IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Review Of Hybrid CNN–Attention Architectures For Brain Tumor Detection And Multi-Class Classification Using MRI

[1]Saundarya Tikhile, [2]Shweta Meshram

[1]M-Tech Student, Electronics and Communication System, Government College of Engineering, Amravati, Maharashtra, India [2]Assistant Professor, Electronics and Communication System, Government College of Engineering, Amravati, Maharashtra, India

Abstract: Brain tumor diagnosis using Magnetic Resonance Imaging (MRI) remains a challenging task due to tumor heterogeneity, complex morphology, and variability across imaging modalities. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved automated detection and classification performance. However, conventional CNNs often struggle to capture global contextual dependencies and long-range spatial relationships. To address these limitations, hybrid architectures integrating attention mechanisms—such as channel, spatial, self-attention, and Transformer-based modules—have emerged as state-of-the-art solutions.

This paper presents a comprehensive review of hybrid CNN–attention frameworks for brain tumor detection and multi-class classification. The review systematically analyzes architectural designs, datasets, preprocessing strategies, evaluation metrics, and performance trends. A taxonomy of hybrid models is proposed, including CNN + attention modules, CNN + Transformer hybrids, and multi-branch fusion networks. Comparative analysis highlights that hybrid attention-based models consistently outperform standalone CNNs, achieving accuracy levels exceeding 98% on benchmark MRI datasets.

Furthermore, this review identifies critical research gaps, including lack of clinical generalization, dataset bias, interpretability limitations, and computational overhead. Future directions focusing on explainable AI, federated learning, and lightweight deployment are discussed. The study aims to provide a structured foundation for researchers developing next-generation intelligent diagnostic systems.

Index Terms - Brain Tumor Classification, MRI, Hybrid CNN, Attention Mechanisms, Vision Transformers, Deep Learning, Medical Imaging.

1. INTRODUCTION

Brain tumors are among the most life-threatening neurological disorders, requiring early and accurate diagnosis for effective treatment planning. Traditional diagnostic approaches rely on manual interpretation of MRI scans, which is time-consuming and prone to inter-observer variability.

Deep learning has revolutionized medical image analysis by enabling automated feature extraction and classification. CNNs, in particular, have demonstrated superior performance compared to handcrafted feature-based methods. However, standard CNN architectures primarily focus on local feature extraction, limiting their ability to model global dependencies and contextual relationships within MRI images.

Recent advancements have introduced attention mechanisms and Transformer-based architectures, which enhance feature representation by focusing on salient regions and capturing long-range dependencies. Hybrid CNN–attention architectures combine the strengths of both paradigms:

- CNN → Local feature extraction
- Attention/Transformer → Global context modeling

These hybrid systems have shown significant improvements in classification accuracy, robustness, and generalization.

The figure illustrates the evolution of brain tumor detection approaches across three stages:

- Traditional Machine Learning: Relies on handcrafted features (e.g., GLCM, wavelets) and classifiers like SVM/KNN, with limited adaptability.
- CNN-Based Deep Learning: Enables automatic feature extraction using architectures such as VGG, ResNet, and DenseNet, improving accuracy.
- Hybrid CNN–Attention Models: Combine CNNs with attention/Transformer mechanisms to capture global context, leading to the highest performance.

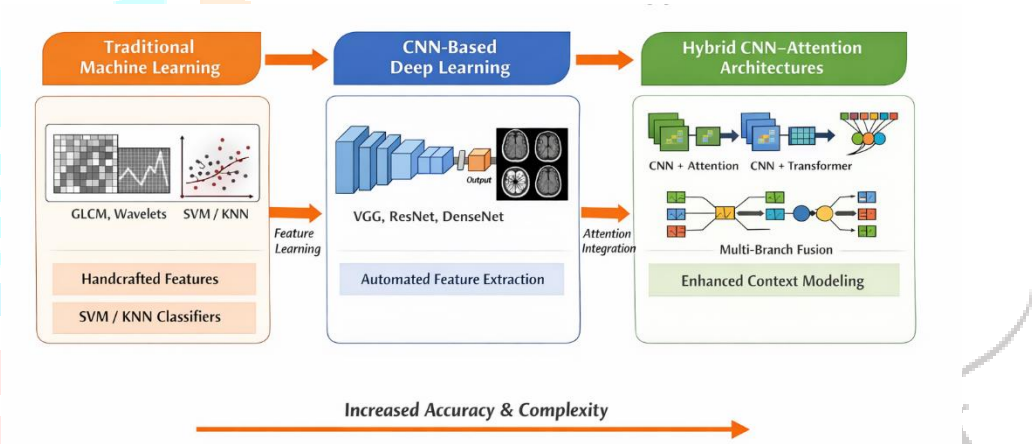

Figure 1: Evolution of Brain Tumor Detection Approaches

Overall, the progression shows that accuracy and model complexity increase from left to right, with hybrid models providing the most advanced feature representation.

2. BACKGROUND AND PROBLEM FORMULATION

2.1. Brain Tumor Classification

Brain tumor classification using MRI plays a fundamental role in computer-aided diagnostic systems, enabling early and accurate identification of tumor types. In clinical practice, brain tumors are broadly categorized into glioma, meningioma, pituitary tumor, and normal (non-tumorous) cases. Among these, gliomas are highly aggressive and exhibit significant heterogeneity, while meningiomas and pituitary tumors are generally more structured and localized.

From a machine learning perspective, this task is formulated as a multi-class classification problem, where the model must distinguish between multiple tumor categories that often exhibit overlapping visual characteristics. This complexity is further amplified by intra-class variability, particularly in glioma cases, where tumor appearance changes across stages and patients. Accurate classification is therefore critical, as treatment strategies such as surgical planning, chemotherapy, or radiotherapy are highly dependent on tumor type and severity.

This figure presents representative MRI images for multi-class brain tumor classification, highlighting visual differences across categories:

- Glioma: Appears as an irregular, heterogeneous mass with diffuse boundaries, indicating high aggressiveness.
- Meningioma: Typically well-defined and localized, often near the outer brain regions.
- Pituitary Tumor: Located near the pituitary gland (base of the brain), usually more compact and structured.
- Normal Brain: No abnormal mass or structural distortion.

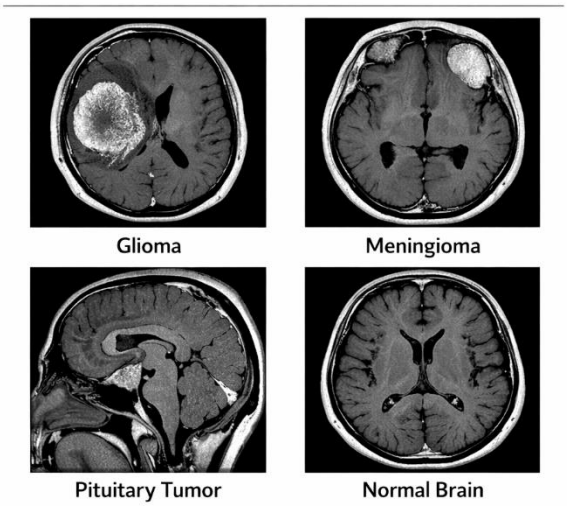

Figure 2: MRI-Based Brain Tumor Classification Examples

Overall, the figure demonstrates inter-class variability and similarity, emphasizing the challenge for models to accurately distinguish tumor types based on subtle visual differences in MRI images.

2.2. Challenges in MRI-Based Diagnosis

Despite the availability of advanced imaging technologies, automated brain tumor classification remains challenging due to several inherent characteristics of MRI data. As outlined in, one of the primary issues is tumor heterogeneity, where variations in shape, size, and texture make it difficult for models to learn consistent feature representations. This is particularly evident in gliomas, which often have irregular and diffused boundaries.

Another major challenge is the limited availability of annotated datasets. Medical image labeling requires expert radiologists, making large-scale, high-quality datasets scarce. This scarcity directly affects deep learning performance, leading to overfitting and reduced generalization.

Additionally, MRI data suffers from modality and intensity variations, as different imaging sequences such as T1, T2, and FLAIR highlight different tissue characteristics. These variations introduce inconsistencies across datasets and contribute to domain shift problems. Class imbalance further complicates the learning process, as certain tumor types are underrepresented, biasing the model toward majority classes.

Finally, the high dimensionality of MRI data increases computational requirements, making training and deployment resource-intensive, especially for advanced architectures.

This figure illustrates the key challenges in MRI-based brain tumor classification:

- Class Similarity: Different tumor types can appear visually similar, making it difficult for models to distinguish between them.
- Tumor Variability: Tumors vary in size, shape, and location across patients, increasing classification complexity.
- Data Imbalance: Some classes (e.g., normal cases) dominate datasets, leading to biased model learning.

- Annotator Discrepancy: Variations in expert labeling can introduce inconsistencies in ground truth data.

Overall, the figure highlights that both data-related and clinical factors significantly impact model accuracy and reliability.

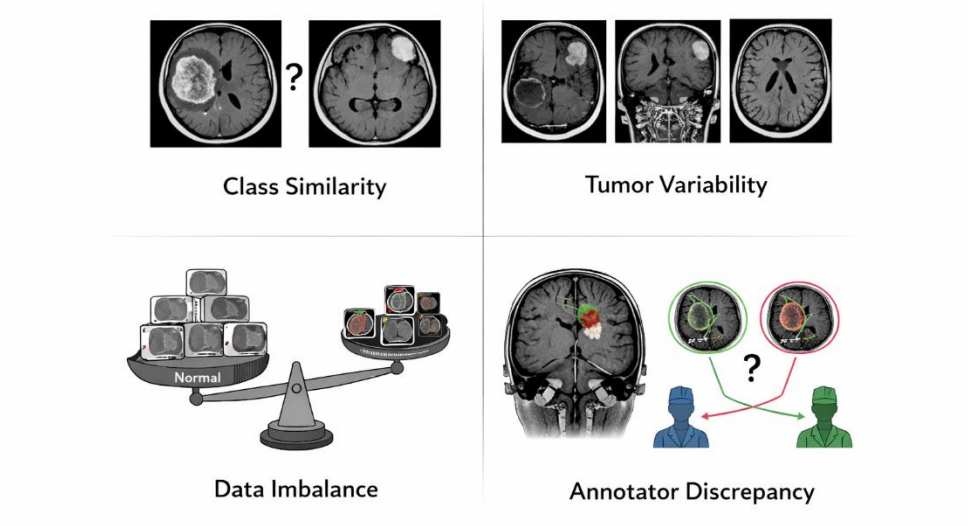

Figure 3: Key Challenges in MRI-Based Brain Tumor Classification

2.3. Limitations of Conventional CNNs

Convolutional Neural Networks have achieved considerable success in medical image analysis; however, their architectural design introduces several limitations when applied to brain tumor classification. One of the primary constraints is their reliance on local receptive fields, which restricts the ability to capture long-range spatial dependencies. In MRI analysis, understanding the global context of tumor regions relative to surrounding tissues is often essential for accurate diagnosis.

Another important limitation is the lack of interpretability. CNNs operate as black-box models, making it difficult to explain their predictions in clinically meaningful terms. This lack of transparency reduces trust and limits adoption in real-world healthcare settings.

Furthermore, CNNs are prone to overfitting when trained on limited medical datasets, as highlighted in . Their large number of parameters allows them to memorize training data rather than generalize effectively. In addition, conventional CNN architectures struggle to capture multi-scale features efficiently, which is critical in brain tumor analysis where lesions can vary significantly in size and structure.

These limitations collectively motivate the development of hybrid CNN–attention architectures, which aim to integrate local feature extraction with global contextual modeling for improved performance.

This figure highlights the key limitations of conventional CNNs in brain tumor MRI analysis:

- Limited Receptive Field: CNNs focus on small local regions, which may miss the full extent of a tumor.
- Lack of Contextual Awareness: They cannot effectively capture relationships between distant regions, reducing diagnostic accuracy.
- Inability to Model Relationships: CNNs struggle to understand interactions between different brain structures, leading to incomplete feature representation.

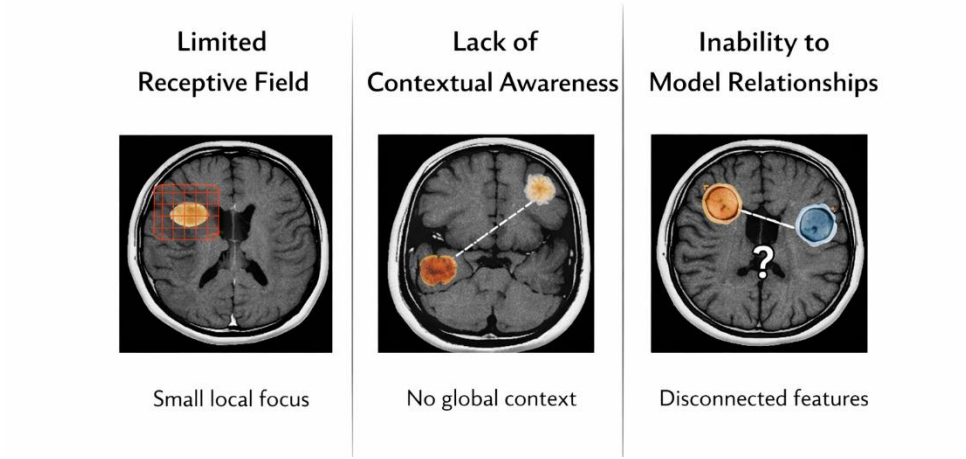

Figure 4: Limitations of CNNs in Brain Tumor MRI Analysis

Overall, the figure shows that CNNs are strong in local feature extraction but weak in global context modeling, which motivates the use of attention and Transformer-based architectures.

3. Evolution of Deep Learning Approaches

3.1. Traditional Machine Learning

Early approaches to brain tumor classification relied heavily on handcrafted feature extraction techniques such as Gray-Level Co-occurrence Matrix (GLCM) and wavelet transforms. These methods aimed to capture texture, intensity, and structural patterns from MRI images, which were subsequently classified using machine learning algorithms such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN).

While these approaches provided initial insights into automated diagnosis, their effectiveness was limited by the dependency on manually engineered features. Such features often fail to generalize across datasets with varying imaging conditions. Furthermore, the lack of scalability and adaptability restricts their performance in complex classification tasks involving high intra-class variability.

3.2. CNN-Based Models

The introduction of Convolutional Neural Networks marked a significant shift toward data-driven feature learning. Architectures such as VGG, ResNet, and DenseNet enabled automatic extraction of hierarchical features directly from MRI data, eliminating the need for handcrafted descriptors.

CNN-based models offer end-to-end learning capabilities, allowing simultaneous feature extraction and classification. This has resulted in substantial improvements in accuracy and robustness compared to traditional methods. However, despite these advantages, CNNs primarily focus on local spatial patterns and often fail to capture long-range dependencies within medical images. This limitation becomes critical in brain tumor analysis, where global context plays a vital role in distinguishing tumor regions from surrounding tissues.

3.3. Emergence of Hybrid Architectures

To overcome the inherent limitations of CNNs, recent research has focused on hybrid architectures that integrate attention mechanisms and Transformer-based models. These approaches combine the strengths of CNNs in local feature extraction with the ability of attention mechanisms to model global dependencies.

Hybridization typically involves incorporating channel and spatial attention modules into CNNs, integrating Vision Transformers for global context modeling, or designing multi-branch architectures that fuse features at different scales. Such models have demonstrated improved performance in terms of classification accuracy, robustness, and interpretability.

This figure presents the evolution of deep learning approaches for brain tumor classification:

- Traditional Machine Learning: Uses handcrafted features and classical classifiers, with limited adaptability.
- CNN-Based Models: Introduce automated feature extraction, improving accuracy through hierarchical learning.
- Hybrid CNN–Attention Architectures: Combine CNNs with attention/Transformer modules to capture both local and global features.

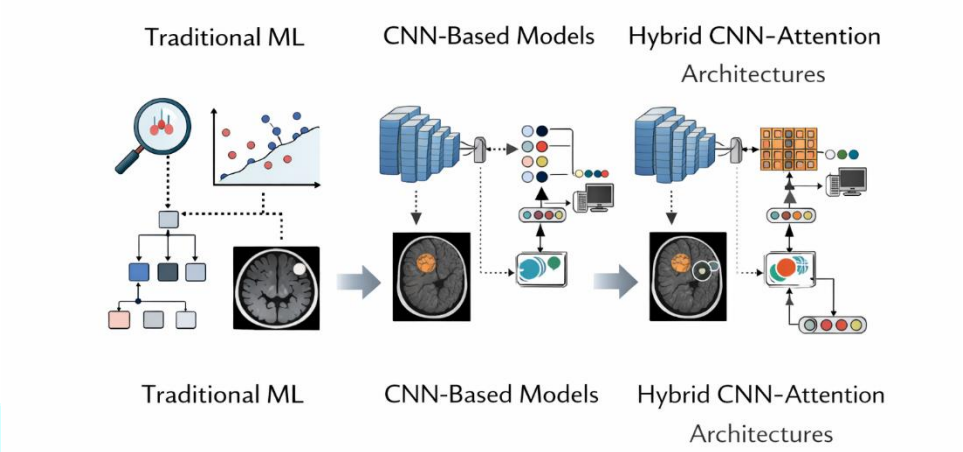

Figure 5: Generic Deep Learning Evolution Pipeline

Overall, the figure shows a clear progression where models become more powerful and accurate as they move from manual feature engineering to hybrid deep learning frameworks.

4. Taxonomy of Hybrid CNN–Attention Architectures

4.1. CNN with Channel and Spatial Attention

Attention mechanisms enhance CNN performance by emphasizing relevant features while suppressing less important information. Channel attention mechanisms, such as Squeeze-and-Excitation (SE) networks, focus on inter-channel relationships, whereas spatial attention mechanisms, such as CBAM, highlight important spatial regions.

The channel attention operation can be expressed as:

$$F' = \sigma\big(W_2 \cdot \delta(W_1 \cdot F)\big) \cdot F \qquad \square\square\square$$

where F denotes the input feature map, δ represents the ReLU activation function, and σσ denotes the sigmoid function. This formulation enables adaptive feature recalibration, improving the representation of tumor-specific regions.

4.2. CNN with Hybrid Attention Mechanisms

Hybrid attention models combine multiple attention strategies, including channel, spatial, and residual attention. This integrated approach allows the model to simultaneously capture "what" features are important and "where" they are located. As a result, these architectures improve feature discrimination and localization accuracy, particularly in complex tumor regions.

4.3. CNN with Vision Transformer (ViT)

Vision Transformers introduce self-attention mechanisms that enable global context modeling. The attention operation is defined as:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \qquad \square2\square$$

This formulation allows the model to capture long-range dependencies across the entire image, addressing one of the primary limitations of CNNs. When combined with CNN-based feature extraction, these hybrid models achieve superior performance in handling tumor heterogeneity and complex spatial relationships.

This figure depicts a two-stage hybrid architecture combining CNN and Transformer modules for brain tumor analysis:

Initially, the MRI image is processed through convolutional layers, which extract low-level and mid-level features such as edges, textures, and tumor boundaries. These layers specialize in capturing localized spatial patterns within the image.

Subsequently, the extracted feature maps are passed to Transformer blocks, which apply self-attention mechanisms to model global dependencies across the entire image. This allows the network to understand relationships between distant regions that CNNs alone cannot capture.

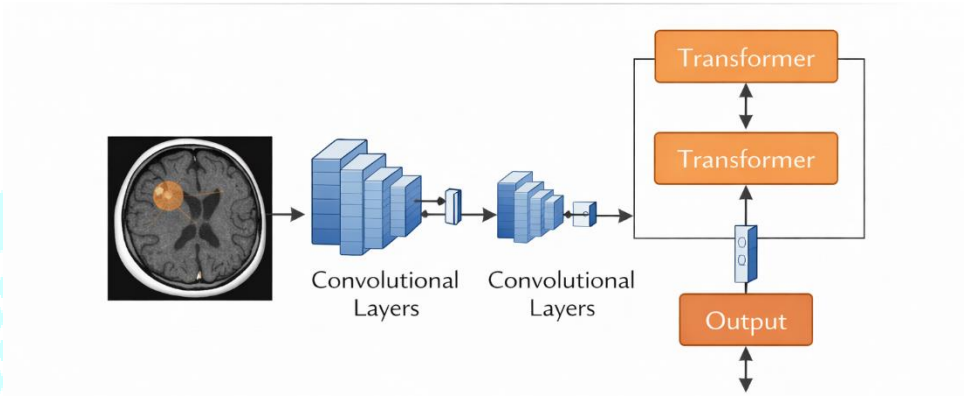

Figure 6: CNN–Transformer Hybrid Architecture

Finally, the integrated features are used to generate the classification output, leveraging both local detail and global context. This architecture effectively combines spatial precision from CNNs with contextual awareness from Transformers, leading to improved diagnostic performance.

4.4. Multi-Branch Hybrid Models

Multi-branch architectures utilize parallel CNN pathways to extract features at different scales, which are then fused using attention mechanisms. This design improves robustness by capturing both fine-grained and high-level features, making it particularly effective for tumors with varying sizes and textures.

4.5. CNN with Attention and Optimization Techniques

Recent studies incorporate optimization strategies such as genetic algorithms and advanced hyperparameter tuning methods into hybrid architectures. These techniques enhance convergence speed and improve generalization performance, especially in scenarios with limited training data.

5. DATASET AND PREPROCESSING STRATEGIES

5.1. Datasets

Commonly used datasets include the Fig share brain tumor dataset, Kaggle MRI datasets, and the BraTS benchmark dataset. These datasets provide diverse imaging conditions and tumor variations, enabling comprehensive evaluation of classification models.

5.2. Preprocessing Pipeline

Preprocessing plays a crucial role in improving model performance. Typical steps include skull stripping to remove non-brain tissues, normalization to standardize intensity values, data augmentation to increase dataset diversity, and region-of-interest (ROI) extraction to focus on tumor regions. These steps collectively enhance feature quality and reduce overfitting.

This figure illustrates the complete MRI preprocessing pipeline used before model training.

Initially, skull stripping removes non-brain tissues (e.g., skull and background), ensuring that only relevant brain regions are retained. Next, normalization standardizes intensity values across images, reducing variations caused by different scanners or acquisition settings.

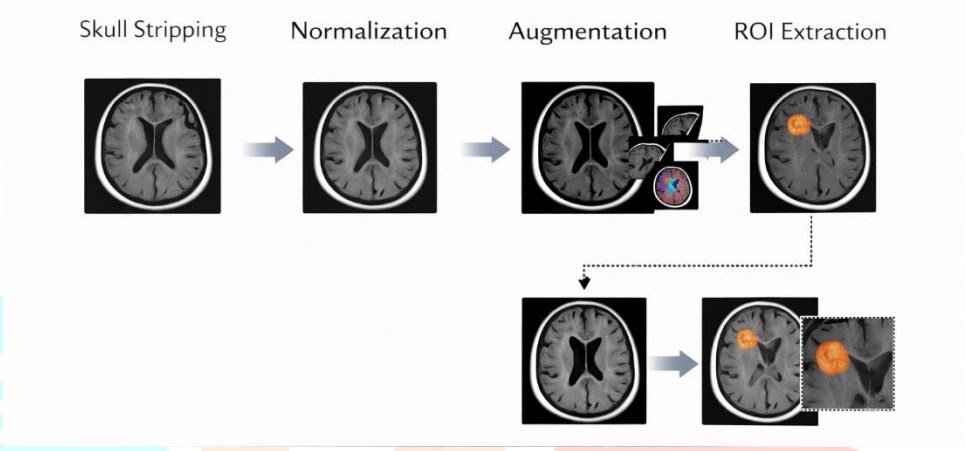

Figure 7: MRI Preprocessing Workflow

The augmentation stage generates additional training samples through transformations such as rotation, flipping, and scaling, which improves model generalization. Finally, ROI (Region of Interest) extraction isolates the tumor region, allowing the model to focus on the most relevant features.

Overall, this pipeline enhances data consistency, robustness, and classification accuracy.

6. PERFORMANCE EVALUATION METRICS

The effectiveness of classification models is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). These metrics provide a comprehensive assessment of model performance, particularly in imbalanced datasets where accuracy alone may be misleading.

This figure presents the ROC curve comparison between CNN and hybrid attention models.

The hybrid attention model consistently lies above the CNN curve, indicating a higher true positive rate for the same false positive rate. This results in a larger area under the curve (AUC), reflecting better overall classification performance.

Overall, the figure demonstrates that hybrid architectures outperform conventional CNNs in accuracy and reliability.

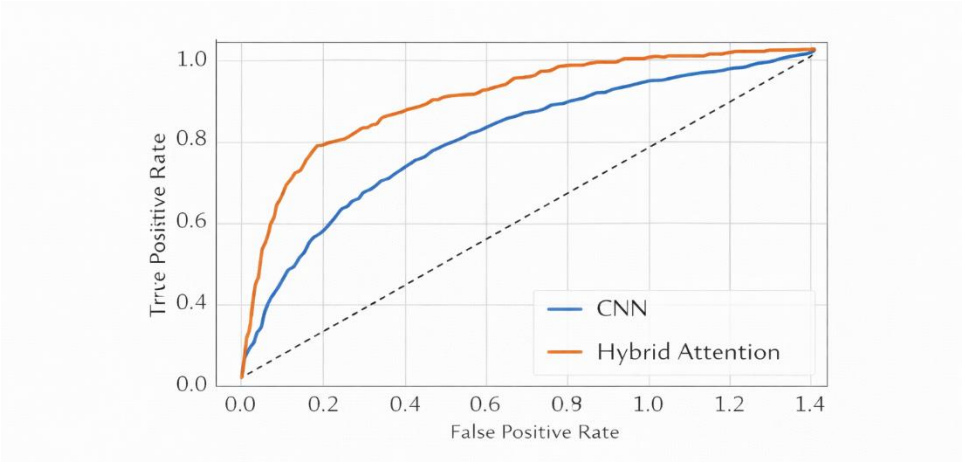

Figure 8: ROC Curve Comparison of Models

7. Comparative Analysis of Hybrid Models

Table 1 compares different architectures for brain tumor classification in terms of performance and complexity. Conventional CNNs provide efficient local feature extraction but are limited by their inability to capture global context, resulting in comparatively lower accuracy.

Incorporating attention mechanisms improves feature refinement and tumor localization, leading to better performance. CNN–ViT hybrid models achieve the highest accuracy by capturing long-range dependencies, although they require significant computational resources. Multi-branch hybrid models further enhance robustness through multi-scale feature fusion but introduce architectural complexity.

Overall, there is a clear trade-off: higher accuracy is achieved with hybrid models at the cost of increased computational and structural complexity.

Table 1 Comparative Analysis of Hybrid Models

Model Type	Key Feature	Accuracy	Strength	Limitation
CNN	Local feature extraction	90–97%	Simplicity	Limited global context
CNN + Attention	Feature refinement	95–98%	Improved localization	Moderate complexity
CNN + ViT	Global modeling	97–99%	High accuracy	Computational cost
Multi-branch hybrid	Feature fusion	98%+	Robustness	Architectural complexity

This figure shows a comparative accuracy analysis of different model architectures.

Traditional machine learning achieves the lowest accuracy, followed by CNN models with improved performance due to automated feature extraction. The inclusion of attention mechanisms further enhances accuracy, while hybrid attention models achieve the highest performance.

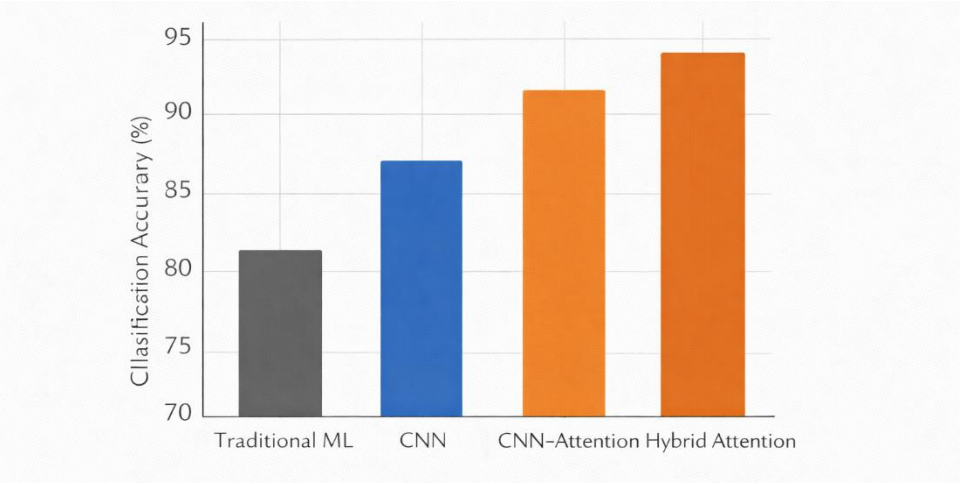

Figure 9: Performance Comparison of Architectures

Overall, the figure highlights a clear trend: as models incorporate attention and hybrid strategies, classification accuracy consistently improves.

8. Critical Discussion

Hybrid CNN–attention architectures demonstrate clear advantages over conventional CNNs by improving feature representation and classification accuracy. The integration of attention mechanisms enhances interpretability through attention maps, enabling better localization of tumor regions.

However, several limitations persist. These models often exhibit reduced generalization when trained on limited datasets and require significant computational resources, particularly in Transformer-based designs. Additionally, dataset bias and lack of standardized evaluation protocols hinder fair comparison across studies. Reproducibility remains a concern due to insufficient reporting of implementation details and limited availability of open-source frameworks.

9. Research Gaps and Future Directions

Despite recent advancements, several research gaps remain. Current models lack sufficient explainability, limiting clinical trust and adoption. Furthermore, most studies rely on single-modality MRI data, ignoring the complementary information available in multi-modal imaging.

Future research should focus on integrating explainable AI techniques such as Grad-CAM and LIME, enabling transparent decision-making. Federated learning offers a promising solution for privacy-preserving model training across distributed medical datasets. Additionally, lightweight architectures are needed for real-time deployment in clinical environments. Multi-modal fusion of MRI sequences and large-scale clinical validation will be essential for translating these models into practical healthcare solutions.

10. Conclusion

Hybrid CNN–attention architectures represent a significant advancement in brain tumor classification by effectively combining local and global feature representations. While these models achieve high accuracy, challenges related to interpretability, generalization, and computational efficiency remain. Addressing these issues is essential for developing clinically reliable and scalable diagnostic systems.

I. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to their respective institution for providing the necessary resources and research environment to carry out this study. The authors also acknowledge the contributions of the research community whose publicly available datasets and prior work have significantly supported this review.

Additionally, we extend our appreciation to peers and reviewers for their valuable insights and constructive feedback, which have helped improve the quality and clarity of this manuscript.

REFERENCES

[1] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," Computers in Biology and Medicine, vol. 111, 2020.

[2] V. R. Parihar, A. Y. Tonge, and P. D. Ganorkar, "Heartbeat and Temperature Monitoring System for Remote Patients using Arduino," International Journal of Advanced Engineering Research and Science (IJAERS), vol. 4, no. 5, pp. 55–58, May 2017.

[3] A. A. Abdusalomov et al., "Brain tumor detection based on deep learning approaches using MRI images," Diagnostics, vol. 13, no. 1, 2023.

[4] P. Tiwari et al., "CNN-based multiclass brain tumor detection using MRI images," Computational Intelligence and Neuroscience, 2022.

[5] M. M. Zahoor et al., "Brain tumor MRI classification using deep residual CNN," Biomedicines, vol. 12, no. 7, 2024.

[6] V. R. Parihar and H. R. Boveiri, "A survey and comparative analysis on image segmentation techniques," in Image Segmentation: A Guide to Image Mining, ICSES, 2018.

[7] M. Ahmed et al., "Hybrid CNN–Transformer framework for brain tumor classification using MRI," Scientific Reports, 2024.

[8] V. R. Parihar, R. S. Nage, and A. S. Dahane, "Image analysis and image mining techniques: A review," Journal of Image Processing and Artificial Intelligence, vol. 3, no. 2, pp. 1–18, 2017.

[9] R. Disci et al., "Advanced deep learning-based brain tumor classification using transfer learning," Healthcare, 2025.

[10] V. R. Parihar, "Image segmentation based on graph theory and threshold," ICSES Transactions on Image Processing and Pattern Recognition (ITIPPR), vol. 4, no. 4, Dec. 2018.

[11] M. Rasheed et al., "Optimized CNN and ResNet101 for multi-class brain tumor classification," Biomedical Signal Processing and Control, 2025.

[12] G. Appasami et al., "Lightweight CNN models for MRI brain tumor classification: A review," Springer, 2025.

[13] L. Ke et al., "Multi-scale channel attention CNN integrated with SVM for brain tumor classification," Scientific Reports, 2026.

[14] Y. Wong et al., "Deep CNN-based multi-class brain tumor classification using MRI datasets," PLOS ONE, 2025.