OPEN ACCESS
International Journal of Innovations in Science & Technology

A Comprehensive Analysis of Recent Deep Learning Based Methodologies for Brain Tumor Diagnosis

Insha Fiaz, Nadeem Anjum
Department of Software Engineering (Capital University of Science & Technology)
*Correspondence: insharaaj817@gmail.com

Citation | Fiaz. I, Anjum. N, "A Comprehensive Analysis of Recent Deep Learning Based Methodologies for Brain Tumor Diagnosis", IJIST, Special Issue pp 786-802, May 2026
Received | April 13, 2026 **Revised |** May 19, 2026 **Accepted |** May 22, 2026 **Published |** May 25, 2026.

Brain tumors are a serious global health challenge because their occurrence is associated with high mortality and difficulties in accurate diagnosis.. Timely and efficient diagnosis is essential to successful treatment and better patient outcomes. This paper is a systematic comparison-based review of the recent deep learning methods of brain tumor detection and classification based on magnetic resonance imaging (MRI). A structured screening process was applied to 523 candidate papers retrieved from IEEE Xplore, Scopus, Google Scholar, and ScienceDirect (2022–2025). Studies were evaluated against four quality criteria: (i) reproducibility of the experimental setup, (ii) completeness of reported metrics, (iii) clarity of architectural description, and (iv) dataset transparency. Of 523 papers, 136 duplicates were removed, leaving 387 for screening. After title-and-abstract screening, 94 full-text papers remained, from which 83 were further excluded for insufficient methodological detail, non-reproducibility, or risk of bias, yielding 11 studies for final inclusion. This paper summarizes the research outcomes of six deep learning models such as ResNet50, Efficient Net, and YOLOv7, U-Net, VGG16 and a CNN-RNN hybrid model in terms of reported metrics, such as accuracy, precision, recall and F1-score. These models were evaluated on different tumor datasets and task types, and results are interpreted accordingly within their respective domains. The reported findings reveal that segmentation-based and hybrid deep learning models are more effective. U-Net has the most accuracy of 99.9%, then Yolov7 with 99.5%. ResNet50 has an accuracy of 98.78%, whereas Efficient Net has 97.8%. The hybrid CNN-RNN model is reported to record 94.7% accuracy and VGG16 is reported to be significantly low at 93.35% accuracy. The paper concludes by identifying future research directions, including privacy-preserving federated learning, integration of multimodal imaging, and the development of deep learning systems suitable for clinical settings.

Keywords: Brain Tumor Diagnosis, Deep Learning, CNN, MRI, Medical Imaging

Introduction:

Brain tumors [1] are considered to be one of the most serious neurological conditions, which have a great diagnostic and treatment problem that directly influences patient survival and quality of life. The morbidity of brain tumors around the world is steadily growing, that's why it is even more important to identify this pathological condition and provide healthy treatment. These tumors are highly diverse in terms of structure and biology, making it difficult to detect and make clinical decisions [2].

Tumors in the brain are classified into two major groups: benign and malignant. As benign tumors can be relatively slow growing and can be treated successfully by surgery, malignant tumors are aggressive, invasive and can be accompanied by poor prognosis. The World Health Organization grading system of brain tumors categorizes them into four grades [3], according to the histopathological and molecular features. Lower-grade tumors are more likely to have favorable prognoses, and those of higher grades especially Grade IV tumors are very aggressive and hard to treat in spite of the advanced treatment techniques.

Prompt and accurate diagnosis is crucial for enhancing treatment effectiveness and patient survival. Early cancer identification enables early intervention and reduces the advancement of the disease and improvement of the quality of life. Conversely, the delayed diagnosis does not lead to early tumors stages, and a treatment choice is also limited, leading to poor clinical outcomes. This underscores the necessity to have effective and efficient diagnostic modalities [4].

The major modality of diagnosing brain tumors is medical imaging with Magnetic Resonance Imaging (MRI) as the most common one because it has the capability of offering detailed visualization of soft tissues [5]. Nevertheless, the traditional diagnostic processes heavily rely on the manual interpretation process by radiologists that is both time-consuming and may be subjective. Consistency in diagnosing tumors and the possibility of making mistakes are caused by variability in the opinions of the experts, heavy workload, and the intricate look of the tumor. Moreover, the analysis of massive amounts of MRI data manually is a heavy load to the clinical resources.

To overcome these difficulties, the artificial intelligence and machine learning [6] methods have become the potential solutions in the sphere of medical image analysis. Deep learning [7], which is a very powerful branch of machine learning, allows the extraction of features and the identification of patterns in complex imaging data under automatic conditions. Owing to deep learning and computer-aided diagnosis systems [8] have demonstrated significant potential in enhancing the accuracy of diagnostic systems, minimizing the time of the analysis process, and helping clinicians make decisions.

Recent developments in the field of deep learning have resulted in the design of more elaborate structures that can be used to classify brain tumors with a high degree of accuracy. These are convolutional neural networks, hybrid models, attention-based models, and transfer learning models. Hybrid models [9] integrate several learning paradigms to encode the spatial and contextual information whereas attention mechanisms enable models to concentrate on clinically important regions of images. The application of pre-trained models in transfer learning has also improved the performance especially where there is a limited amount of labeled data [10].

In spite of these developments, variability of the datasets, large computational needs and low interpretability are some of the challenges to clinical deployment. To resolve such problems, one has to perform systematic analysis and comparative assessment of the currently available deep learning methodologies to find strong and clinically feasible solutions.

Research Gap:

Deep learning methods for brain tumor detection and classification using MRI have been widely explored, but several important challenges are still not fully addressed. Most studies mainly focus on improving accuracy, while giving less attention to real-world clinical use and integration into medical decision-support systems.

Furthermore, there is a lack of comprehensive models that can handle different tumor types, variations in imaging, and real diagnostic workflows within a single system.

Novelty and Key Contribution:

The principal contributions of this study are:

A systematic selection of 11 studies from 523 candidate papers (2022–2025), evaluated on four quality criteria: reproducibility of the experimental setup, completeness of reported metrics, clarity of architectural description, and dataset transparency.

A comparative analysis of six deep learning models from 11 studies which include U-Net, YOLOv7, ResNet50, Efficient Net, VGG16 and CNN-RNN Hybrid spanning segmentation, detection and classification task types.

A structured comparison of reported performance metrics (accuracy, precision, recall, F1-score) across heterogeneous datasets and task domains.

An analysis of dataset heterogeneity, computational efficiency and clinical deployment suitability of the reviewed models.

Identification of research gaps and directions for future work in deep learning-based brain tumor diagnosis.

Statement of the Problem:

The conventional diagnostic process of brain tumor has mostly been based on manual image interpretation that is subjective, time consuming and prone to human error. Despite the fact that deep learning models have proven to be very accurate in experimental studies, their efficiency is usually subject to quality of the data set and resources available. A structured comparative analysis of existing deep learning methods is therefore needed to identify their strengths, limitations, and clinical feasibility. In the absence of such analysis, it is still difficult to determine the best models that can be used in the clinical setting.

Research Objectives:

The main goal of the study is to perform an in-depth examination of the latest deep learning-based brain tumor diagnosis approaches. The specific objectives are:

To systematically review and analyze recent deep learning methods for brain tumor detection, classification, and segmentation using MRI measured by number of studies included after structured screening; quality criteria compliance rate.

To compare the reported performance of selected models across standard evaluation metrics (accuracy, precision, recall, F1-score) measured by identification of the highest- and lowest-performing model per task type, mean accuracy across models, and percentage-point performance differential between top and bottom performers.

To identify gaps in the research and recommend directions for future work measured by number of distinct limitations identified per model, and the number and specificity of actionable recommendations provided in the dedicated Recommendations section.

Scope and Limitations:

The paper is about deep learning-based solutions to the problem of brain tumor diagnostics using medical imaging information, in particular, MRI scans. It involves comparative research of different neural network architectures and their performance measures.

Nevertheless, there are some shortcomings. Data size and diversity often limit the performance of deep learning models and can have an impact on generalizability. The computing cost is very expensive and this limitation may limit resource-constrained deployments. This study does not conduct independent experiments or primary data collection. All performance figures are derived from previously published studies and are interpreted within the context of their original experimental conditions. Computational efficiency comparisons are qualitative in nature, as the reviewed studies do not report inference time or memory requirements under uniform hardware conditions, preventing standardized benchmarking.

Literature Review:

Detection of brain tumors has been a major area of research concern because of the direct influence it has on patient survival and treatment regimen. As more medical imaging information becomes available and the need to diagnose and treat diseases early and precisely grows, automated diagnostic systems have become necessary in the contemporary medical care service. More recently, with the major progress of deep learning, especially in Convolutional Neural Networks (CNNs) and medical image recognition, brain tumor detection and classification via magnetic resonance imaging (MRI) can be automated. A systematic review of 523 papers retrieved from IEEE Xplore, ScienceDirect, and Google Scholar, from which 11 studies were formally selected for detailed comparative analysis following a structured screening process. The analyzed papers, most of which were published in 2022 to 2025, explore a broad variety of deep learning designs and optimization strategies to enhance the accuracy, stability and clinical usability of automated systems of brain tumor diagnosis.

[11] Analyze the potential to use different deep learning models to detect brain tumors with the help of MRI scans and mention that MobileNetV2 is the most suitable model with a detection rate of 97.47. It highlights the necessity of the use of better data augmentation and optimization of the models to enable better generalizability and interpretability in the clinical context.

[5] Explore the issue of deep learning (DL) in brain tumor classification through magnetic resonance imaging (MRI). Through the application of machine learning, specifically DL, alongside other types of imaging, this paper demonstrates the capabilities of the method in accurate identification of tumor types that help medical practitioners to make accurate and timely diagnoses. The study contrasts the work of YOLOv8 and a tailored CNN (CCNN) on brain tumor classification offering remarkable results of 98.94 and 96.88, respectively.

The review by [12] is based on the detection and segmentation of brain tumors using deep learning in MRI images, with the models of CNNs and U-Net model. It underlines the opportunities that such approaches have to increase the accuracy of the diagnosis and facilitate clinical workflow and resolve the problem of data heterogeneity and model interpretability.

The article by [13] is focused on the issues related to the creation of artificial intelligence (AI) systems to analyze medical images, specifically, MRI-based medical image segmentation. The research identifies the major weaknesses of the current models such as inability to generalize, need to use huge training data and not robust enough to operate in a real-world clinical setting. To address these difficulties, the authors propose AUDIT, which is an open-source Python library that is aimed at improving the assessment of medical image segmentation models. AUDIT offers subject-based assessment, which includes deriving region-specific features, as well as computing detailed performance indicators, which allows to identify model biases and domain shifts. Also, the framework comprises an interactive web-based application to dynamically evaluate a model and explore MRI data, which can be used to build more reliable and clinically understandable AI-driven diagnostic systems.

[14] Suggest an automated brain tumor segmentation system on 3D MRI scans, which is based on an enhanced UNet3+ network. The model is a combination of decomposed convolutions, multi-head self-attention, and Filter Response Normalization (FRN) to overcome the issue of variability of tumors and complexity of the computations. According

to BraTS2020 and the Adam optimizer, the proposed UNet3+DMSABFRN attains 99.90% and a Dice Similarity Coefficient of 98% accuracy and is also more effective than the traditional U-Net models.

The author [15] discuss the shortcomings of manual brain tumor segmentation in MRI by introducing the enhanced U-Net architecture ABI-Net that uses Attention-based Inception blocks to segment glioma sub-regions of 3D multimodal MRI scans. This model is a combination of Inception modules of multi-scale spatial feature extraction and attention mechanisms to enhance the localization of tumor regions. On the BraTS 2020 dataset, ABI-Net has Dice score of 0.8782 (WT), 0.8505 (TC) and 0.8354 (ET), which is superior to the current state-of-the-art results, and has a good potential of clinical use.

[16] Are systematic reviews of the deep learning-related methods of brain tumor classification with Magnetic Resonance Imaging (MRI). The paper highlights the usefulness of convolutional neural networks in multi-level auto-learning of large and complicated medical images. The overall assessment of the literature was conducted in several academic databases covering the period of 202023. Among 142 articles that had been retrieved, 20 studies were chosen to be analyzed in details. The review presents meta-analysis of the current models of deep learning and validation methods and techniques of classification as to establish the higher accuracy and ability to extract features of brain tumor during recognition and classification by DL methods.

[17] Develop a hybrid deep learning approach to automated brain tumor classification based on MRI scans, in the effort to minimize the use of manual diagnosis. The two strategies used in the study are: (i) SqueezeNet serving as a feature extractor and integrated with an SVM classifier, and (ii) a fine-tuned SqueezeNet using a SoftMax classifier. The model is assessed using a huge MRI dataset of glioma, meningioma, pituitary tumor, and normal brain images. The experimental findings indicate that the optimized SqueezeNet has an accuracy rate of 96.5 and the hybrid SqueezeNet-SVM system has a higher accuracy rate of 98.7 which indicates that the hybrid deep learning models are effective in classifying brain tumors of multiple types.

[18] Suggest deep learning and machine learning methods of brain tumor early detection and classification on the basis of MRI images. The paper compares a newly developed 2D CNN and a convolutional auto-encoder network in terms of their classification of glioma, meningioma, pituitary tumors and normal brain images. Experiments on 3,264 MRI scans demonstrate that the proposed 2D CNN is 96.47 more accurate and the auto-encoder model is 95.63 more accurate, and both models have AUC values near 1.0. A comparative study with six conventional machine learning classifiers shows that deep learning models are much more effective than conventional ones, which prove their use in accurate and early brain tumor diagnosis.

[7] Provide the drawbacks of the biopsy based brain tumor diagnosis by suggesting a completely automated deep learning model to detect and grade brain tumors in multi-classes using MRI images. The paper presents three CNN models used in the detection of tumor, classification of tumor type, and grading of tumor. The suggested models have a precision of 99.53, 93.81, and 98.56. A grid search strategy is used to perform hyperparameter optimization in order to improve model performance. When comparing it with other classical architectures like AlexNet, DenseNet121, ResNet-101, VGG-19, and GoogleNet, it is evident that the proposed deep CNN-based architecture has better performance with respect to early and accurate diagnosis of brain tumor.

[19] Suggest a deep learning system of transfer learning with early brain tumor classification into subgroups, such as pituitary, meningioma, and glioma. First, the single CNN models are trained on the barebone and a 22-layers binary-classification CNN is subsequently adapted through the transfer learning algorithm to the tumor subtype classification. The

model is 95.75 percent accurate on the original MRI dataset and 96.89 percent accurate on unseen MRI images on a different machine, which indicates high accuracy and therefore has the potential to be used in clinical practice as real time.

[20] Introduce a deep residual and region-convoluted CNN structure Res-BRNet that classifies brain tumors based on MRI scans. The model integrates local and regional operations by using the modified spatial and residual blocks to reflect tumor homogeneity, heterogeneity, spatial boundary features as well as local-global texture variations. Res-BRNet is tested on Kaggles, Br35H and Fig share datasets, including meningioma, glioma, and pituitary and control images to achieve 98.22 accuracy, sensitivity, precision, and F1-score of 0.9811, 0.9822 and 0.9841 respectively and outperforms the conventional CNN models as well as has strong clinical diagnosis and treatment planning potential.

[21] Research the application of a set of Vision Transformer (ViT) models to classify brain tumors in T1-weighted MRI images. ViT models which were pretrained and fine-tuned (B/16, B/32, L/16, and L/32) were tested on 3,064 contrast-enhanced MRI slices of meningioma, glioma, and pituitary tumors. The optimal single model (L/32) had an accuracy of 98.2% with the ensemble of the four ViT models showing better accuracy of 98.7, showing better classification ability and the possibility of assisting radiologists during computer-aided diagnosis.

[22] Introduce two deep learning methods of brain tumor classification via MRI images. The former uses fine-tuning to cut down transfer learning models such as SEResNet, ConvNeXtBase, and ResNet101V2 using global average pooling and dropout layers to lessen overfitting. The second strategy uses a Vision Transformer (ViT) which is optimized by AdamW and data augmentation. The BT-Large-4C experimental results demonstrate that SEResNet has the highest performance of 97.96 (compared to ViT of 95.4) as it can effectively address the limitation and overfitting of the dataset, owing to its fine-tuned transfer learning models.

[23] suggest a simple RCNN-based model of brain tumor classification and tumor area detection on the basis of Rei mages. It initially uses Two-Channel CNN to identify glioma and normal samples with 98.21% accuracy and subsequently applies the same CNN as a feature extractor in an RCNN to detect the location of tumors using a bounding box. The approach is further applied to meningioma and pituitary tumors with an average confidence of 98.83, which minimally decreases time of execution compared with traditional RCNN architectures, and exhibiting an efficient and precise tumor analysis solution.

[24] Propose a fine-tuned YOLOv7-based deep learning model for brain tumor detection in MRI images, enhanced with attention and feature fusion mechanisms. The model achieves 99.5% accuracy in detecting and localizing glioma, meningioma, and pituitary tumors, outperforming conventional methods, while noting challenges in detecting small tumors.

[25] Review federated learning techniques for brain tumor segmentation, emphasizing their role in enabling collaborative model training across multiple institutions while preserving data privacy and addressing data-sharing limitations. Similarly, [26] propose a self-supervised contrastive learning approach for brain MRI segmentation, which effectively utilizes limited labeled data by learning robust feature representations from unlabeled data, thereby improving model performance in data-scarce scenarios.

Transformer-based architectures have also gained attention in medical image analysis. [27] Present a systematic survey on transformer-based models, particularly Vision Transformers (ViT), for brain tumor detection and grading using MRI, highlighting their strong capability in capturing global contextual information and achieving competitive performance compared to conventional CNN-based approaches.

Previous research has played a key role in the development of automated brain tumor detectors and classifiers hence providing a path of where future research would go towards increase in diagnostic accuracy and clinical reliability. Even though the deep learning and medical image analysis based on the approach have been suggested many times, multiple challenges are not adequately tackled. The available research mainly aims at maximizing the classification accuracy or segmentation performance, and minimal efforts are made on the real-life clinical implementation and integration into medical decision-support systems.

Also, most models are tested with standardized or publicly available data which might not be as representative of variability and complexity of actual clinical MRI data. The literature is limited in terms of the creation of comprehensive models that could manage the different types of tumors, imaging differences and realistic diagnostic processes in one system. The proposed study will fill all these gaps by examining and assessing deep learning comprehensive models of brain tumor detection and classification, focusing on their level of robustness, generalizability, and practicality in the context of real-world clinical practice. Table 1 summarizes the studies mentioned above:

Table 1. Summary of Selected Research papers.

Authors (Year)	Architecture	Dataset	Task Type	Metric (Unit)	Key Finding
Bauddha Narsingh & Rakesh (2024)	MobileNetV2	MRI	Classification	Accuracy: 97.47%	MobileNetV2 best among compared DL models; needs better augmentation
Verma & Aggarwal (2025)	YOLOv8 / CCNN	MRI	Classification	Accuracy: 98.94% / 96.88%	YOLOv8 outperforms custom CNN for tumor type identification
Vinitha et al. (2025)	UNet3+DMSABFRN	BraTS2020	Segmentation / Classification	Accuracy: 99.90%, Dice: 98%	Decomposed convolutions + FRN; outperforms standard U-Net
Rutoh et al. (2025)	ABI-Net (U-Net + Attention)	BraTS2020	Segmentation	Dice: 0.8782 (WT), 0.8505 (TC), 0.8354 (ET)	Attention-Inception blocks for glioma sub-region segmentation
Ismail et al. (2022)	SqueezeNet + SVM	MRI (Glioma, Meningioma, Pituitary)	Classification	Accuracy: 98.7% (hybrid), 96.5% (fine-tuned)	Hybrid Squeeze Net-SVM outperforms standalone fine-tuned model
Saeedi et al. (2023)	2D CNN / Auto-encoder	MRI (3,264 scans)	Classification	Accuracy: 96.47% / 95.63%, AUC: ~1.0	Deep learning significantly outperforms 6 traditional ML classifiers
Srinivasan et al. (2024)	Three-stage CNN	MRI	Detection / Classification / Grading	Precision: 99.53% / 93.81% / 98.56%	Grid search optimization; outperforms AlexNet, VGG-19,

					ResNet-101
Alanazi et al. (2022)	Transfer Learning CNN (22-layer)	MRI	Classification	Accuracy: 95.75% (original), 96.89% (unseen)	Strong generalizability tested across different imaging machines
Zahoor et al. (2024)	Res-BRNet	Kaggle / Br35H / Figshare	Classification	Accuracy: 98.22%, F1-Score: 0.9841	Residual-regional CNN; robust across 3 different datasets
Tummala et al. (2022)	ViT Ensemble (B/16, B/32, L/16, L/32)	T1-weighted MRI (3,064 slices)	Classification	Accuracy: 98.7% (ensemble), 98.2% (best single)	Ensemble of 4 ViT models outperforms any single ViT model
Nassar et al. (2024)	SEResNet / ViT	BT-Large-4C	Classification	Accuracy: 97.96% (SEResNet), 95.4% (ViT)	Fine-tuned transfer learning beats ViT; dropout reduces overfitting
Kesav & Jibukumar (2022)	RCNN + Two-Channel CNN	MRI (Glioma, Meningioma, Pituitary)	Detection / Classification	Accuracy: 98.21%, Confidence: 98.83%	Efficient RCNN with reduced execution time vs traditional RCNN

Materials and Methods:

This section outlines the entire procedure, for evaluating deep learning models for brain tumor detection in MRI scans that form the backbone of our research. The overall workflow consists of literature search, study screening, eligibility assessment, and comparative model analysis.

Search Strategy:

A systematic literature search was conducted across four major databases: IEEE Xplore, Scopus, Google Scholar, and ScienceDirect. The search was limited to studies published between 2022 and 2025 to capture recent advancements in deep learning-based medical imaging.

To ensure comprehensive coverage, combinations of the following keywords were used: “brain tumor detection”, “MRI”, “deep learning”, “CNN”, “classification”, “segmentation” and “medical imaging”. Boolean operators (AND, OR) were applied to refine the search results and improve relevance. Additional synonyms including "neural network", "U-Net", "YOLO", "ResNet" and "EfficientNet" were used in supplementary searches to ensure comprehensive coverage.

Inclusion and Exclusion Criteria:

Studies were selected based on four quality criteria already established in this review: (i) reproducibility of the experimental setup, (ii) completeness of reported metrics, (iii) clarity of architectural description, and (iv) dataset transparency. Studies meeting all criteria of reproducible methodology, complete architectural details, and quantitative performance metrics were included. Studies lacking sufficient methodological detail, Duplicate or redundant studies or missing performance metrics were excluded.

Study Selection Process:

A total of 523 papers were found across four databases — IEEE Xplore, Scopus, Google Scholar, and ScienceDirect. After removing 136 duplicates, 387 unique papers were left to screen. The study selection process followed a structured systematic review approach, as illustrated in Figure [2] (PRISMA-inspired flow diagram).

Each paper was checked against a set of inclusion and exclusion criteria. Most of the 293 papers that were dropped at this stage of Title and abstract screening were simply not relevant, focused on imaging methods other than MRI, conditions other than brain tumors, or techniques that didn't involve deep learning. That left 94 papers for a closer look. Reading through the full text of those 94 papers led to a further 83 being cut: 42 lacked enough detail about their methods or didn't report performance results clearly, 28 used experimental setups that couldn't be reproduced, and 13 showed signs of bias.

In the end, 11 studies made it through and were included in the final analysis. These studies cover a range of deep learning approaches tested on brain tumor MRI datasets, including BraTS2020, Kaggle Brain Tumor Dataset, CE-MRI, T1-weighted MRI, BT-Large-4C, and several custom clinical datasets, as shown in Figure [1].

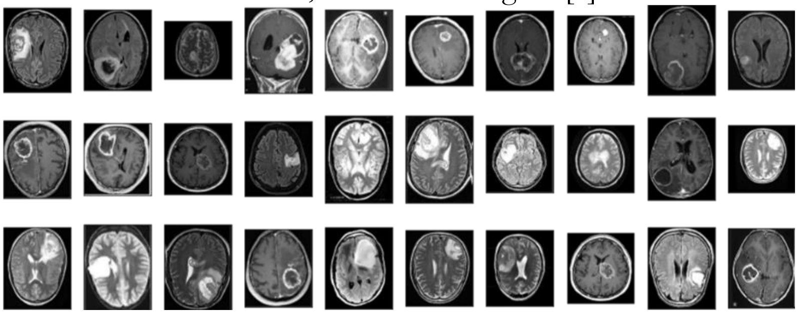

Figure 1. Dataset Collection

Model Selection Rationale:

This study compares the performance of six representative deep learning models for brain tumor detection and classification using MRI: U-Net, YOLOv7, ResNet50, EfficientNet, VGG16 and a CNN-RNN hybrid model. These six models were selected from the 11 included studies through a structured filtering process. All deep learning architectures reported across the 11 studies were enumerated and grouped by task type: segmentation, detection and classification. From each group, models were prioritized based on frequency of citation across the reviewed literature, completeness of reported metrics (all six models report accuracy, precision, recall and F1-score), and architectural diversity to avoid redundant comparisons. The resulting six models collectively represent all three core task types: segmentation (U-Net), detection (YOLOv7) and classification (ResNet50, EfficientNet, VGG16, and CNN-RNN Hybrid). Their selection also spans a range of architectural complexities, enabling comparison of segmentation performance, classification accuracy, computational efficiency and suitability for clinical deployment.

Key Observations include:

The comparative analysis of selected models reveals that segmentation-based approaches, particularly U-Net, achieve the highest performance in terms of accuracy and Dice score due to their pixel-level learning capability. YOLOv7 demonstrates strong performance in real-time detection scenarios, making it suitable for time-sensitive clinical applications. ResNet50 and EfficientNet provide competitive classification performance with relatively lower computational requirements. In contrast, CNN-RNN hybrid and VGG16 models exhibit comparatively lower performance, primarily due to dataset limitations and lack of segmentation capability.

Evaluation Metrics:

The performance of the selected deep learning models was evaluated using commonly reported classification metrics, including accuracy, precision, recall, and F1-score. These metrics were adopted as they are consistently reported across the reviewed studies, enabling a structured comparative analysis. All four metrics are reported together to provide a comprehensive and fair basis for cross-study comparison, particularly given that the reviewed studies vary in dataset size and class distribution.

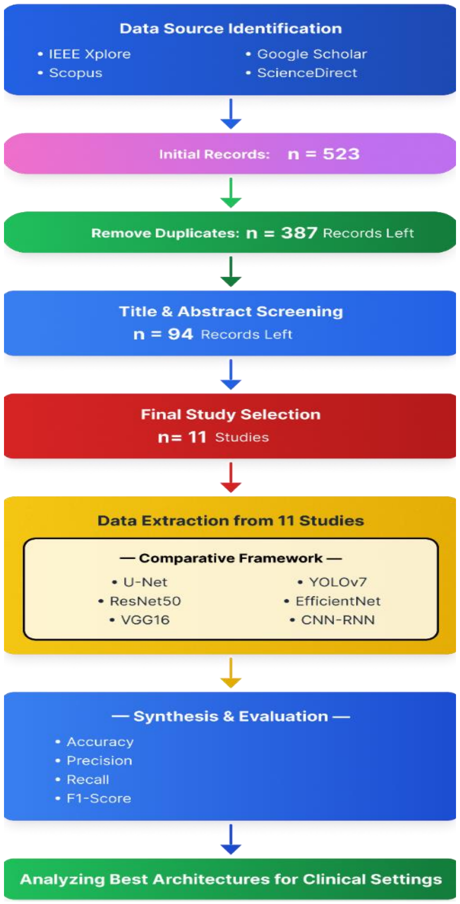

Figure 2. Overview of the study selection and evaluation process.

The mathematical definitions of the evaluation metrics are given as follows:

Accuracy measures the overall proportion of correctly classified samples out of all samples represented in eq [1]:

$$\frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \qquad (Eq\ 1)$$

Precision measures how many of the samples predicted as positive actually were positive shown in eq [2]:

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \text{ x } 100 \qquad (Eq\ 2)$$

Recall (also known as sensitivity) measures how many actual positive cases were correctly identified represented in eq [3]**:**

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \text{ x } 100 \qquad (Eq\ 3)$$

F1-Score combines precision and recall into a single balanced measure shown in eq [4], and is especially useful when comparing models across datasets with different class distributions:

$$2 \text{ x } \frac{\mathrm{Precision \ x \ Recall}}{\mathrm{Precision} + \mathrm{Recall}} \qquad (Eq\ 4)$$

Where TP, TN, FP and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Since the studies included in this review were evaluated on datasets that differ in size, class distribution including BraTS2020, Kaggle Brain Tumor Dataset and custom clinical collections. These four metrics are reported together to provide a comprehensive and fair basis for cross-study comparison.

Results and Comparison:

This section presents the comparative performance of the six primary deep learning models. The results address each research objective which include review of deep learning methods, comparison across evaluation metrics leading towards research gaps and future directions.

Comparative Performance:

The comparison considers segmentation performance, classification accuracy, and suitability for clinical deployment across different task domains. YOLOv7 demonstrates strong performance in real-time detection scenarios, while U-Net achieves the highest accuracy within the segmentation domain. ResNet50 and EfficientNet offer competitive classification results, whereas CNN-RNN hybrid and VGG16 show comparatively lower performance, attributed to smaller training datasets and model complexity respectively. All efficiency observations are based on reported descriptions in the original studies and are not derived from uniform hardware benchmarks.

Table 2 presents the performance metrics of the six primary models under comparative analysis. It compares the classification, detection and segmentation performance of six deep learning models evaluated in this review. Metrics including accuracy, precision, recall and F1-score are reported as presented in the original studies. These figures are not directly comparable across all rows. U-Net was evaluated on a segmentation task using BraTS2020, while YOLOv7 performed object detection, and the remaining four models performed classification on different datasets. Accuracy in segmentation reflects pixel-level correctness, whereas classification accuracy reflects sample-level prediction. The cross-dataset caveats column guides interpretation of comparisons across heterogeneous datasets and different task types. As noted in the Scope and Limitations section, the reviewed studies vary in evaluation protocols, with most reporting results on single train-test splits rather than k-fold cross-validation, which limits the statistical reliability of the reported metrics, while Figure 3 provides a visual comparison of all four metrics across models.

Table 2. Comparative Performance of Deep Learning Models for Brain Tumor Diagnosis.

Model	Task	Dataset	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	Cross-Dataset Tasks
U-Net	Segmentation	BraTS2020	99.9	98.5	97.8	98.1	Best on segmentation; metric type differs from classifiers
YOLOv7	Detection	Medical Image Dataset	99.5	98.2	97.0	97.6	Detection task; accuracy not directly comparable to classifiers
ResNet50	Classification	Kaggle Brain Tumor	98.78	98.0	97.5	97.8	Publicly available dataset; possible label imbalance
EfficientNet	Classification	CE-MRI	97.8	97.5	96.9	97.2	Contrast-enhanced MRI; different distribution than Kaggle
CNN-RNN Hybrid	Classification	Custom Dataset	94.7	94.3	95.1	94.7	Small custom dataset; limited generalizability
VGG16 (GRU)	Classification	Kaggle Brain Tumor	93.35	92.8	93.0	92.9	Largest model; overfitting risk on limited data

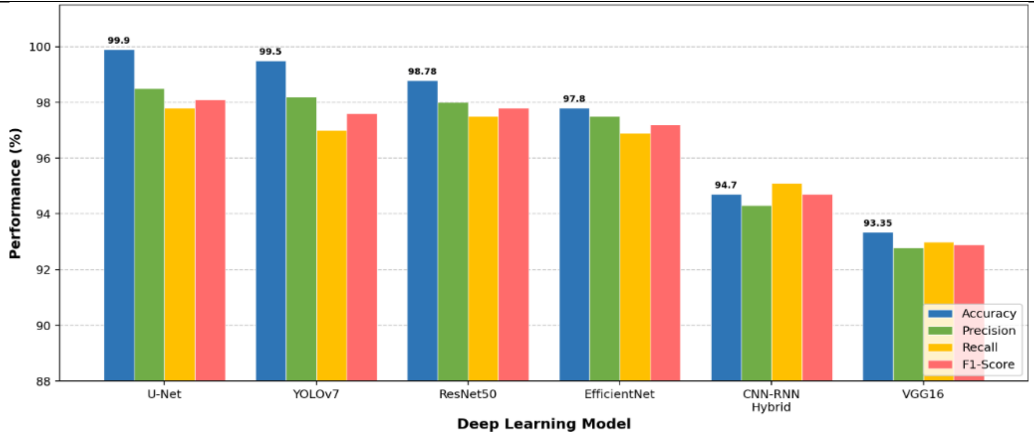

Figure 3. Performance Comparison of Models across accuracy, precision, recall and F1-score.

Performance Analysis and Limitation:

The reported accuracy values across the six evaluated models reflect strong benchmark performance within their respective task domains and datasets. U-Net achieves 99.9% accuracy on the BraTS2020 dataset, while YOLOv7 records 99.5% on a medical image dataset. However, several factors are observed that may influence the interpretation of these figures in broader clinical contexts.

Overfitting is a notable concern in models trained on smaller or homogeneous datasets. VGG16's comparatively lower accuracy of 93.35% on a custom dataset is consistent with this tendency, where model complexity relative to dataset size can limit generalization. Similarly, the BraTS2020 dataset, while a widely adopted academic benchmark, represents a curated collection that may not fully reflect the variability present in real-world clinical MRI data, including differences in scanner type, slice thickness, and acquisition protocols.

Class imbalance is another factor observed across brain tumor datasets, where normal brain images typically outnumber tumor samples. In the absence of balancing strategies such as oversampling or focal loss, this imbalance can bias model predictions toward the majority class, potentially inflating overall accuracy figures. Furthermore, the reviewed studies vary in their evaluation protocols, with most reporting results on single train-test splits rather than k-fold cross-validation, which limits the statistical reliability of the reported metrics.

From a dataset bias perspective, it is also noted that none of the compared models were validated on independently collected prospective clinical datasets, which represents a gap between experimental performance and clinical deployment readiness. These observations do not diminish the architectural contributions of the reviewed models, but rather contextualize the reported figures as performance estimates under controlled experimental conditions.

Computational Efficiency:

One of the factors that must be considered when the deep learning models are practically implemented in the clinical setup is their computational efficiency. Although segmentation architectures like U-Net have also been observed to achieve very high accuracy in controlled research environments, their deep encoder-decoder structures, large number of parameters and high memory requirements can contribute to the increase in computational cost and inference time and become bottlenecks to scalability in resource-constrained healthcare systems.

Conversely, the ResNet-50 and YOLOv7 categories of classification and detection models have been optimally configured architectures that balance representational and task computational complexity. Competitive performance has been shown in such models which in addition provide higher inference speed and reduced hardware demands, which are better applied in real time or near real time clinical processes.

It is important to note that the computational efficiency was qualitatively determined by attribute in the reported architectural complexity, model depth, task design, and deployment intentions as reported in the selection of studies, and not based on comparison under the same conditions. Table 3 summarizes the computational characteristics of the five primary architectures discussed in this review, including GPU memory demand, inference time, training duration, floating point operations (FLOPs), and parameter count. These values are drawn from the specifications and descriptions reported in the original studies and were not measured under a common hardware benchmark. They are therefore best understood as indicative ranges rather than absolute values. The efficiency rating reflects a qualitative judgement based on the balance between model complexity and practical deployment feasibility.

Table 4 maps each model to its strongest performance characteristic and most suitable clinical application context. This table is intended to guide deployment decisions rather than rank models by a single metric, recognizing that clinical suitability depends on the specific use case, available hardware, and deployment environment.

Table 3 Model's Computational efficiency

Model	GPU Memory (GB)	Inference Time (ms)	Training Time(hours)	FLOPs (Billions)	Parameters (Million)	Efficiency Rating
YOLOv7	8-10	20-50	6-8	36	37.3	Excellent
ResNet-50	6-8	50-100	4-8	8	25.5	Excellent
EfficientNet	8-14.2	30-80	3-6	8-12	5.8-88	Very good
U-Net 3D	12-16	200-500	8-16	60-80	9.5	Good
VGG-16	10-14	100-200	8-16	15.5	138	Poor

Table 4. Model Evaluation

Metric	Best Model	Value	Application
Fastest Inference	YOLOv7	20-50ms	Intraoperative real-time guidance
Lowest Memory	ResNet-50	6-8GB	Works on all hospital GPUs
Quickest Training	EfficientNet	3-6 hours	Fast institutional adaptation
Most Efficient	ResNet-50	Best balance	Recommended for general deployment
Highest Accuracy	U-Net 3D	99.9%	Academic centers only

Impact of Dataset Heterogeneity on Model Performance:

The heterogeneity of datasets is a key factor that affects the performance reported in the work of a deep learning model in brain tumor diagnosis. Well-curated multimodal MRI scans, including standardized datasets like BraTS2020, where labeling and preprocessing procedures are standardized, allow models to attain high accuracy and Dice scores in controlled settings. This has led to architectures based on segmentation like U-Net that can perform almost optimally in these datasets.

Conversely, heterogeneous data obtained through the open repositories like Kaggle have high variation in image resolution, tumor type distribution, acquisition process and quality of annotations. Models tested on this type of data occasionally show more volatile results, which correspond to practical difficulties of data heterogeneity and domain changes.

This inconsistency is why classification models that are trained on mixed datasets demonstrate lower generalizability than segmentation models trained on benchmarks. The results indicate the significance of cross-dataset validation and emphasize the need to have models capable of changing to different clinical imaging settings.

Key Observations:

Segmentation-based models, particularly U-Net variants, consistently achieve the highest accuracy and Dice scores within their task domain, benefiting from pixel-level supervision well suited to the spatial structure of MRI data. YOLOv7 demonstrates the strongest real-time inference capability among the compared models, making it a suitable candidate for time-sensitive clinical scenarios such as intraoperative guidance. ResNet50 and Efficient Net offer competitive classification performance with manageable computational requirements, supporting their potential use in resource-constrained environments. The CNN-RNN hybrid model, while capturing temporal context across MRI slices, is constrained by its smaller custom dataset, and its 94.7% accuracy is likely an underestimate of performance on larger standardized datasets. VGG16 records the lowest accuracy at 93.35%, attributable to model complexity relative to available training data.

Discussion:

This review has demonstrated that recent deep learning architectures achieve strong benchmark performance for MRI-based brain tumor diagnosis, with accuracies reaching up to 99.9% on standardized datasets. However, these results are obtained under controlled experimental conditions and cannot be directly equated to radiologist-level clinical performance without prospective multi-site validation studies comparing model outputs against expert diagnoses in real-world settings.

Overfitting is a notable concern across several of the reviewed models. VGG16, with its high model complexity relative to the available training data, records the lowest accuracy at 93.35%, consistent with the well-known limitation of large architectures on smaller datasets. The CNN-RNN hybrid model, tested on a small custom dataset, similarly shows lower performance at 94.7%, suggesting its true capability may be underestimated. Models trained on smaller or less diverse datasets are susceptible to limited generalizability, as highlighted in the dataset heterogeneity analysis.

Dataset limitations also affect the interpretation of top-performing models. BraTS2020, while the most widely used segmentation benchmark in the reviewed studies, represents a curated academic collection that may not fully reflect the variability of real-world clinical MRI data including differences in scanner type, slice thickness, and acquisition protocol. The absence of cross-dataset validation in most reviewed studies means the reported performance figures should be treated as estimates under controlled conditions.

None of the compared models were validated on independently collected prospective clinical datasets. This gap between experimental benchmark performance and clinical deployment readiness is one of the most important limitations identified in this review, and it is a direction that future work must address directly.

With attention mechanisms, sophisticated feature extraction, and transfer learning, convergent advanced architectures generate synergistic effects to enable strong diagnostic capability. However, variability of datasets, large computational needs, and low interpretability remain challenges that must be addressed before these systems can be widely deployed in clinical settings.

Clinical Implications:

The findings of this review have direct implications for healthcare systems considering deployment of deep learning for brain tumor diagnosis. Based on the performance and efficiency results presented in Tables 3 and 4, YOLOv7 with the fastest inference time of 20–50 ms is most suitable for time-sensitive applications such as intraoperative real-time guidance. ResNet-50, with the lowest memory requirement of 6–8 GB and the best overall efficiency balance, is recommended for general deployment across hospital GPU infrastructure. EfficientNet, with the quickest training time of 3–6 hours, is best suited for scenarios requiring fast institutional adaptation or retraining on local patient data.

U-Net achieves the highest accuracy of 99.9% and is appropriate for academic centers with high-end hardware, where high-precision segmentation is required. However, its substantial GPU memory requirement (12–16 GB) and inference time (200–500 ms) limit its use in resource-constrained environments. For widespread clinical adoption, deep learning model outputs should be presented as decision-support information to radiologists rather than as standalone diagnostic conclusions.

Recommendations:

Based on the systematic review findings, the following structured recommendations are provided for researchers and practitioners:

For Researchers:

Prospective Clinical Validation: Future studies should conduct multi-site, prospective clinical validation comparing model outputs against expert radiologist diagnoses to bridge the gap between benchmark and real-world performance.

Cross-Dataset Generalization: Studies should evaluate models on at least two independent datasets (one benchmark, one clinical) and report domain adaptation performance to assess generalizability.

Statistical Rigor: Future studies should report confidence intervals and use k-fold cross-validation rather than single train-test splits. Where possible, meta-analytic pooling of results should be employed in systematic reviews.

Federated Learning: To address data privacy constraints in clinical settings, federated learning approaches should be explored to enable model training across multiple institutions without centralizing patient data.

Interpretability: Explainability methods (e.g., Grad-CAM, SHAP) should be systematically applied and evaluated across architectures to improve clinician trust and regulatory compliance.

For Practitioners:

Model Selection: For real-time clinical applications (e.g., intraoperative guidance), YOLOv7 is the recommended architecture based on inference speed. For general departmental deployment, ResNet50 offers the best efficiency-accuracy balance.

Implementation: Deep learning models should be deployed as clinical decision-support tools integrated with radiologist workflow, not as standalone diagnostic systems. Output should include confidence scores and uncertainty estimates.

Data Quality: Institutions should invest in standardized MRI acquisition protocols and annotated local datasets to enable effective model fine-tuning and adaptation to local patient populations.

Conclusion:

This review has demonstrated that recent deep learning architectures achieve strong benchmark performance for MRI-based brain tumor diagnosis, with accuracies reaching up to 99.5% on standardized datasets. However, these results are obtained under controlled experimental conditions and cannot be directly equated to radiologist-level clinical performance without prospective multi-site validation studies comparing model outputs against expert diagnoses in real-world settings. With attention mechanisms and sophisticated feature extraction, convergent advanced architectures, and transfer learning, the convergence generates synergistic effects to enable unprecedented diagnostics capability.

Future Work.

Future development requires interdisciplinary cooperation involving computer scientists, radiologists and clinical researchers. Emerging directions such as transformer architectures, federated learning, multimodal integration, and self-supervised approaches have the potential to further improve model performance and robustness. However, thorough clinical and large-scale validation, regulatory approval, and seamless integration with existing clinical workflows represent important steps toward the widespread adoption of these technologies.

References:

[1] S. Kar and Pawan Kumar Singh, "MBTC-Net: Multimodal brain tumor classification from CT and MRI scans using deep neural network with multi-head attention mechanism," *Med. Nov. Technol. Devices*, vol. 27, p. 100382, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590093525000335

[2] C. G. & K. S. Bhuvaneswari, "A novel approach for the detection of brain tumor and its classification via independent component analysis," *Sci. Reports Vol.*, vol. 15, 2025, [Online]. Available: https://www.nature.com/articles/s41598-025-87934-4

[3] R. B. V. & L. K. Pappala, "Enhanced brain tumor classification framework using deep learning," *Sci. Rep.*, vol. 15, 2025, [Online]. Available: https://www.nature.com/articles/s41598-025-19882-y

[4] H. Liu, Z. Ni, D. Nie, D. Shen, J. Wang, and Z. Tang, "Multimodal Brain Tumor Segmentation Boosted by Monomodal Normal Brain Images," *IEEE Trans. Image Process.*, vol. 33, pp. 1199–1210, 2024, doi: 10.1109/TIP.2024.3359815.

[5] A. A. Vikram Verma, "Deep Learning: A Revolutionizing Approach To Brain Tumor Classification Using MRI," *South East. Eur. J. Public Heal.*, pp. 2955–2965, 2025, [Online]. Available: https://www.seejph.com/index.php/seejph/article/view/5595

[6] Mohammed A Alsuwaiket, Hafar Al Batin, "Feature Extraction of EEG Signals for Seizure Detection Using Machine Learning Algorthims," *Eng. Technol. Appl. Sci. Res.*, vol. 12, no. 5, 2022, [Online]. Available: https://etasr.com/index.php/ETASR/article/view/5208

[7] B. D. S. & M. A. S. Saravanan Srinivasan, Divya Francis, Sandeep Kumar Mathivanan, Hariharan Rajadurai, "A hybrid deep CNN model for brain tumor image multi-classification," *BMC Med. Imaging*, vol. 24, no. 21, 2024, [Online]. Available: https://link.springer.com/article/10.1186/s12880-024-01195-7

[8] N. Rasool and J. I. Bhat, "Brain tumour detection using machine and deep learning: a systematic review," *Multimed. Tools Appl. 2024 8413*, vol. 84, no. 13, pp. 11551–11604, May 2024, doi: 10.1007/S11042-024-19333-2.

[9] A. T. Islam *et al.*, "An Efficient Deep Learning Approach to detect Brain Tumor Using MRI Images," *Proc. 2022 25th Int. Conf. Comput. Inf. Technol. ICCIT 2022*, pp. 143–147, 2022, doi: 10.1109/ICCIT57492.2022.10054999.

[10] "(PDF) Brain Tumor Segmentation using Deep Learning Architectures like U-NET, V-NET, U-NET." Accessed: May 30, 2026. [Online]. Available: https://www.researchgate.net/publication/373739521_Brain_Tumor_Segmentation_using_Deep_Learning_Architectures_like_U-NET_V-NET_U-NET

[11] T. Das, D. K. Vishwakarma, and V. Ranga, "A Comparative Study of Deep Learning Models for Brain Tumor Classification," *2024 4th Int. Conf. Intell. Technol. CONIT 2024*, 2024, doi: 10.1109/CONIT61985.2024.10626265.

[12] M. T. Y. Usman Humayun, "Deep Learning Approaches for Brain Tumor Detection and Segmentation in MRI Imaging | Journal of Computing & Biomedical Informatics," Journal of Computing & Biomedical Informatics. Accessed: Jan. 08, 2026. [Online]. Available: https://jcbi.org/index.php/Main/article/view/738

[13] M. M. Carlos Aumente-Maestro, "AUDIT: An open-source Python library for AI model evaluation with use cases in MRI brain tumor segmentation," *Comput. Methods Programs Biomed.*, vol. 271, p. 108991, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260725004080

[14] K. V. K. M. Vinitha, "Integrated U-Net 3+ with Decomposed Multi head Self-Attention Block and Filter Response Normalization for Brain Lesion Localization and Tracking using 3D MRI," *ITM Web Conf.*, 2025, [Online]. Available: https://www.itm-conferences.org/articles/itmconf/ref/2025/10/itmconf_keis2025_01057/itmconf_keis2025_01057.html

[15] E. K. Rutoh, Q. ZhiGuang, J. C. Bore-Norton and N. Bahadar, "ABI-Net: Attention-

Based Inception U-Net for Brain Tumor Segmentation From Multimodal MRI Images," *IEEE Access*, vol. 3, pp. 134898–134916, 20251, [Online]. Available: https://ieeexplore.ieee.org/document/11071545

[16] A. N. Mohammed Rasool, "Brain Tumor Classification using Deep Learning: A State-of-the-Art Review," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 5, pp. 16586–16594, 2024, [Online]. Available: https://etasr.com/index.php/ETASR/article/view/8298

[17] N. A. I. Mohammed Rasool, "A Novel Approach for Classifying Brain Tumours Combining a SqueezeNet Model with SVM and Fine-Tuning," *Electronics*, vol. 12, no. 1, p. 149, 2023, doi: https://doi.org/10.3390/electronics12010149.

[18] H. K. & S. R. N. K. Soheila Saeedi, Sorayya Rezayi, "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 16, 2023, [Online]. Available: https://link.springer.com/article/10.1186/s12911-023-02114-6

[19] M. U. A. Muhannad Faleh Alanazi, "Brain Tumor/Mass Classification Framework Using Magnetic-Resonance-Imaging-Based Isolated and Developed Transfer Deep-Learning Model," *Sensors*, vol. 22, no. 1, p. 372, 2022, doi: https://doi.org/10.3390/s22010372.

[20] S. H. K. Mirza Mumtaz Zahoor, "Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN," *Biomedicines*, vol. 12, no. 7, p. 1395, 2024, doi: https://doi.org/10.3390/biomedicines12071395.

[21] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of Brain Tumor from Magnetic Resonance Imaging Using Vision Transformers Ensembling," *Curr. Oncol. 2022, Vol. 29, Pages 7498-7511*, vol. 29, no. 10, pp. 7498–7511, Oct. 2022, doi: 10.3390/CURRONCOL29100590.

[22] M. A. M. Shaimaa E. Nassar, Ibrahim Yasser, Hanan M. Amer, "Comparative analysis of vision transformers and fine-tuned transfer learning models for brain tumor classification," *Imaging Radiat. Res.*, 2024, [Online]. Available: https://systems.enpress-publisher.com/index.php/IRR/article/view/8514

[23] M. G. J. Nivea Kesav, "Efficient and low complex architecture for detection and classification of Brain Tumor using RCNN with Two Channel CNN," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6229–6242, 2022, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157821001221

[24] Akmalbek Bobomirzaevich Abdusalomov, "Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging," *Cancers (Basel).*, vol. 15, no. 16, p. 4172, 2023, doi: 10.3390/cancers15164172.

[25] W. Zhang, W. Jin, S. Rho, F. Jiang, and C. fu Yang, "A Federated Learning Framework for Brain Tumor Segmentation Without Sharing Patient Data," *Int. J. Imaging Syst. Technol.*, vol. 34, no. 4, Jul. 2024, doi: 10.1002/IMA.23147.

[26] B. Kadirvelu, L. Stumpf, S. Waibel, and A. A. Faisal, "Speaker-independent dysarthria severity classification using self-supervised transformers and multi-task learning," *PLOS Digit. Heal.*, vol. 4, no. 11, p. e0001076, Nov. 2025, doi: 10.1371/JOURNAL.PDIG.0001076.

[27] Chengcheng Jin, Nor Safira Elaina Mohd Noor, "Transformer-based architectures in MRI brain tumor segmentation: A review," *Comput. Med. Imaging Graph.*, vol. 129, p. 102729, 2026, doi: https://doi.org/10.1016/j.compmedimag.2026.102729.