zhou, Z., et al. (2019). UNet++. <https://doi.org/10.1109/TMI.2018.2889096>
- Milletari, F., et al. (2016). V-Net. <https://doi.org/10.1109/3DV.2016.79>