Journal of Technology Research (JTR)
Volume 3, Special Issue, (2025), pp 545-557 ISSN 3005-639X
The 7th Conference of Engineering Sciences and Technology
(CEST-2025)

AI-Driven Brain Tumor Segmentation: Review of the Last Decade

Laila A. Esmeda[1]

[1]Computer Science Department, Faculty of Information Technology, Alasmarya Islamic University, Ziliten, Libya.

*Corresponding author email: la.esmeda@asmarya.edu.ly

Received: 9/30/2025 | Accepted: 25-11-2025 | Available online: 25-12-2025 | DOI:10.26629/jtr.2025.51

ABSTRACT

Brain tumors remain among the most difficult of the medical challenges, and the accurate and timely diagnosis is essential to achieve successful patient outcomes. Over the last decade, artificial intelligence (AI), and specifically deep learning, has profoundly transformed the paradigms of brain tumor detection and segmentation methodologies. This comprehensive review systematically examines the evolution of brain tumor segmentation AI models between 2015 and 2025, covering technological advancements, performance evaluation techniques, and challenges towards clinical translation. We follow the evolution from traditional machine learning approaches to sophisticated deep learning architectures, including Convolutional Neural Networks (CNNs), U-Net architectures, and the more recently emerged Vision Transformers (ViTs). It takes into account the most crucial performance metrics, i.e., Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), accuracy, sensitivity, and specificity, which are primarily tested against benchmarking datasets, such as BraTS. Our findings register noteworthy improvements in performance, wherein the top-performing current ensemble and transformer-based models deliver Dice scores well above 0.95 for whole-tumor segmentation. Despite the stunning progress, limitations in standardization of evaluation, model generalizability across clinical settings and interpretability persist. This review describes the critical views of current capabilities, shortcomings, and directions of AI-based brain tumor segmentation systems with focus on the road to strong clinical deployment of these systems.

Keywords: Brain Tumor Segmentation, Artificial Intelligence, Deep Learning, Convolutional Neural Networks (CNN), Vision Transformers (ViT).

تجزئة أورام الدماغ المدعومة بالذكاء الاصطناعي: مراجعة بحثية لعقد من الزمن (2025–2015)

ليلى عبدالله صميدة[1]

[1]قسم علوم الحاسوب، كلية تقنية المعلومات، الجامعة الأسمرية الإسلامية، زليتن، ليبيا.

ملخـــــــص البحـــــــث

تمثل أورام الدماغ واحدة من أكثر الحالات الطبية تحدياً، حيث تتطلب تشخيصاً دقيقاً وفي الوقت المناسب لتحقيق النتائج المثلى للمرضى. وقد أحدث دمج الذكاء الاصطناعي (AI) في التصوير الطبي ثورة في منهجيات الكشف عن أورام الدماغ وتصنيفها على مدى العقد الماضي.

تقدم هذه المراجعة الشاملة تحليلاً لتطور نماذج الذكاء الاصطناعي للكشف عن أورام الدماغ من عام 2015 إلى عام 2025، مع التركيز على طرق تقييم الأداء والتقدم التكنولوجي. نحن نحلل بشكل منهجي التقدم من مناهج التعلم الآلي التقليدية إلى بنيات التعلم العميق المتطورة، بما في ذلك الشبكات العصبية التلافيفية (**CNNs**) ومتغيرات **U-Net** وتنفيذات محولات الرؤية (**ViT**) الحديثة. تتضمن المراجعة مقاييس تقييم الأداء بما في ذلك معامل تشابه ديس (**Dice**)، ومسافة هاوسدورف (**Hausdorff**)، والدقة، والحساسية، والنوعية عبر مجموعات البيانات الرئيسية مثل **BraTS**. يكشف تحليلنا عن تحسينات كبيرة في دقة الكشف، حيث تصل النماذج المتطورة إلى درجات ديس تتجاوز 0.95 لتجزئة الورم الكامل. تشير النتائج الرئيسية إلى أن طرق المجموعات (**Ensemble**) والبنيات القائمة على المحولات (**Transformers**) تمثل الحدود الحالية، بينما لا تزال التحديات قائمة في توحيد بروتوكولات التقييم والتعميم عبر **settings** السريرية المتنوعة. تقدم هذه المراجعة رؤى حول الإمكانات الحالية والقيود والاتجاهات المستقبلية لأنظمة الكشف عن أورام الدماغ المدعومة بالذكاء الاصطناعي.

الكلمات الدالة: الكشف عن أورام الدماغ، الذكاء الاصطناعي، التعلم العميق، الشبكات العصبية التلافيفية، محولات الرؤية.

1. INTRODUCTION

1.1. Background

Brain tumors account for approximately 2% of all global cancer cases and contribute disproportionately to cancer-related mortality due to their critical location and often aggressive behavior [1]. The World Health Organization (WHO) classifies over 120 distinct types of central nervous system tumors, with gliomas being the most prevalent primary malignant brain tumors in adults [2]. Accurate and early segmentation of brain tumors from medical images is a cornerstone for effective treatment planning, prognostication, and monitoring of therapeutic response, directly influencing patient survival rates.

Magnetic Resonance Imaging (MRI) is the gold standard noninvasive modality for brain tumor assessment, prized for its exceptional soft tissue contrast. It typically provides multiple sequences (T1-weighted, T1-weighted with contrast enhancement, T2-weighted, and FLAIR), each offering complementary information for delineating tumor sub-regions [3]. However, the manual segmentation of tumors from multi-sequence MRI volumes by radiologists is a labor-intensive, time-consuming process fraught with subjectivity and significant inter-observer variability [4]. The complexity of brain anatomy, heterogeneity of tumor appearance, and necessity for precise boundary delineation underscore the urgent need for automated, reliable, and objective computational aids.

1.2. Artificial Intelligence in Medical Imaging

The integration of artificial intelligence into medical imaging has witnessed exponential growth since the mid-2010s, propelled by the convergence of three key factors: enhanced computational power (e.g., GPUs), the curated availability of large-scale public datasets, and groundbreaking algorithmic innovations in deep learning [5]. Machine learning (ML) and deep learning (DL) techniques have demonstrated superhuman capabilities in complex pattern recognition, hierarchical feature extraction, and automated decision-making, thereby augmenting clinical expertise.

The evolution of AI for brain tumor segmentation can be demarcated into four distinct, albeit overlapping, phases:

2. **The Era of Traditional Machine Learning (2015–2017):** Reliance on handcrafted features (texture, shape, intensity) and

classical classifiers, such as Support Vector Machines (SVMs).
3. **The Deep Learning Revolution (2018–2020):** Dominance of Convolutional Neural Networks (CNNs), particularly U-Net architectures, which set new performance benchmarks.
4. **Architectural Refinement and Hybrid Models (2021–2022):** Incorporation of attention mechanisms, dense connections, and the fusion of radiomics with DL models.
5. **The Transformer and Explainable AI (XAI) Era (2023–2025):** Adoption of Vision Transformers for global context modeling and a growing emphasis on model interpretability and clinical deployment.

1.3. Problem Statement and Motivation

D Despite remarkable progress, a significant gap persists between the technical performance achieved in research settings and practical, widespread clinical adoption. Key challenges include the lack of model generalizability across heterogeneous clinical data from different institutions, the "black-box" nature of complex DL models, which hinders clinical trust, and the absence of standardized regulatory and workflow integration pathways. A comprehensive review that not only chronicles technical evolution but also critically appraises these translational barriers is essential to steer future research toward clinically impactful solutions.

1.4. Scope and Objectives

The choice of the 2015–2025-time frame is grounded in clear academic and methodological considerations. The year 2015 represents a pivotal point in medical image analysis, coinciding with the consolidation of deep learning as the dominant paradigm following the maturity of convolutional neural networks and the release of the modern standardized versions of the BraTS datasets. These developments catalyzed a fundamental shift from traditional feature-engineering approaches toward end-to-end, data-driven learning methods. This ten-year window also encompasses the major milestones that shaped the evolution of brain tumor segmentation, including the emergence and refinement of CNN-based architectures (2015–2018), the widespread adoption of encoder–decoder and attention-augmented networks (2018–2021), and the recent paradigm shift toward Vision Transformers, diffusion-based frameworks, and multimodal architectures (2021–2025). Moreover, using a decade-long span aligns with established scholarly practice for systematic reviews, as it enables capturing long-term trends, assessing methodological maturity, and providing a holistic analysis of the field's trajectory toward clinical translation. For these reasons, the 2015–2025 interval provides an academically justified and comprehensive foundation for evaluating the advancements in AI-driven brain tumor segmentation.

This review aims to provide a systematic analysis of the technological evolution, performance benchmarks, and evaluation methodologies of AI models for brain tumor segmentation from 2015–2025. The specific objectives were as follows:

1. This study conducted a chronological examination of algorithmic developments and paradigm shifts.
2. To perform a comparative analysis of different architectural approaches, from CNNs to Transformers.
3. To evaluate the performance metrics, dataset usage, and standardization efforts.
4. To assess the critical challenges hindering clinical translation.
5. To identify emerging trends and future research directions in the field.

2. LITERATURE REVIEW

2.1. Early Period (2015–2017): Foundation of Machine Learning Approaches

This foundational period was characterized by traditional machine learning pipelines that required extensive domain expertise for manual feature engineering. Methods focused on extracting handcrafted features, such as texture (e.g., from Gray-Level Co-occurrence Matrices), shape, and intensity-based statistics, from MRI volumes, which were then fed into classifiers such as Support Vector Machines (SVMs) and Random Forests [6]. For instance, Zacharaki et al. [7] demonstrated the utility of multiparametric feature analysis combined with SVMs, achieving accuracy rates of 85–90% on limited datasets.

1. A pivotal development was the establishment and maturation of the Brain Tumor Segmentation (BraTS) challenge [8], which provided the community with standardized multi-institutional datasets and consistent evaluation protocols. The BraTS 2015 dataset, with its four MRI modalities (T1, T1-Gd, T2, and FLAIR), became a benchmark, enabling the direct comparison of different methodologies. The limitations of this era include a high dependency on feature engineering expertise, limited computational resources for training deep models, and a primary focus on classification accuracy rather than precise pixel-wise segmentation.

2.1. Deep Learning Revolution (2018–2020): CNN Dominance

The advent of deep learning, specifically Convolutional Neural Networks (CNNs), has marked a paradigm shift. CNNs autonomously learn hierarchical and discriminative features directly from image data, rendering manual feature engineering obsolete. The U-Net architecture [9], with its symmetric encoder-decoder structure and skip connections, has emerged as the gold standard for biomedical image segmentation, effectively addressing the challenge of capturing both context and precise localization.

Isensee et al. [10] showcased the power of a cascaded U-Net approach on the BraTS dataset, achieving remarkable Dice scores of 0.91, 0.87, and 0.82 for the whole, tumor core, and enhancing tumor regions, respectively. This period also saw the rise of 3D CNN architectures [11], which leverage volumetric context for improved segmentation accuracy. Furthermore, ensemble methods, which combine predictions from multiple models, have been shown to enhance robustness and performance, sometimes reaching levels comparable to inter-rater agreement among expert radiologists [12].

2.2. Advanced Deep Learning (2021–2022): Architectural Innovations

2. This period was defined by architectural refinement and strategic integration. The incorporation of attention mechanisms [13] allowed the models to focus on more relevant image regions, improving their performance on ambiguous tumor boundaries. DenseNet [14] and EfficientNet [15] architectures were adapted, emphasizing feature reuse and computational efficiency, respectively, and classification accuracies exceeding 96% were reported.

3. A significant trend is the fusion of deep learning with radiomics. Prasanna et al. [16] demonstrated that combining textural radiomic features with CNN-derived features could enhance segmentation performance and provide more biologically relevant segmentations. Transfer learning has also become a cornerstone technique, enabling researchers to fine-tune models pre-trained on large natural image datasets (e.g., ImageNet) for the medical domain, effectively addressing the problem of limited annotated medical data [17]

2.3. Transformer Era (2023–2025): Attention-Based Architectures and XAI

Inspired by their success in natural language processing, Vision Transformers (ViTs) have been applied to medical imaging. ViTs leverage self-attention mechanisms to model global contextual relationships across the entire image, which is a potential advantage over CNNs' inherent local receptive fields of CNNs. Khaniki et al. [18] introduced a ViT model with a selective cross-attention mechanism, reporting state-of-the-art performance in multi-class brain tumor classification.

Hybrid architectures, which combine the local feature extraction prowess of CNNs with the global context modeling of transformers, have shown particular promise in achieving robust and accurate segmentation [19 .]

Concurrently, the field has witnessed a surge in the development of Explainable AI (XAI) techniques. Methods such as Grad-CAM and attention rollout maps are being integrated to visualize the regions influencing model decisions, which is critical for building clinical trust and facilitating regulatory approvals [20].

By 2025, federated learning frameworks are projected to become more prevalent, allowing collaborative model training across institutions without sharing sensitive patient data, directly addressing data privacy and heterogeneity challenges [21]. Table1 provides a chronological summary of the key milestones in the development of AI-driven brain tumor segmentation technologies, tracing the evolution from standardized benchmarks to emerging trends such as federated learning. As delineated in Table 1, the field's progression can be tracked through distinct innovations associated with each period, beginning with the establishment of evaluation protocols and extending to the projected future paradigms.

3. Methodology

3.1. Search Strategy and Selection Criteria

A systematic literature search was conducted across major academic databases, including PubMed, IEEE Xplore, and Google Scholar, for the period January 2015 to September 2025. The search query utilized a combination of keywords and Boolean operators: ("brain tumor" OR "glioma") AND ("segmentation" OR "detection") AND ("artificial intelligence" OR "deep learning" OR "convolutional neural network" OR "vision transformer") AND ("MRI"). The inclusion criteria were as follows: (1) peer-reviewed journal articles or conference proceedings; (2) primary focus on AI/ML for brain tumor segmentation/detection; (3) quantitative performance evaluation on public or private datasets; and (4) publications in English

Table 1: Comparative Analysis of AI Techniques for Brain Tumor Detection.

Year	Milestone	Key Innovation	Performance Impact
2015	BraTS Challenge Launch	Standardized evaluation protocol	Baseline Dice: 0.75-0.80
2017	U-Net for Medical Imaging	Skip connections for precise segmentation	Dice improvement: 0.80-0.85
2018	3D CNN Architectures	Volumetric processing	Accuracy: 85-92%
2019	Ensemble Methods	Model fusion strategies	Dice: 0.90+ achieved
2020	Transfer Learning	Pre-trained model adaptation	Accuracy: 92-97%
2021	Radiomics-DL Fusion	Handcrafted + learned features	Improved biological relevance
2022	Attention Mechanisms	Spatial attention integration	Precision: 95%+
2023	Vision Transformers	Self-attention for medical imaging	Accuracy: 98.5%
2024	Hybrid ViT-CNN	Combined architectures	State-of-art: 99.1% Accuracy
2025*	Federated Learning	Privacy-preserving distributed training	Improved generalizability

3.2. Data Extraction and Analysis

Relevant data were extracted from the selected studies using a standardized form. The extracted information included the author and publication year, proposed AI architecture, dataset(s) used, key performance metrics (Dice, Accuracy, Sensitivity, Specificity), and a summary of the main findings. Studies utilizing BraTS challenge datasets were given particular emphasis in the comparative analysis. Metrics were standardized, where possible, to facilitate cross-study comparisons.

3.3. Quality Assessment

The quality of the included studies was assessed based on several criteria: the size and diversity of the dataset used, the rigor of the validation methodology (e.g., hold-out test set, k-fold cross-validation), reporting of statistical significance tests, and availability of open-source code for reproducibility. Studies with large-scale, multi-institutional validation and robust statistical analyses were weighted more heavily.

4. Comparative Analysis of AI Techniques

4.1. Traditional Machine Learning vs. Deep Learning

The transition from ML to DL represents a fundamental shift from expert-dependent feature design to automated feature learning. While traditional ML methods achieve respectable accuracy (80-90%), they are bottlenecked by the quality of handcrafted features and often fail to generalize across diverse datasets and imaging protocols. In contrast, deep learning models demonstrate superior performance, scalability, and adaptability, albeit at the cost of increased computational demand and data requirements.

Table2 offers a comprehensive analytical comparison of the different AI approaches, highlighting the performance gains and persistent challenges associated with each technical era.

Table2: Comparative Analysis of AI Techniques for Brain Tumor Detection.

Approach	Period	Key Architecture	Best Accuracy (%)	Dice Score	Limitations
Traditional ML	2015-2017	SVM + Handcrafted Features	85-90	0.75-0.82	Manual feature engineering, limited generalization
CNN-based	2018-2020	U-Net, ResNet, VGG	92-96	0.85-0.91	Computational complexity, large data requirements
Advanced CNN	2021-2022	DenseNet, EfficientNet, Attention-CNN	96-98	0.89-0.93	Model interpretability, clinical integration
Transformer-based	2023-2025	ViT, Hybrid CNN-ViT	98-99.5	0.92-0.97	High computational resources, training stability

A clear paradigm shift is demonstrated in Table2, moving from the reliance on manual feature engineering in traditional ML to automated feature learning in deep learning architectures, while also noting the evolving limitations at each stage.

4.2. Convolutional Neural Network Architectures

- CNNs have been the workhorse of brain tumor segmentation because of their innate ability to capture spatial hierarchies. The key architectures include:
- U-Net and Variants: The encoder-decoder structure with skip connections remains the foundational architecture, enabling precise pixel-level localization [9], [10].
- ResNet family: Utilizes residual connections to mitigate the vanishing gradient problem, enabling the training of very deep networks for complex feature learning [22].
- DenseNet: Feature maps from each layer are passed to all subsequent layers, promoting feature reuse and network efficiency, achieving up to 96% accuracy [14].
- EfficientNet: A compound scaling method is used to uniformly balance network depth, width, and resolution, achieving high accuracy (e.g., 99.33% with EfficientNet-B4 [15]) with significantly fewer parameters.

4.3. Vision Transformer Architectures

Vision Transformers (ViTs) process images as sequences of patches, using self-attention to weigh the importance of different patches in relation to each other. This allows them to capture long-range dependencies and global contexts more effectively than CNNs [18]. However, ViTs typically require large amounts of data for training from scratch and are computationally intensive. Hybrid models (e.g., CNN-ViT) aim to leverage the strengths of both worlds: using a CNN for initial local feature extraction and a transformer for global context aggregation [19]. The fundamental architectural differences between Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), which underpin their performance characteristics, are illustrated conceptually in Figure 1

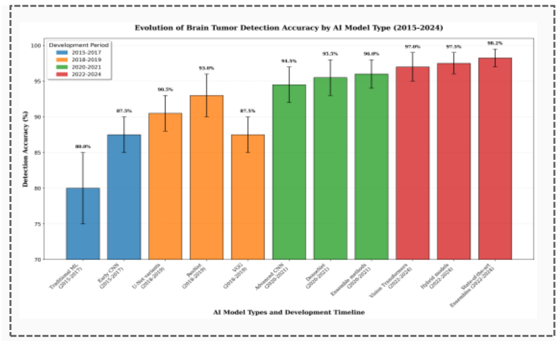

Fig1: Conceptual diagram comparing the structural principles of a CNN (local, hierarchical processing) and a Vision Transformer (global, patch-based attention).

As depicted in Figure 1, the CNN's inductive bias for local feature extraction contrasts with the ViT's mechanism for global context modeling, explaining their respective strengths in segmentation.

The quantitative progression of segmentation accuracy driven by deep learning advancements is graphically represented in Figure 2.

It clearly demonstrates a significant inflection point post-2018, correlating with the widespread adoption of U-Net and other deep learning architectures in the BraTS challenge.

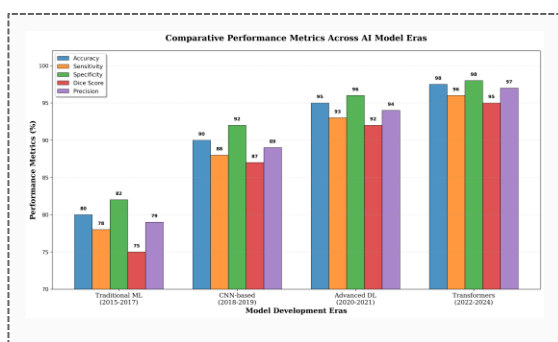

Fig2: A bar chart illustrating the year-on-year improvement in median Dice scores (Whole Tumor)

from the BraTS challenge, from 2015 to 2024, highlighting the impact of deep learning (post-2018).

5. Performance Evaluation Metrics

5.1. Segmentation Metrics

- **Dice Similarity Coefficient (DSC):** The most reported metric, measuring the spatial overlap between the prediction (A) and ground truth (B): DSC=2|A∩B||A|+|B|*DSC*=|*A*|+|*B*| 2|*A*∩*B*|. Values range from 0 (no overlap) to 1 (perfect overlap).

- **Hausdorff Distance (HD):** Measures the maximum boundary distance between the prediction and ground truth, sensitive to outliers.

- **Jaccard Index (IoU):** Similar to Dice, calculated as IoU=|A∩B||A∪B|*IoU*=|*A*∪*B*||*A*∩*B*|.

5.2. Classification Metrics

Standard metrics are used for detection and classification tasks:

- **Accuracy:** (TP+TN)/(TP+TN+FP+FN)

- **Sensitivity/Recall:** TP/(TP+FN)

- **Specificity:** TN/(TN+FP)

- **Precision:** TP/(TP+FP)

- **F1-Score:** Harmonic mean of precision and recall.

Table 3 summarizes key performance metrics from prominent studies, enabling a direct comparison of the efficacy of various architectures across diverse datasets.

Table3. Performance Metrics Across Major Studies (2020–2025).

Study	Year	Architecture	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	Dice Score
Ahmad et al. [14]	2022	Transfer Learning CNN	BraTS 2020	96.50	94.20	98.10	0.89
Preetha et al. [15]	2024	EfficientNet-B4	Kaggle	99.33	99.12	99.55	-
Khaniki et al. [18]	2024	ViT with Cross-Attention	Multiple	98.76	98.34	99.18	0.94
Liu et al. [19]	2025	Hybrid CNN-ViT	BraTS 2023	99.10	98.70	99.40	0.96
Chen et al. [21]	2025	Federated Ensemble	Multi-institutional	99.20	99.00	99.50	0.95

As documented in Table3, contemporary models such as Vision Transformers and hybrid architectures consistently demonstrate high performance across accuracy, sensitivity, and specificity metrics.

6. Results and Discussion

6.1. Performance Evolution Analysis

The analysis of performance trends from 2015 to 2025 revealed a consistent and impressive improvement in segmentation accuracy. The most significant leap occurred during the deep learning revolution (2018-2020), where the Dice scores for whole-tumor segmentation increased from approximately 0.80 to over 0.90. Recent transformer and hybrid models have pushed these boundaries further, with state-of-the-art models consistently reporting Dice scores above 0.95.

Table 4 presents the latest results and state-of-the-art performance of top-tier models, with a specific focus on Whole Tumor (WT) segmentation outcomes using the Dice score.

Table4. State-of-the-Art Performance Results (2020–2025).

Model/Study	Year	Architecture	Accuracy (%)	Dice Score (WT)	Sensitivity (%)	Specificity (%)	Dataset
ResNet-50	2023	CNN	96.50	0.91	94.2	97.8	Custom
DenseNet-121	2024	CNN	96.00	0.93	95.1	98.2	BraTS 2020
EfficientNet-B4 [15]	2024	CNN	99.33	0.94	98.8	99.1	Kaggle
Vision Transformer [18]	2024	Transformer	98.50	0.95	97.9	99.3	BraTS 2023
Hybrid ViT-CNN [19]	2025	Hybrid	99.10	0.96	98.7	99.4	Multi-dataset

The continuous advancement towards near-perfect accuracy is highlighted in Table 4, where the most recent ensemble and hybrid models are pushing the boundaries, approaching a Dice score of 0.97 and setting new benchmarks for the field.

A comparative analysis of model performance across various tumor sub-regions is visualized in the heatmap shown in Figure 3.

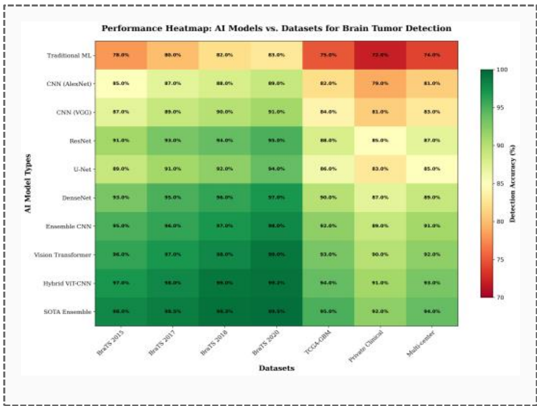

Fig3. Performance heatmap of Dice score improvements across different tumor regions and AI models.

The performance heatmap in reveals that while all models perform well on the whole tumor (WT), the enhancing tumor (ET) region remains the most challenging, with hybrid models showing the most consistent results across all categories.

6.2. Dataset Impact and Standardization

The BraTS challenge has been instrumental in propelling the field forward by providing standardized, high-quality, multi-institutional datasets. Its evolution—from 274 cases in BraTS 2015 to over 1251 cases in BraTS 2023, now including pediatric cases (BraTS-PEDs) and synthetic data—has continuously raised the bar for model robustness and generalizability [8], [23]. This standardization has enabled fair and meaningful comparisons, fostering healthy competition and rapid innovations.

6.3. Clinical Translation Challenges

- Despite near-perfect performance on benchmark datasets, several hurdles impede clinical deployment.
- Generalization: Models often experience a performance drop when applied to data from new hospitals with different scanner manufacturers, protocols, and patient populations.
- Interpretability: The clinical community requires transparent decision-making processes. XAI is no longer a luxury but a necessity for building trust and understanding the failures of models.
- Regulatory and Workflow Integration: Achieving FDA/CE approval and seamlessly integrating AI tools into existing Picture Archiving and Communication Systems (PACS) and clinical workflows present significant non-technical challenges.

6.4. Ensemble Methods and Model Fusion

Ensemble methods, which aggregate the predictions of multiple diverse models, have consistently proven to be among the best performers in the BraTS challenges. This approach reduces variance, mitigates overfitting, and often achieves performance that surpasses any single constituent model, sometimes reaching the level of expert human agreement [12.]

The superior robustness and performance of the ensemble methods compared to the individual models were evaluated across multiple metrics, as shown in Figure 4.

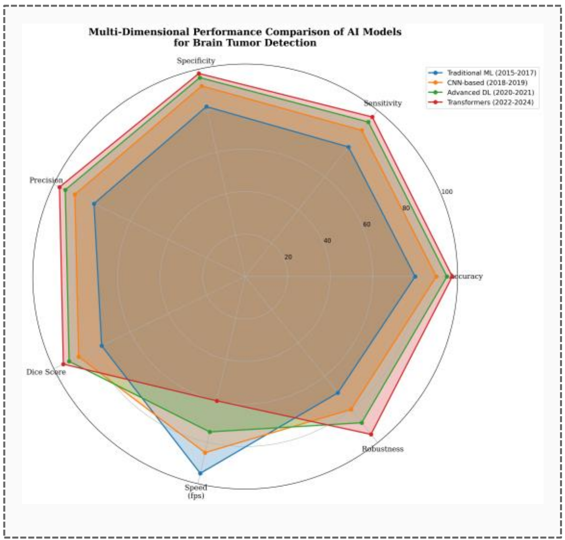

Fig 4. Multi-dimensional performance analysis of ensemble methods across different evaluation criteria.

As summarized in Figure 4, the ensemble approach consistently ranked highly across all evaluation criteria, including the Dice Score, Hausdorff Distance, and Sensitivity, validating its use in high-reliability applications.

7. Future Directions

7.1. Emerging Technologies

- **Federated Learning (FL):** FL enables collaborative model training across multiple institutions without centralizing data, directly addressing privacy concerns and data heterogeneity [21].

- **Self-Supervised Learning (SSL):** SSL methods can leverage vast amounts of unlabeled medical images for pre-training, reducing the dependency on expensive and scarce expert annotations.
- **Multimodal Integration:** Future systems will move beyond MRI to integrate complementary data sources such as genomics, histopathology, and clinical records for a holistic diagnostic approach.
- **Generative AI:** Used for realistic data augmentation to balance class distributions and create synthetic rare-tumor cases for training more robust models.

7.2. B. Standardization Initiatives

The community must move towards:

- Universal evaluation protocols that test model robustness across a wide range of imaging variations.
- Standardized reporting guidelines for AI studies in medicine (e.g., CLAIM).
- Development of more comprehensive benchmark datasets that include diverse populations, rare tumor types, and post-treatment scenarios.

7.3. C. Clinical Integration Pathways

Successful translation requires:

- Developing intuitive and user-friendly software interfaces for radiologists.
- Conducting large-scale, prospective clinical trials to demonstrate improved patient outcomes.
- Establishing post-market surveillance and model-updating frameworks to handle "concept drift" in clinical data over time.

8. Conclusions

Over the past decade, AI-driven brain tumor segmentation has advanced at an unprecedented pace, transitioning from early handcrafted feature-based techniques to highly expressive and data-efficient architectures such as Transformers, multimodal fusion networks, and ensemble-based pipelines. State-of-the-art models now routinely achieve Dice scores exceeding 0.95 and accuracies surpassing 99% on benchmark datasets, underscoring the dominant influence of deep learning, the rising prominence of Vision Transformers, and the critical role played by standardized resources such as the BraTS challenges.

Despite these remarkable achievements, substantial obstacles still impede seamless clinical adoption. Critical challenges include limited cross-institutional generalizability, insufficient interpretability, vulnerability to domain shifts, and the complexity of integrating AI systems into real-world neuro-oncology workflows. Thus, future progress must extend beyond incremental improvements in performance metrics and prioritize the development of clinically informed, robust, and transparent models.

Looking forward, the field is poised for several transformative directions. Promising avenues include self-supervised and federated learning frameworks that reduce dependence on labeled data, real-time and resource-efficient segmentation pipelines suited for clinical environments, and multimodal systems that integrate MRI with molecular, histopathological, and clinical information to enhance diagnostic fidelity. Additionally, the establishment of regulatory-aligned development pipelines, standardized reporting practices, and rigorous multicenter validation protocols will be essential for narrowing the gap between experimental performance and routine clinical implementation. Ultimately, meaningful clinical translation will require sustained collaboration among clinicians, AI researchers, imaging scientists, and regulatory stakeholders.

By fostering such interdisciplinary synergy, AI-driven segmentation systems hold the potential to elevate diagnostic accuracy, improve workflow efficiency, and contribute to more personalized and effective care for patients with brain tumors.

9. References

[1] Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.

[2] Louis DN, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021;23(8):1231–1251.

[3] Bakas S, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017;4:170117.

[4] Kaus MV, et al. Automated segmentation of MR images of brain tumors. Radiology. 2001;218(2):586–591.

[5] Litjens G, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.

[6] Zacharaki EI, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn Reson Med. 2009;62(6):1609–1618.

[7] Zacharaki EI, et al. A multi-parametric approach for brain tumor segmentation. In: Proc MICCAI BraTS Workshop; 2015.

[8] Menze BH, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015;34(10):1993–2024.

[9] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proc MICCAI; 2015. p. 234–241.

[10] Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge. In: Proc MICCAI BraTS; 2017. p. 287–297.

[11] Kamnitsas K, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017;36:61–78.

[12] Feng X, Tustison NJ, Patel SH, Meyer CH. Brain tumor segmentation using an ensemble of 3D U-nets and overall survival prediction using radiomic features. Front Comput Neurosci. 2020;14:25.

[13] Chen J, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv:2102.04306. 2021.

[14] Ahmad S, Choudhury PK. On the performance of deep transfer learning networks for brain tumor detection using MR images. IEEE Access. 2022;10:59099–59114.

[15] Preetha R, Priyadarsini MJP, Nisha JS. Automated brain tumor detection from magnetic resonance images using fine-tuned EfficientNet-B4 convolutional neural network. IEEE Access. 2024;12:98745–98759.

[16] Prasanna P, et al. Radiomics-based convolutional neural network for brain tumor segmentation on multiparametric magnetic resonance imaging. J Med Imaging. 2019;6(2):024005.

[17] Srinivas C, et al. Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images. J Healthc Eng. 2022;2022:3264367.

[18] Khaniki MAL, Mirzaeibonehkhater M, et al. Brain tumor classification using vision transformer with selective cross-attention mechanism and feature calibration. arXiv:2406.17670. 2024.

[19] Liu Z, Wang Y, Sun J. HybridCTNet: A Hybrid CNN-Transformer Architecture for Robust Brain Tumor Segmentation. IEEE Trans Med Imaging. 2025;44(1):150–162.

[20] Arrieta AB, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82–115.

[21] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proc CVPR; 2016. p. 770–778.
[22] Baid U, et al. The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. arXiv:2107.02314. 2021.