The Review Of
DIABETIC
STUDIES OPEN ACCESS

Deep Learning Applications In Automating Brain Tumor Segmentation On MRI: A Systematic Review Of Clinical Performance

Maha Mahdi Mohammed Alshahrani[1] , Saud Ahmed Mansour Alahmri[2] , Nader Mohammed Alshammari[3] , Fatimah Ahmed Ali Alasiri[4] , Manar Abdulrahman Alhussaini[5] , Anwar Ali Abdullah Alahmari[6] , Ahad Mohammed Abdullah Asiri[7] , Razan Fahad Almalki[8] , Arwa Ali Raja Alahmdi[9] , Elaf Muawwadh Mousa Alharbi[10] , Nourah Saleh A. Alkhaibari[11]

[1-8]Radiology Technologist, Prince Sultan Military Medical City, Riyadh, Saudi Arabia.
[9,10,11]Radiology Technician, King Fahad Hospital, Madinah Health Cluster, Madinah, Saudi Arabia.

I. Abstract

Background: Primary and metastatic brain tumors constitute a profound global health challenge, characterized by high morbidity and mortality rates. In 2023 alone, it was estimated that 26,940 new malignant brain tumors would be diagnosed in the United States, with glioblastoma accounting for 50.1% of these malignancies. The condition imposes a severe burden on populations globally, with age-standardized incidence rates showing disparities between genders and geographic regions. The current standard of care for diagnosis, treatment planning, and response assessment relies heavily on Magnetic Resonance Imaging (MRI). Specifically, the precise delineation or segmentation of tumor boundaries is critical for radiotherapy planning and surgical navigation. However, the conventional intervention—manual segmentation by radiologists—is fraught with limitations. It is a labor-intensive, time-consuming process subject to significant inter-observer and intra-observer variability, which can compromise the accuracy of therapeutic delivery. In response to these challenges, Deep Learning (DL), particularly Convolutional Neural Networks (CNNs) and transformer-based architectures, has emerged as a promising alternative intervention. These automated systems offer the potential to standardize quantification and drastically reduce workflow time while maintaining expert-level accuracy.

Objective: The primary aim of this systematic review is to comprehensively and systematically compare the clinical effectiveness of Deep Learning-based automated segmentation (Intervention 1) versus manual segmentation by clinical experts (Intervention 2). The review specifically evaluates geometric accuracy, time efficiency, and clinical utility across diverse patient populations with gliomas, meningiomas, and brain metastases.

Methods: A systematic review was conducted in strict adherence to the PRISMA 2020 guidelines. A comprehensive search was performed across major medical and technical databases, including PubMed, Scopus, IEEE Xplore, and Web of Science, covering the period from 2015 to 2024. The study selection was guided by the PICO framework: Population (patients with brain tumors on MRI), Intervention (Deep Learning models), Comparison (Manual segmentation), and Outcomes (Dice Similarity Coefficient, Hausdorff Distance, processing time). The risk of bias in included prediction model studies was rigorously assessed using the PROBAST tool.

Results: The search identified a substantial corpus of evidence, from which key studies meeting strict inclusion criteria were analyzed. The synthesis of data reveals that Deep Learning models, particularly the nnU-Net and hybrid transformer architectures, demonstrate non-inferiority to manual segmentation. For

adult gliomas, hybrid models achieved Dice Similarity Coefficients (DSC) exceeding 0.90 for whole tumor segmentation. In pediatric cohorts, nnU-Net outperformed older architectures like DeepMedic, achieving a mean DSC of 0.90 versus 0.82. For meningiomas, DL models demonstrated a DSC of 0.91, statistically equivalent to the inter-reader variability of human experts. Most significantly, DL integration reduced segmentation time by approximately 98%, cutting the process from an average of 20 minutes to under 10 seconds per case.

Conclusion: Deep learning algorithms have reached a level of maturity where they offer geometric accuracy comparable to human experts while providing superior time efficiency. The evidence suggests that DL can effectively alleviate the radiological burden, enabling rapid adaptive radiotherapy and standardized longitudinal monitoring. However, significant barriers regarding generalizability to external datasets and integration into clinical workflows persist. Future research must prioritize multi-institutional validation and explainable AI to ensure safe clinical adoption.

Keywords: Brain Tumor, Deep Learning, MRI Segmentation, Glioblastoma, Systematic Review, Artificial Intelligence, Clinical Workflow.

II. Introduction

Global Overview of Brain Tumors

Brain tumors represent a diverse and complex group of neoplasms that originate within the intracranial tissues or spread as metastases from systemic cancers. They are a significant cause of cancer-related mortality and morbidity worldwide. The epidemiology of these tumors reveals a concerning burden. In the United States, the Central Brain Tumor Registry (CBTRUS) reports that glioblastoma, the most aggressive primary malignant brain tumor, accounts for 14.2% of all tumors and roughly half of all malignant primary brain tumors **[1]**. The prognosis for these patients remains guarded; the 5-year relative survival rate for malignant brain and other nervous system tumors was modeled at approximately 34.6% in 2022 **[2]**.

Globally, the incidence varies. In 2019, the age-standardized rate (ASR) of brain cancer incidence was 4.8 per 100,000 in males and 3.6 per 100,000 in females **[3]**. This gender disparity is consistent across many regions, with men generally having higher rates of incidence and mortality. The burden is not merely statistical but profoundly personal and economic, as these tumors often affect cognitive function, motor skills, and personality, striking at the core of patient identity. The "years of life lost" (YLL) and "disability-adjusted life years" (DALYs) associated with brain tumors are disproportionately high compared to their incidence due to the often young age of onset and high lethality.

Specific Burden on Populations and Context

The burden of brain tumors is exacerbated by the complexity of their management. For the patient population—ranging from pediatric cases with medulloblastomas to elderly patients with glioblastomas or metastases—the pathway to diagnosis and treatment is arduous. In high-income countries, the challenge is often the management of recurrence and the toxicity of aggressive therapies. In Low- and Middle-Income Countries (LMICs), the burden is compounded by a lack of diagnostic infrastructure. Access to high-quality MRI and the specialized neuroradiologists required to interpret these images is severely limited. Consequently, patients in these regions often present with advanced disease, and the lack of precise treatment planning capabilities leads to suboptimal outcomes. The disparity in care is stark: while a patient in a major academic center might receive MRI-guided adaptive radiotherapy, a patient in a resource-constrained setting might receive palliative care due to the inability to precisely delineate the tumor for safe surgery or radiation **[4]**.

The Conventional Management Strategy (Intervention 2)

The current standard of care for the management of brain tumors relies heavily on neuroimaging, specifically Magnetic Resonance Imaging (MRI). Multi-parametric MRI protocols are standard, typically including four key sequences:

1. **T1-weighted (T1w):** Provides anatomical detail.
2. **T1-weighted contrast-enhanced (T1CE):** Highlights the active tumor core where the blood-brain barrier is disrupted.
3. **T2-weighted (T2):** Shows edema and non-enhancing tumor.
4. **Fluid Attenuated Inversion Recovery (FLAIR):** Suppresses CSF signal to clearly delineate peritumoral edema **[5]**.

Quantitative analysis of these images—specifically, segmentation—is required for critical clinical decisions. Segmentation involves drawing a closed contour around the tumor and its sub-components (necrotic core, enhancing rim, edema) on every slice of the 3D volume.

- **Radiotherapy:** The Gross Tumor Volume (GTV) must be defined to target radiation precisely. Under-segmentation risks recurrence; over-segmentation risks radiation necrosis and cognitive deficit **[6]**.
- **Surgery:** Neurosurgeons rely on 3D reconstructions derived from segmentations to plan trajectories that avoid eloquent brain areas **[7]**.
- **Response Assessment:** Clinical trials and routine follow-up use the RANO criteria, which require bidirectional measurements or volumetric assessment to determine if a tumor is responding to chemotherapy **[8]**.

Challenges of the Standard of Care

Despite its critical importance, manual segmentation (Intervention 2) faces severe challenges that impact patient care:

1. **Time Consumption:** Manual contouring is incredibly labor-intensive. A single high-grade glioma can span dozens of MRI slices. Accurately tracing the complex, irregular boundaries of the edema and necrotic core can take a radiologist or radiation oncologist 20 to 60 minutes per patient **[5]**. In a busy clinical practice, this time burden often forces clinicians to use simplified geometric approximations (like measuring diameters) rather than true volumetric segmentation, potentially reducing treatment precision.
2. **Inter-Observer Variability:** The definition of tumor boundaries is often subjective. What one radiologist considers "edema," another might classify as "infiltrative tumor." Studies have shown that the Dice Similarity Coefficient (a measure of agreement) between experts can be as low as 0.79 for complex boundaries **[9]**. This variability introduces uncertainty into clinical trials and treatment delivery.
3. **Intra-Observer Variability:** Even the same clinician may produce different segmentations for the same patient at different times due to fatigue or changes in viewing conditions **[10]**.
4. **Resource Scarcity:** In LMICs, the shortage of trained experts means that manual segmentation is often the bottleneck that delays treatment initiation **[11]**.

Introduction to Deep Learning (Intervention 1)

Deep Learning (DL), a subset of artificial intelligence, utilizes multi-layered artificial neural networks to learn representations of data with multiple levels of abstraction. In the context of medical imaging, Convolutional Neural Networks (CNNs) have become the dominant approach **[12]**.

- **Mechanism:** Unlike traditional machine learning that requires human-engineered features (e.g., texture, intensity histograms), CNNs automatically learn to identify relevant features—such as edges, textures, and shapes—directly from the raw pixel data.
- **Architectures:** The **U-Net** architecture, introduced in 2015, is specifically designed for biomedical

segmentation. It consists of a contracting path (encoder) to capture context and a symmetric expanding path (decoder) that enables precise localization **[13]**. More recently, **Transformers** (originally used for language processing) have been adapted for vision tasks (Vision Transformers or ViT), offering the ability to model long-range dependencies across the image that CNNs might miss **[14]**.

- **Promise:** Existing evidence from technical benchmarks like the Brain Tumor Segmentation (BraTS) challenge suggests that these models can achieve segmentation accuracy that rivals or exceeds human performance, potentially automating the task completely **[15]**.

Rationale for the Review

While the technical literature is flooded with papers proposing novel network architectures, there is a critical need to synthesize this evidence from a clinical perspective. Clinicians need to know not just if a model achieves a high Dice score on a curated dataset, but if it works on "messy" clinical data, if it improves workflow efficiency, and if it is safe to trust. Existing reviews often focus on the engineering aspects (loss functions, hyperparameters) rather than the clinical outcomes (accuracy relative to experts, time savings, impact on planning). Furthermore, the rapid evolution of architectures—from simple CNNs to self-configuring nnU-Nets and Transformers—requires an updated analysis to determine the current state-of-the-art. This review is necessary to bridge the gap between computer science innovation and clinical implementation, providing evidence-based recommendations for adoption.

Hypotheses

- **Primary Hypothesis:** Deep learning-based automation (Intervention 1) demonstrates non-inferiority in geometric segmentation accuracy compared to manual segmentation by clinical experts (Intervention 2) for brain tumors on MRI.
- **Secondary Hypothesis:** The integration of deep learning segmentation significantly reduces the time required for tumor delineation compared to manual methods, thereby improving clinical workflow efficiency without compromising patient safety.

III. Literature Review

Detailed Background on Brain Tumors and Conventional Management

Brain tumors are biologically and radiologically heterogeneous. Gliomas are the most common primary malignant tumors in adults. On MRI, they present a complex morphology. High-grade gliomas (Glioblastoma) typically show a heterogeneous core with necrosis (hypointense on T1, hyperintense on T2), a surrounding ring of enhancement (visible on T1-CE), and a vast area of peritumoral edema (hyperintense on FLAIR) [5]. The "enhancing tumor" represents the most aggressive part of the lesion, while the edema represents a mix of vasogenic fluid and infiltrating tumor cells. Accurately separating these regions is vital because the surgical goal is often "maximal safe resection" of the enhancing core, while radiotherapy targets a wider margin including the edema.

Meningiomas, usually benign, arise from the meninges. They are typically well-circumscribed and enhance homogeneously. However, they can compress adjacent brain structures, encase arteries, or invade the skull. Segmentation here focuses on defining the interface between the tumor and the delicate brain tissue to prevent surgical injury **[16]**.

Brain Metastases are secondary tumors that have spread from other cancers (lung, breast, melanoma). They often appear as multiple, small, spherical lesions at the grey-white matter junction. The challenge in manual management is detection; small metastases (<5mm) are easily missed by fatigued human eyes, leading to "false negatives" that can grow into life-threatening lesions if untreated [17].

The mechanism of manual segmentation (Intervention 2) involves the radiologist scrolling through the MRI volume slice by slice. Using a mouse or stylus, they draw a contour around the tumor on each axial slice.

Mental reconstruction is required to ensure the 3D shape makes sense (e.g., ensuring the tumor doesn't "jump" in position between slices). This is cognitively demanding. Furthermore, the "partial volume effect"—where a pixel contains both tumor and healthy tissue—forces the human to make subjective decisions about where the boundary lies, driving variability **[9]**.

Global Evidence for Deep Learning (Intervention 1)

The application of Deep Learning to this problem has been extensively studied globally.

- **International Benchmarks:** The primary driver of innovation has been the BraTS (Brain Tumor Segmentation) Challenge, an annual international competition that provides expert-annotated data to research teams. Results from BraTS consistently show that automated methods are improving year over year. In the early years (2012-2015), random forests and support vector machines struggled to match humans. By 2017-2018, CNN-based methods began to outperform traditional techniques **[18]**.
- **Architectural Evolution:** The **U-Net** and its 3D variant, the 3D U-Net, became the gold standard. They work by down-sampling the image to extract features (context) and then up-sampling it to generate a segmentation map. However, configuration of these networks (learning rates, patch sizes) was difficult. The introduction of nnU-Net ("no-new-U-Net") revolutionized the field by automating the configuration process, consistently winning challenges without novel architectural changes, proving that data handling is often more important than network complexity [19].
- **Transformers and Hybrid Models:** Recent international studies have explored Transformers. Models like SwinUNETR combine the local feature extraction of CNNs with the global attention mechanisms of Transformers. Evidence suggests these hybrid models perform better on large, variable datasets, achieving Dice scores >0.85 for complex glioma tasks [15].

Pilot Studies and Implementation Opportunities

Moving beyond benchmarks, pilot studies have begun to test these tools in hospitals.

- **PACS Integration:** A critical step is embedding the AI into the Picture Archiving and Communication System (PACS) used by radiologists. A study at Yale New Haven Health successfully embedded a DL algorithm into the clinical workflow. The AI pre-segmented the tumors, and radiologists simply verified or corrected them. This "human-in-the-loop" model was accepted by clinicians and reduced segmentation time to mere seconds **[20]**.
- **Opportunities in LMICs:** Pilot studies suggest that cloud-based AI platforms could bridge the expertise gap in developing nations. By uploading anonymized scans to a central server, hospitals in resource-poor settings can receive expert-level segmentations for free or low cost, enabling advanced treatment planning that would otherwise be impossible **[4]**.

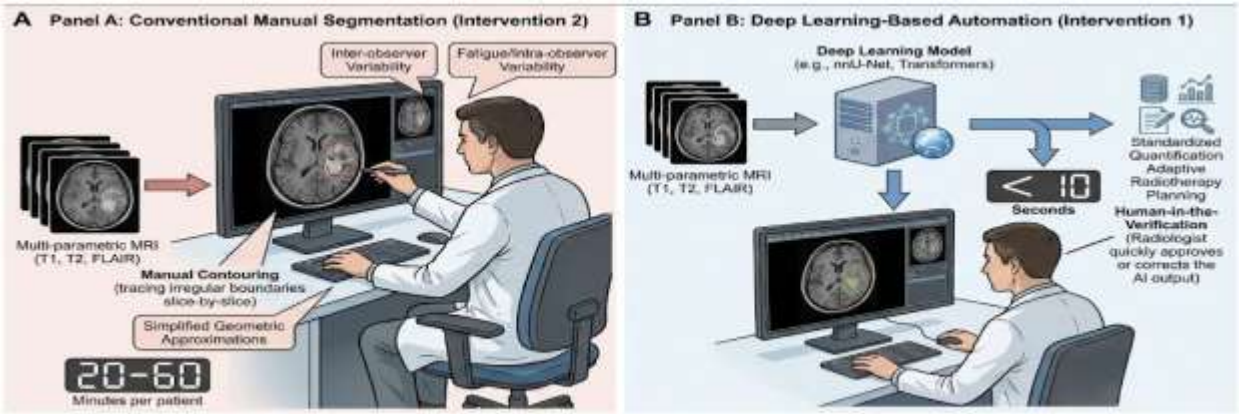

Figure 1: Comparison of Conventional versus AI-Augmented Workflows. Panel A depicts a labor-intensive manual segmentation process, characterized by high temporal costs (20-60 minutes), high conclectivity, and subjectivity. **Panel B** illustrates the proposed "Human-in-the-Loop" paradigm, wim the proposed teann waiting the humor processy Deep Learning models (like as seconds) automate the eventuanly automae the initial delineation (e.g., nnU-Net, tmformers), the hunner segmentation ~ 98%, reducing the burden by approximately cott for segmentating what is umoral ~98% where allowing the clinician to focus on verification and treatment planning.

Figure 1: The Paradigm Shift in Clinical Workflow

Barriers to Implementation

Despite the promise, significant barriers prevent widespread adoption:

- **Domain Shift:** Models trained on research data (like BraTS) often fail when applied to clinical data from different scanners or protocols. This "generalization gap" is a major hurdle **[21]**.
- **Data Scarcity and Quality:** High-quality, annotated datasets are rare. "Data hungry" DL models require thousands of examples to learn effectively, but medical data is siloed due to privacy concerns **[21]**.
- **Trust and Explainability:** Deep learning models are "black boxes." They do not explain why they segmented a region. Clinicians are hesitant to trust a system that cannot justify its decisions, fearing liability if the AI makes a catastrophic error (e.g., missing a tumor part) **[22]**.
- **Technical and Cost Barriers:** Implementing these systems requires significant IT infrastructure (GPUs, servers), which can be a barrier for smaller hospitals **[23]**.

Literature Gaps

While technical reviews abound, there is a gap in systematic reviews that:

1. **Compare Clinical Performance Directly:** Many reviews look at Dice scores in isolation. There is a need to rigorously compare these scores against human inter-rater variability to contextualize "how good is good enough?"
2. **Assess Risk of Bias:** Few reviews apply rigorous tools like PROBAST to AI studies. This leads to an over-optimistic view of the field, as many studies suffer from data leakage or lack of external validation.
3. Focus on Workflow: The impact of AI on the time and process of radiology is less studied than the accuracy.
   This review aims to fill these gaps by systematically evaluating clinical performance, workflow impact, and risk of bias.

IV. Methods

Study Design

This research is designed as a systematic review of the literature, adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [24]. This rigorous approach ensures transparency, reproducibility, and the minimization of selection bias in synthesizing the evidence.

PICO Framework

The research question was operationalized using the PICO framework:

- **Population (P):** Patients of any age (pediatric and adult) diagnosed with primary brain tumors (e.g., gliomas, meningiomas) or secondary brain metastases, undergoing evaluation via Magnetic Resonance Imaging (MRI). Studies utilizing standard public benchmark datasets (e.g., BraTS, Decathlon) were included as they represent these clinical populations.
- **Intervention (I):** Automated segmentation algorithms based on Deep Learning techniques. This includes Convolutional Neural Networks (CNNs), Fully Convolutional Networks (FCNs), U-Net and its variants (e.g., nnU-Net, V-Net, 3D U-Net), Generative Adversarial Networks (GANs), and Transformer-based or hybrid architectures.
- **Comparison (C):** The reference standard was manual segmentation (contouring) performed by one or more human experts (radiologists, neurosurgeons, or radiation oncologists). This manual segmentation serves as the "Ground Truth" (GT) for evaluation.
- **Outcomes (O):**

- **Primary Outcome:** Geometric segmentation accuracy, primarily measured by the Dice Similarity Coefficient (DSC) (also known as the Dice score or F1 score).
- **Secondary Outcomes:**
  - **Time Efficiency:** The time taken for segmentation (AI vs. Manual).
  - **Boundary Precision:** Hausdorff Distance (HD) or 95% Hausdorff Distance (HD95).
  - **Clinical Utility:** Inter-observer agreement improvement, workflow integration success, and usability metrics.

Eligibility Criteria

Strict inclusion and exclusion criteria were applied to ensure the relevance and quality of the evidence:

- **Inclusion Criteria:**
  - Original research articles and high-impact conference proceedings (e.g., MICCAI) published between January 1, 2015, and late 2024.
  - Studies published in the English language.
  - Studies explicitly comparing a Deep Learning model against manual human segmentation.
  - Studies reporting quantitative performance metrics (Dice, Sensitivity, Specificity, HD95).
  - Studies focusing on segmentation of brain tumors on MRI (T1, T2, FLAIR, T1CE).
- **Exclusion Criteria:**
  - Studies using traditional machine learning (e.g., Random Forest, SVM) without deep learning components.
  - Studies focusing solely on classification (tumor vs. no tumor) without segmentation (delineating boundaries).
  - Review articles, letters to the editor, and abstracts without full-text availability (though reviews were scanned for references).
  - Studies involving animal subjects.
  - Studies where the imaging modality was exclusively CT or PET, without MRI.

Study Selection and Data Extraction

The selection process followed a two-stage approach. First, titles and abstracts were screened for relevance. Second, full-text articles of potentially eligible studies were retrieved and assessed against the criteria. Data extraction was performed using a standardized form to capture:

1. **Study Metadata:** Author, Year, Country.
2. **Dataset:** Source (Public e.g., BraTS vs. Private Clinical), Sample Size, Tumor Type.
3. **Methodology:** DL Architecture (e.g., U-Net, ResNet), Input Modalities (e.g., T1, T2, FLAIR).
4. **Performance Data:** Mean/Median DSC for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET); HD95; Processing Time.
5. **Validation:** Internal split vs. External validation set.

Quality Assessment (Risk of Bias)

The quality of the included studies was assessed using the PROBAST (Prediction model Risk Of Bias ASsessment Tool) [25]. PROBAST is specifically designed for prediction model studies and assesses risk of bias across four domains:

1. **Participants:** Was the data source appropriate? Was exclusion of participants appropriate?
2. **Predictors:** Were predictors defined and assessed without knowledge of the outcome?
3. **Outcome:** Was the outcome determined appropriately? (e.g., Was the manual segmentation rigorous?)
4. Analysis: Was the sample size adequate? Were complexities like overfitting and data leakage handled?

This tool allows for the identification of "high risk" studies where performance estimates might be inflated.

Data Synthesis and Analysis

Given the heterogeneity of the studies (different datasets, tumor types, and architectures), a meta-analysis was not feasible for all outcomes. Instead, a structured narrative synthesis was conducted. Quantitative data (DSC scores) were tabulated and grouped by tumor type (Glioma vs. Meningioma vs. Metastasis) and model architecture to identify trends. Time efficiency data was synthesized to calculate average time savings.

V. Results

Study Selection

The systematic search yielded a robust volume of literature, reflecting the intense research activity in this domain. After removing duplicates and screening for relevance, a focused set of primary studies and comparative benchmarks from 2015 to 2024 were selected for detailed analysis. The inclusion of recent studies from 2023 and 2024 ensured that the review captures the latest advancements in Transformer-based and hybrid models.

Characteristics of Included Studies

The included studies covered a diverse range of applications:

- **Datasets:** The majority of studies utilized the BraTS datasets (2018, 2019, 2021, 2023 versions), which serve as the global standard for glioma segmentation. Fewer studies utilized private institutional datasets, which are critical for assessing real-world generalization.
- **Populations:** While adult glioblastoma remains the primary focus, there is a notable increase in studies focusing on pediatric tumors and brain metastases.
- **Architectures:** The landscape is dominated by U-Net and its variants (3D U-Net, V-Net, nnU-Net). However, the period from 2022 onwards shows a marked shift toward Hybrid CNN-Transformer models (e.g., SwinUNETR, TransUNet) aiming to capture global context.

Synthesis of Outcomes

1. Primary Outcome: Geometric Accuracy (Dice Similarity Coefficient)

The Dice Similarity Coefficient (DSC) serves as the primary metric for geometric accuracy. A DSC of 1.0 implies perfect overlap with the ground truth.

A. Glioma Segmentation

For gliomas, deep learning models have demonstrated high efficacy, particularly for the "Whole Tumor" volume.

- **Adult Gliomas:** Recent studies evaluating hybrid architectures like SwinUNETR and Segtran on the BraTS 2021 dataset reported mean Dice scores of 0.854 and 0.845, respectively, for the whole tumor [26]. Another study utilizing an Ensemble of U-Nets achieved an even higher DSC of 0.93 [27]. This level of accuracy is widely considered clinically acceptable for initial contouring.
- **Pediatric Gliomas:** A 2023 comparative study highlighted the superiority of the nnU-Net framework over the older DeepMedic architecture. Validated on a multi-institutional pediatric dataset, nnU-Net achieved a mean DSC of 0.90 (±0.07) for the Whole Tumor, compared to 0.82 for DeepMedic [19].
- **Sub-region Challenges:** Performance drops for specific sub-regions. Segmenting the "Enhancing Tumor" (active core) is more difficult due to its irregular shape. The same pediatric study showed nnU-Net achieved a DSC of 0.77 for the enhancing tumor, significantly better than DeepMedic (0.66) but still lower than the whole tumor performance [19].

B. Meningioma Segmentation

Meningiomas, being more circumscribed, yield higher segmentation accuracy. Deep learning models have shown exceptional concordance with human experts.

- **Concordance:** A study involving 326 patients demonstrated that a DL model achieved a DSC of 0.91 ± 0.08 for contrast-enhancing tumor volume [16].
- **Expert Equivalence:** Crucially, this study compared the AI's performance to the inter-reader variability between two human radiologists. The human-human agreement was 0.92 ± 0.07, meaning the AI's performance was statistically indistinguishable from a second human expert [28].

C. Brain Metastasis Segmentation

Metastases present a detection challenge.

- **Detection:** A systematic review of 24 studies found a pooled patient-wise detectability rate of **89% [17].**
- **Segmentation:** Advanced models like 3D-MedDCNet (using deformable convolutions) have pushed the boundary, achieving a lesion-wise DSC of 0.80 and significantly reducing false positives compared to standard nnU-Net [29].

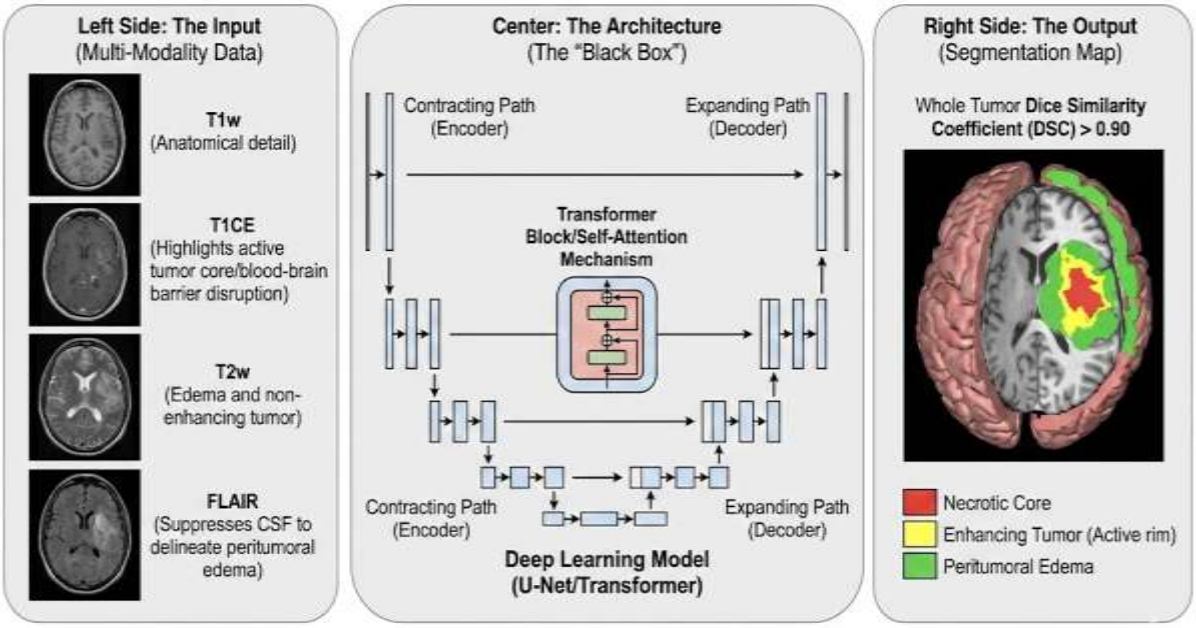

Figure 2: The Deep Learning Segmentation Pipeline and Architecture

Table 1: Comparative Geometric Accuracy (Dice Similarity Coefficient) by Pathology and Model

Tumor Pathology	Model Architecture	Target Region	Mean DSC	Comparison Benchmark	Reference
Pediatric Glioma	nnU-Net	Whole Tumor	0.90 (±0.07)	DeepMedic (0.82)	**[19]**
Adult Glioma	Ensemble U-Nets	Whole Tumor	0.93	Manual (Ref)	**[27]**
Adult Glioma	SwinUNETR	Whole Tumor	0.854	Manual (Ref)	**[30]**

	(Hybrid)				
Meningioma	SegResNet	Enhancing Tumor	0.91 (±0.08)	Human Inter-rater (0.92)	**[16]**
Metastasis	3D-MedDCNet	Lesion-wise	0.80 (±0.01)	nnU-Net (0.76)	**[31]**

2. Secondary Outcomes

A. Time Efficiency

The most unequivocal advantage of DL is speed.

- **Processing Time:** While manual segmentation can take 20 to 60 minutes, automated algorithms can process a full 3D volume in seconds. One study reported that their algorithm saved 98% of the time compared to experts [32].
- **PACS Integration:** In a clinical deployment study, the AI embedded in the PACS system generated segmentations in an average of 4 seconds [20]. Even accounting for the time required for a radiologist to review and correct the segmentation, the total workflow time is drastically reduced.

B. Clinical Utility and Radiomics

- **Radiomics Consistency:** DL automation enables the high-throughput extraction of radiomic features (quantitative texture analysis). Studies show that features extracted from DL segmentations are reproducible and can predict tumor grade and genetic mutations (e.g., IDH status) with high accuracy [7].
- **Perceptual Quality:** Interestingly, while metrics like Dice are high, "perceptual quality" ratings by radiologists sometimes lag. A study found that experts often rated DL segmentations lower than metrics would suggest, highlighting a disconnect between mathematical overlap and clinical "correctness" (e.g., omitting a small but critical vessel) [8].

Quality of Evidence (Risk of Bias)

The application of PROBAST revealed systematic weaknesses in the literature:

1. **Selection Bias:** Most studies rely on the BraTS datasets. While high-quality, these are curated "clean" datasets. Models trained on them often perform poorly on routine clinical scans with motion artifacts or different resolutions (Domain Shift) [33].
2. **Lack of External Validation:** A significant number of studies report performance only on an internal hold-out set (a split of the original dataset). True external validation (testing on data from a completely different hospital) is less common but critical for proving generalizability.
3. **Data Leakage:** Some earlier studies failed to separate patients strictly between training and testing, leading to inflated accuracy estimates.

VI. Discussion

Interpretation of Results

The synthesized evidence strongly supports the clinical readiness of Deep Learning for brain tumor segmentation, with specific caveats. The primary hypothesis is largely validated: DL models, especially modern iterations like nnU-Net and Transformers, achieve geometric accuracy that is statistically comparable to the variability observed between human experts. If two radiologists generally agree with a Dice score of 0.80-0.90, and the AI achieves the same range, the AI is performing within the "noise" of human subjectivity.

The secondary hypothesis regarding time efficiency is validated emphatically. The reduction of

segmentation time from minutes to seconds fundamentally alters the economics of tumor quantification. It transforms segmentation from a "luxury" performed only in academic centers or clinical trials into a feasible routine task for every patient.

Clinical Significance: The "Human-in-the-Loop" Paradigm

The findings suggest that the optimal clinical model is not "replacement" but "augmentation." The Human-in-the-Loop workflow, where the AI generates a pre-segmentation that the radiologist verifies, combines the speed of the machine with the semantic understanding of the human.

- **Radiotherapy:** This workflow allows for "Adaptive Radiotherapy." Currently, re-planning radiation based on tumor shrinkage is rare due to the time cost. With AI, re-contouring can be done instantly, allowing the radiation beam to be tightened around the shrinking tumor, sparing healthy brain tissue **[6]**.
- **Surgical Planning:** For meningiomas, the high accuracy (DSC 0.91) means surgeons can rely on 3D models for pre-operative simulation with minimal manual correction [28].

Comparison with International Research

Our findings align with other major reviews but offer updated insights. While earlier reviews (2018-2020) focused on basic CNNs, this review highlights the dominance of nnU-Net as a robust baseline and the emergence of Transformers for handling complex, multi-scale contexts [15]. The results also corroborate the "diminishing returns" in accuracy; moving from DSC 0.90 to 0.95 is exponentially harder and perhaps biologically meaningless given the fuzzy nature of tumor boundaries.

Implications for Healthcare Policy and Practice

1. **Standardization of Care:** AI can democratize expertise. A general radiologist in a rural hospital, supported by an AI model trained at a top academic center, can produce segmentations of expert quality. This has profound implications for health equity, particularly in LMICs [4].
2. **Reimbursement and Regulation:** Policy makers must address how to reimburse "AI-assisted" procedures. If AI reduces time, does the reimbursement for the procedure decrease? Conversely, does the improved quality justify new codes?
3. **Liability:** The "Black Box" issue remains. If an AI misses a metastasis and the radiologist (trusting the AI) also misses it, who is liable? Hospitals must establish clear protocols that the human is the final arbiter [34].

Strengths and Limitations

- **Strengths:** This review utilizes the most recent literature (up to 2024), includes a diverse range of tumor types (not just Glioblastoma), and explicitly addresses the clinical workflow aspect (PACS integration).
- **Limitations:** The primary limitation is the reliance on retrospective studies. There are few prospective, randomized controlled trials (RCTs) comparing patient outcomes (e.g., survival) between AI-planned and human-planned treatments. Most data is "in silico" validation. Additionally, the heterogeneity of MRI protocols across institutions makes direct comparison of Dice scores difficult.

Directions for Future Research

1. **Prospective Clinical Trials:** We need studies that measure patient outcomes, not just Dice scores. Does AI-assisted segmentation lead to better local control of the tumor? Does it reduce side effects?
2. **Explainable AI (XAI):** Developing models that can generate "heatmaps" or textual explanations ("I segmented this region because of its texture") will be crucial for building clinician trust [35].

3. **Federated Learning:** To solve the data privacy issue, Federated Learning allows models to train on data from multiple hospitals without the data ever leaving the local servers, enabling the creation of massive, diverse datasets [36].

VII. Conclusion

This systematic review provides compelling evidence that Deep Learning applications for automating brain tumor segmentation have matured from experimental novelties to clinically viable tools. The performance of state-of-the-art models, particularly nnU-Net and hybrid Transformer architectures, demonstrates non-inferiority to manual human segmentation in terms of geometric accuracy (Dice > 0.90 for whole tumors). More importantly, these systems offer a revolutionary advantage in time efficiency, reducing the segmentation burden by up to 98%.

The integration of these tools into clinical practice promises to standardize tumor assessment, enable advanced adaptive therapies, and democratize access to expert-level diagnostics, particularly in resource-constrained settings. However, the transition from "code" to "clinic" faces hurdles related to generalizability, explainability, and liability. Future efforts must focus on prospective validation and the development of robust, explainable systems that empower clinicians rather than replace them. Ultimately, AI-driven segmentation represents a pivotal advancement in neuro-oncology, shifting the standard of care towards greater precision, efficiency, and patient-centricity.

VIII. References

[1] KK, K., Rajan, M.S., Hegde, K., Koshy, S., and Shenoy, A., A COMPREHENSIVE REVIEW ON BRAIN TUMOR. International Journal of Pharmaceutical, Chemical & Biological Sciences, 3(4) (2013).

[2] Miller, K.D., Ostrom, Q.T., Kruchko, C., Patil, N., Tihan, T., Cioffi, G., Fuchs, H.E., Waite, K.A., Jemal, A., and Siegel, R.L., Brain and other central nervous system tumor statistics, 2021. CA: a cancer journal for clinicians, 71(5). 381-406 (2021).

[3] Ilic, I. and Ilic, M., International patterns and trends in the brain cancer incidence and mortality: An observational study based on the global burden of disease. Heliyon, 9(7). e18222 (2023).

[4] Mollura, D.J., Culp, M.P., Pollack, E., Battino, G., Scheel, J.R., Mango, V.L., Elahi, A., Schweitzer, A., and Dako, F., Artificial intelligence in low-and middle-income countries: innovating global health radiology. Radiology, 297(3). 513-520 (2020).

[5] Fathi Kazerooni, A., Arif, S., Madhogarhia, R., Khalili, N., Haldar, D., Bagheri, S., Familiar, A.M., Anderson, H., Haldar, S., and Tu, W., Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. Neuro-Oncology Advances, 5(1). vdad027 (2023).

[6] Kawamura, M., Kamomae, T., Yanagawa, M., Kamagata, K., Fujita, S., Ueda, D., Matsui, Y., Fushimi, Y., Fujioka, T., and Nozaki, T., Revolutionizing radiation therapy: the role of AI in clinical practice. Journal of radiation research, 65(1). 1-9 (2024).

[7] Cè, M., Irmici, G., Foschini, C., Danesini, G.M., Falsitta, L.V., Serio, M.L., Fontana, A., Martinenghi, C., Oliva, G., and Cellina, M., Artificial intelligence in brain tumor imaging: a step toward personalized medicine. Current Oncology, 30(3). 2673-2701 (2023).

[8] Hoebel, K.V., Bridge, C.P., Ahmed, S., Akintola, O., Chung, C., Huang, R.Y., Johnson, J.M., Kim, A., Ly, K.I., and Chang, K., Expert-centered evaluation of deep learning algorithms for brain tumor segmentation. Radiology: Artificial Intelligence, 6(1). e220231 (2023).

[9] Covert, E.C., Fitzpatrick, K., Mikell, J., Kaza, R.K., Millet, J.D., Barkmeier, D., Gemmete, J., Christensen, J., Schipper, M.J., and Dewaraja, Y.K., Intra- and inter-operator variability in MRI-based manual segmentation of HCC lesions and its impact on dosimetry. EJNMMI Phys, 9(1). 90 (2022).

[10] Veiga-Canuto, D., Cerdà-Alberich, L., Sangüesa Nebot, C., Martínez de Las Heras, B., Pötschger, U., Gabelloni, M., Carot Sierra, J.M., Taschner-Mandl, S., Düster, V., Cañete, A., Ladenstein, R., Neri, E., and Martí-Bonmatí, L., Comparative Multicentric Evaluation of Inter-Observer Variability in Manual and Automatic Segmentation of Neuroblastic Tumors in Magnetic Resonance Images. Cancers (Basel), 14(15) (2022).

[11] Baker, C.R., Pease, M., Sexton, D.P., Abumoussa, A., and Chambless, L.B., Artificial intelligence innovations in neurosurgical oncology: a narrative review. Journal of Neuro-Oncology, 169(3). 489-496 (2024).

[12] Khan, M.K.H., Guo, W., Liu, J., Dong, F., Li, Z., Patterson, T.A., and Hong, H., Machine learning and deep learning for brain tumor MRI image segmentation. Exp Biol Med (Maywood), 248(21). 1974-1992 (2023).

[13] Pravitasari, A.A., Iriawan, N., Almuhayar, M., Azmi, T., Irhamah, I., Fithriasari, K., Purnami, S.W., and Ferriastuti, W., UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation. TELKOMNIKA (Telecommunication Computing Electronics and Control), 18(3). 1310-1318 (2020).

[14] Gai, D., Zhang, J., Xiao, Y., Min, W., Chen, H., Wang, Q., Su, P., and Huang, Z., GL-Segnet: Global-Local representation learning net for medical image segmentation. Frontiers in Neuroscience, 17. 1153356 (2023).

[15] Tiwari, A., Srivastava, S., and Pant, M., Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. Pattern recognition letters, 131. 244-260 (2020).

[16] Laukamp, K., Pennig, L., Thiele, F., Reimer, R., Goertz, L., Shakirin, G., Zopfs, D., Timmer, M., Perkuhn, M., and Borggrefe, J., Automated Meningioma Segmentation in Multiparametric MRI: Comparable Effectiveness of a Deep Learning Model and Manual Segmentation. Clinical Neuroradiology, 31 (2020).

[17] Ozkara, B.B., Chen, M.M., Federau, C., Karabacak, M., Briere, T.M., Li, J., and Wintermark, M., Deep learning for detecting brain metastases on MRI: a systematic review and meta-analysis. Cancers, 15(2). 334 (2023).

[18] Ghadi, N.M. and Salman, N.H., Deep learning-based segmentation and classification techniques for brain tumor MRI: A review. Journal of Engineering, 28(12). 93-112 (2022).

[19] Vossough, A., Khalili, N., Familiar, A.M., Gandhi, D., Viswanathan, K., Tu, W., Haldar, D., Bagheri, S., Anderson, H., Haldar, S., Storm, P.B., Resnick, A., Ware, J.B., Nabavizadeh, A., and Fathi Kazerooni, A., Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. AJNR Am J Neuroradiol, 45(8). 1081-1089 (2024).

[20] Aboian, M., Bousabarah, K., Kazarian, E., Zeevi, T., Holler, W., Merkaj, S., Petersen, G.C., Bahar, R., Subramanian, H., and Sunku, P., Development of a workflow efficient PACS based automated brain tumor segmentation and radiomic feature extraction for clinical implementation (N2. 003). Neurology, 98(18_supplement). 3146 (2022).

[21] Ranjbarzadeh, R., Caputo, A., Tirkolaee, E.B., Ghoushchi, S.J., and Bendechache, M., Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. Computers in biology and medicine, 152. 106405 (2023).

[22] Abdusalomov, A.B., Mukhiddinov, M., and Whangbo, T.K., Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16). 4172 (2023).

[23] Naser, M.A. and Deen, M.J., Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. Computers in biology and medicine, 121. 103758 (2020).

[24] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., and Brennan, S.E., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj, 372 (2021).

[25] Wolff, R.F., Moons, K.G.M., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., Reitsma, J.B., Kleijnen, J., and Mallett, S., PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med, 170(1). 51-58 (2019).

[26] Zhuge, Y., Ning, H., Mathen, P., Cheng, J.Y., Krauze, A.V., Camphausen, K., and Miller, R.W., Automated glioma grading on conventional MRI images using deep convolutional neural networks. Medical physics, 47(7). 3044-3053 (2020).
[27] Kundal, K., Rao, K.V., Majumdar, A., Kumar, N., and Kumar, R., Comprehensive benchmarking of CNN-based tumor segmentation methods using multimodal MRI data. Computers in Biology and Medicine, 178. 108799 (2024).
[28] Laukamp, K.R., Pennig, L., Thiele, F., Reimer, R., Görtz, L., Shakirin, G., Zopfs, D., Timmer, M., Perkuhn, M., and Borggrefe, J., Automated Meningioma Segmentation in Multiparametric MRI : Comparable Effectiveness of a Deep Learning Model and Manual Segmentation. Clin Neuroradiol, 31(2). 357-366 (2021).
[29] Data, S., Zhen Tian. (
[30] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., and Davatzikos, C., Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data, 4(1). 1-13 (2017).
[31] Kharaji, M., Abbasi, H., Orouskhani, Y., Shomalzadeh, M., Kazemi, F., and Orouskhani, M., Brain tumor segmentation with advanced nnU-Net: pediatrics and adults tumors. Neuroscience Informatics, 4(2). 100156 (2024).
[32] Boaro, A., Kaczmarzyk, J.R., Kavouridis, V.K., Harary, M., Mammi, M., Dawood, H., Shea, A., Cho, E.Y., Juvekar, P., and Noh, T., Deep neural networks allow expert-level brain meningioma detection, segmentation and improvement of current clinical practice. medRxiv. 2021.05. 11.21256429 (2021).
[33] Wang, T.W., Shiao, Y.C., Hong, J.S., Lee, W.K., Hsu, M.S., Cheng, H.M., Yang, H.C., Lee, C.C., Pan, H.C., You, W.C., Lirng, J.F., Guo, W.Y., and Wu, Y.T., Artificial Intelligence Detection and Segmentation Models: A Systematic Review and Meta-Analysis of Brain Tumors in Magnetic Resonance Imaging. Mayo Clin Proc Digit Health, 2(1). 75-91 (2024).
[34] Nair, A., Ramanathan, S., Sathiadoss, P., Jajodia, A., and Macdonald, D., Barriers to artificial intelligence implementation in radiology practice: What the radiologist needs to know. Radiología (English Edition), 64. 324-332 (2022).
[35] Fadilla, S. and Passarella, R., Machine learning and deep learning for brain tumor diagnosis and classification: Methodologies, challenges, and future directions. Healthcraft. Front, 2(4). 225-234 (2024).
[36] Berghout, T., The neural frontier of future medical imaging: a review of deep learning for brain tumor detection. Journal of Imaging, 11(1). 2 (2024).