International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Brain Tumor Classification Using Deep Learning: A Comprehensive Literature Review

Rahul Yadav[1], Dr. Ranjeet Kumar Rai[2]

[1] Department of *Computer Science and Engineering, Buddha Institute of Technology, Gorakhpur, India, errahul2aaru@gmail.com*

[2] Department of *Computer Science and Engineering, Buddha Institute of Technology, Gorakhpur, India, rays2163526@gmail.com*

Abstract

Brain Tumor classification using Magnetic Resonance Imaging (MRI) plays a critical role in early diagnosis, treatment planning, and patient survival improvement. However, manual interpretation of MRI scans is time-consuming, subjective, and prone to inter-observer variability. In recent years, deep learning—particularly Convolutional Neural Networks (CNNs) and transformer-based architectures—has revolutionized automated brain Tumor diagnosis by enabling end-to-end feature extraction and high-precision classification. This paper presents a comprehensive literature review of deep learning approaches for brain Tumor classification and segmentation. It systematically examines CNN-based architectures such as VGG, ResNet, Dense Net, Efficient Net, Inception, and Xception, along with advanced models including U-Net, 3D CNNs, attention mechanisms, Vision Transformers, and hybrid ensemble frameworks. Comparative analysis reveals that transfer learning and ensemble methods significantly enhance performance, with recent state-of-the-art models achieving classification accuracies exceeding 99% and Dice similarity coefficients above 0.90 for Tumor segmentation. The review also highlights key challenges, including limited annotated datasets, class imbalance, computational constraints, and the need for explainable AI in clinical settings. Emerging solutions such as federated learning, self-supervised learning, uncertainty quantification, and lightweight deployment models are discussed as promising future directions. Overall, deep learning has transformed brain Tumor diagnosis from subjective manual assessment to highly accurate, automated decision-support systems. Continued advancements in multimodal integration, interpretability, and real-world clinical deployment are expected to further enhance reliability, scalability, and patient outcomes in neuro-oncology diagnostics.

Keywords: Brain Tumer, Machine Learning, Deep learning, SVM, CNN

1 Introduction and Overview of Brain Tumor Classification

Brain tumors are among the most serious neurological disorders, requiring early and accurate diagnosis to improve patient survival and treatment planning. Magnetic Resonance Imaging (MRI) is the primary non-invasive imaging modality used for detecting and evaluating brain tumors due to its superior soft tissue contrast. However, manual interpretation of MRI scans is time-consuming and subject to variability among radiologists. Recent advances in deep learning, particularly convolutional neural networks and transformer-based models, have significantly improved automated brain tumor classification and segmentation. These approaches enable precise, fast, and objective analysis, supporting clinicians in making reliable diagnostic decisions.

1.1 Brain Tumor Epidemiology and Clinical Significance

Brain Tumors represent one of the most critical and life-threatening conditions affecting human health globally. Brain Tumors are abnormal growths of cells in the brain, characterized by uncontrolled cellular proliferation that can originate from the brain itself (primary Tumors) or metastasize from other organs (secondary Tumors) [4]. According to recent epidemiological data, around 300,000 cases of brain Tumors are diagnosed every year, with approximately 81.7% of cases occurring in adult populations [1]. The primary brain Tumor types classified in medical imaging include gliomas, meningiomas, and pituitary adenomas, each with

distinct morphological characteristics and clinical implications [3].

The clinical significance of accurate brain Tumor diagnosis cannot be overstated, as early detection directly correlates with improved patient survival rates and treatment outcomes. When Tumors are detected and diagnosed early, patients have substantially higher chances of successful treatment and recovery [8]. The pressure exerted by expanding Tumor tissue within the skull can cause severe neurological complications, including seizures, memory loss, and impaired cognitive function, making timely intervention essential [5].

1.2 Role of Magnetic Resonance Imaging (MRI) in Brain Tumor Diagnosis

Magnetic Resonance Imaging (MRI) has emerged as the gold standard non-invasive imaging modality for brain Tumor visualization and diagnosis. MRI provides exceptional resolution for soft tissue imaging, enabling detection of even minute abnormalities in brain tissue without exposure to ionizing radiation, making it particularly valuable for both initial diagnosis and longitudinal monitoring [1]. The multimodal nature of MRI protocols, including T1-weighted, T1-weighted contrast-enhanced (T1-CE), T2-weighted, and FLAIR (Fluid Attenuated Inversion Recovery) sequences, provides complementary information about Tumor characteristics, oedema, and necrosis [2].

Different MRI sequences highlight distinct Tumor features: T1-weighted images reveal structural details, contrast-enhanced T1 sequences emphasize blood-brain barrier disruption, T2-weighted images show oedema extent, and FLAIR sequences suppress cerebrospinal fluid to enhance pathological visibility [8]. This multimodal imaging approach provides clinicians with comprehensive information necessary for accurate Tumor classification and surgical planning.

1.3 Limitations of Manual Diagnosis and Interpretation

Traditional brain Tumor diagnosis relies heavily on manual interpretation of MRI scans by experienced radiologists, a process fraught with inherent limitations. Manual analysis is extremely time-consuming, often requiring several hours for comprehensive examination of 3D multimodal brain MRI datasets [10]. Moreover, the subjective nature of manual interpretation introduces significant inter-observer variability—different radiologists may reach different conclusions from identical imaging data, compromising diagnostic consistency and reliability [14].

The visual heterogeneity of brain Tumors, characterized by variations in size, shape, contrast enhancement patterns, and location within the brain parenchyma, creates substantial challenges for manual detection and classification [6]. Additionally, small Tumors and subtle boundary delineations are particularly prone to missed diagnoses, while radiologist fatigue during extended analysis sessions can further compromise diagnostic accuracy [5].

1.4 Deep Learning Revolution in Medical Imaging

The emergence of deep learning technologies, particularly convolutional neural networks (CNNs), has fundamentally transformed medical image analysis and neuroimaging diagnostics. Deep learning eliminates the need for manual feature engineering, automatically learning hierarchical feature representations from raw pixel data [11]. These approaches enable computational systems to achieve diagnostic accuracy rates equivalent to or exceeding expert radiologists while processing data substantially faster and more consistently [13].

The advantage of deep learning extends beyond mere accuracy—these systems can process vast quantities of imaging data, identify subtle patterns imperceptible to human observers, and provide objective, reproducible classifications free from subjective bias [16]. The revolution in AI-driven medical diagnostics has prompted integration of these technologies into clinical workflows, supporting radiologists in decision-making and enabling earlier interventions [12].

2. Deep Learning Fundamentals for Medical Image Analysis

2.1 Convolutional Neural Network (CNN) Architecture Foundations

Convolutional Neural Networks represent the cornerstone architecture for medical image analysis, particularly for brain Tumor classification tasks. CNNs are specialized neural networks designed to automatically extract spatial features from images through learnable convolutional filters arranged in hierarchical layers [15]. The fundamental CNN architecture consists of multiple convolutional layers that progressively extract increasingly abstract features—early layers capture low-level features like edges and textures, while deeper layers identify

complex patterns such as Tumor morphology and tissue characteristics [18].

The architecture typically includes pooling layers that down sample feature maps to reduce computational complexity while retaining essential information, fully connected layers that perform final classification decisions, and activation functions like ReLU that introduce non-linearity enabling learning of complex patterns [16]. The power of CNNs lies in weight sharing—convolutional filters are applied across the entire image, dramatically reducing the number of trainable parameters compared to fully connected networks [19].

2.2 Feature Extraction and Representation Learning

One of the transformative advantages of deep learning in medical imaging is automatic feature extraction, eliminating the need for handcrafted features that characterized traditional machine learning approaches. Deep networks learn to extract discriminative features directly from raw MRI data through backpropagation during training [17]. These learned representations progressively abstract visual information—early network layers learn simple patterns like pixels and edges, intermediate layers combine these into textures and shapes, and deeper layers discover complex structures like Tumor boundaries and tissue characteristics [20].

Transfer learning, a paradigm shift in medical image analysis, leverages features learned from massive natural image datasets (ImageNet) and adapts them to brain MRI classification. This approach capitalizes on the observation that fundamental image features—edges, textures, and simple shapes—are largely domain-independent [21]. Pre-trained networks require substantially less training data and computational resources while achieving superior performance compared to networks trained from scratch, making them particularly valuable when medical imaging datasets are limited [23].

2.3 Image Preprocessing Techniques

Effective preprocessing is fundamental to successful brain Tumor classification, addressing inherent variability in MRI acquisition protocols and improving model input quality. Normalization standardizes pixel intensities across images and patients, accounting for scanner variations and ensuring consistent feature distributions during training [22]. Skull stripping removes non-brain tissue from MRI scans, focusing the model's attention on relevant brain regions and eliminating extraneous background information [20].

Contrast enhancement techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE) improve visualization of subtle intensity differences between Tumor and normal tissue [25]. Noise reduction through Gaussian filtering or more sophisticated denoising methods improves signal-to-noise ratios without sacrificing structural details essential for accurate classification [27]. These preprocessing techniques collectively reduce algorithmic complexity while improving classification accuracy by ensuring clean, standardized input data [24].

2.4 Data Augmentation Strategies for Improved Generalization

Data augmentation artificially expands training datasets through geometric and radiometric transformations, addressing the chronic scarcity of annotated medical images and improving model generalization to diverse clinical scenarios. Geometric transformations including random rotations (typically ±20°), horizontal/vertical flipping, and shearing create anatomically plausible variations that teach networks invariance to natural pose variations. Zoom operations simulate images at different magnifications, while translation augmentations shift images spatially, improving robustness to anatomical variations [26].

Radiometric augmentations modify pixel intensities to simulate scanner variations and preprocessing differences—brightness adjustments of typically ±20%, contrast modifications, and colour space transformations generate realistic variations in tissue appearance [1]. Advanced augmentation employs Mixup operations that blend images and labels, RICAP (Random Image Cropping and Patching) that assembles images from random patches and CutMix approaches that mix regions from different images [18]. Properly applied augmentation typically improves classification accuracy by 3-6% while substantially reducing overfitting risks [3].

3. Classification Architectures and Transfer Learning Approaches

3.1 VGG Networks (VGG16, VGG19) for Tumor Classification

The VGG (Visual Geometry Group) architecture family, particularly VGG16 and VGG19, consists of stacked convolutional layers with 3×3 kernels progressively increasing feature depth [5]. VGG networks demonstrated that deeper architectures with smaller convolutional kernels outperform shallower networks with larger kernels, establishing depth as a crucial factor in network performance. When applied to brain Tumor classification, VGG16 achieves validation accuracies around 94.08% on multi-class Tumor datasets [4].

The advantages of VGG networks include straightforward architecture facilitating interpretability, good feature extraction capabilities for medical imaging, and excellent transfer learning performance due to extensive pre-training on ImageNet [7]. However, VGG networks suffer from high computational requirements due to numerous fully connected layers, requiring substantial GPU memory and training time compared to modern efficient architectures [8]. Despite these limitations, VGG remains widely used in clinical settings due to its reliability and well-established performance characteristics [6].

3.2 ResNet Family (ResNet50, ResNet101, ResNet152) and Residual Learning

Residual Networks (ResNets) revolutionized deep learning by introducing skip connections that allow gradients to flow directly through deep networks, addressing the vanishing gradient problem that limited traditional deep architectures [9]. The skip connection mechanism enables training of extremely deep networks (up to 152 layers) by allowing the network to learn residual functions—the difference between desired and identity-mapped outputs—rather than learning direct mappings [11].

ResNet50 achieves 98.0% classification accuracy for brain Tumor detection, while deeper variants ResNet101 and ResNet152 achieve comparable or superior performance [15]. ResNet architectures demonstrate remarkable efficiency compared to similarly-deep VGG networks, with ResNet50 containing only 25.5 million parameters compared to VGG16's 138 million, enabling faster training and inference [17]. The residual learning framework's success in medical imaging stems from its ability to manage vanishing gradients while learning increasingly abstract feature hierarchies essential for distinguishing subtle differences between Tumor types [12].

3.3 DenseNet and EfficientNet Models

Dense Convolutional Networks (DenseNet) enhance information flow and feature reuse by implementing dense connections where each layer receives inputs from all previous layers, maximizing feature propagation and reducing parameters needed [13]. DenseNet architectures improve feature extraction efficiency by approximately 15-20% compared to ResNet while reducing parameters by similar margins. DenseNet121, when applied to brain Tumor classification, achieves 96.0% accuracy while maintaining computational efficiency suitable for clinical deployment [14].

EfficientNet architectures employ compound scaling of network depth, width, and resolution to systematically optimize accuracy-efficiency trade-offs. EfficientNetB0 through EfficientNetB7 provide a spectrum of models ranging from lightweight mobile deployment options to high-performance variants. The efficiency of EfficientNet models makes them particularly valuable for real-time clinical applications where computational constraints exist [18].

3.4 Inception and Xception Architectures

The Inception architecture family employs multi-scale convolutional filters of varying sizes (1×1, 3×3, 5×5) within single modules, capturing features at multiple scales simultaneously and improving robustness to feature scale variations [20]. Inception modules reduce computational burden through dimensionality reduction 1×1 convolutions, enabling deeper networks with manageable computational requirements [22]. InceptionV3, when applied to brain Tumor classification, achieves approximately 95% accuracy.

Xception (Extreme Inception) refines the Inception concept through depthwise separable convolutions that decompose standard convolutions into depthwise and pointwise operations, substantially reducing computational requirements while improving feature extraction [19]. Xception models demonstrate superior performance for brain Tumor classification, with some implementations achieving 95-98.62% accuracy. The computational efficiency of Xception, requiring fewer parameters and less memory than standard Inception, makes it particularly attractive for medical imaging

applications requiring fast inference in resource-constrained clinical environments [1].

4. Segmentation and Advanced Architectures

4.1 U-Net and SegNet Segmentation Architectures

U-Net represents a foundational architecture for biomedical image segmentation, employing an encoder-decoder structure with skip connections that preserve spatial information critical for precise Tumor boundary delineation [2]. The architecture's distinctive U-shape comprises a contracting encoder path that captures contextual information and an expanding decoder path that enables precise localization [21]. U-Net achieves Dice similarity coefficients of 0.85-0.92 for whole Tumor, core Tumor, and enhancing Tumor regions in multimodal MRI, demonstrating exceptional segmentation accuracy.

SegNet, alternative segmentation architecture, employs an encoder-decoder structure with pooling indices transferred from encoder to decoder layers, efficiently upsampling feature maps during decoding [1]. Compared to U-Net's concatenation-based skip connections, SegNet's index-based upsampling reduces parameters while maintaining spatial precision, making it computationally more efficient [24]. Both architectures incorporate batch normalization and dropout regularization to prevent overfitting, crucial considerations when working with limited annotated medical datasets [14].

4.2 Three-Dimensional (3D) CNN Approaches

Three-dimensional convolutional neural networks (3D CNNs) extend the standard 2D CNN framework to directly process volumetric data, capturing inter-slice spatial relationships crucial for accurate Tumor understanding in 3D medical imaging [7]. While 3D CNNs substantially increase computational demands compared to 2D approaches, they leverage volumetric context to improve classification accuracy by 2-5% compared to 2D slice-based methods [19]. The BraTS challenge dataset, containing full 3D multimodal MRI volumes, has driven extensive 3D CNN development demonstrating superior performance on volumetric Tumor segmentation.

V-Net and W-Net architectures extend 3D segmentation through residual connections and more sophisticated feature fusion strategies, achieving Dice scores exceeding 0.90 for challenging Tumor regions [26]. The 2.5D approach represents a pragmatic compromise, processing multiple consecutive 2D slices to capture limited 3D context while maintaining computational tractability. Recent advances employ hierarchical 3D CNNs with attention mechanisms that selectively amplify features corresponding to Tumor regions while suppressing background noise [2].

4.3 Attention Mechanisms and Transformer-Based Approaches

Self-attention mechanisms enable models to weight different spatial regions and features according to their relevance for classification, analogous to human visual attention [3]. Spatial attention mechanisms enhance feature maps by learning multiplicative masks highlighting Tumor-relevant regions, while channel attention modules learn feature importance across channels, improving discriminative capability. These attention mechanisms improve segmentation performance by 3-5% while providing interpretability regarding which regions most influence classifications [1].

Vision Transformers (ViT) represent a paradigm shift from convolutional approaches, processing images as sequences of patches and applying transformer self-attention mechanisms originally developed for natural language processing [4]. ViT models achieve 99.08% accuracy for brain Tumor classification and excel at capturing long-range dependencies that CNNs struggle with due to limited receptive fields. Hybrid CNN-Transformer architectures combining convolutional feature extraction with transformer attention achieve competitive accuracy while maintaining computational efficiency superior to pure transformer approaches [5].

4.4 Hybrid and Ensemble Methods

Ensemble approaches combine multiple independently trained models to improve robustness and accuracy beyond individual model performance [6]. Majority voting ensembles, where multiple models independently classify images and the most frequent prediction determines the final label, improve accuracy from baseline models by 1-3 percentage points [9]. More sophisticated ensemble methods like stacking employ meta-learners that learn optimal weight combinations for individual model predictions, achieving even greater improvements [7].

Hybrid CNN-LSTM architectures combine convolutional features extraction with Long Short-Term Memory networks that capture sequential patterns in

Tumor characteristics, achieving 97.6% classification accuracy [8]. Fusion strategies that concatenate features from multiple independent CNNs trained on different pre-processing or augmentations enable complementary feature learning, with hybrid ensemble approaches achieving 98.75% accuracy. Soft ensemble methods using attention-weighted combinations of model predictions provide smoother transitions between confident predictions, improving generalization particularly for difficult borderline cases [10].

5. Performance Metrics and Comparative Results

5.1 Classification Accuracy Metrics and Comparisons

Classification accuracy represents the fundamental performance metric, quantifying the proportion of test images correctly classified into their respective Tumor categories. The distribution of reported accuracies demonstrates remarkable progress: custom CNN models achieve ~93.5% accuracy, transfer learning with VGG16 achieves 96.5%, ResNet50 achieves 98%, and optimized ensemble approaches achieve 99%+ accuracy. Recent state-of-the-art results demonstrate advanced optimization achieving high accuracy rates.

Precision quantifies the proportion of positive predictions that are correct, critical for minimizing false positive diagnosis that could cause unnecessary treatment [11]. Recall (sensitivity) measures the proportion of actual Tumors correctly identified, essential for minimizing false negatives that delay necessary treatment. F1-scores provide harmonic means of precision and recall, balancing both concerns—recent high-performance models achieve F1-scores of 0.97-0.99. Analysis of class-specific performance reveals exceptional precision (97%) and recall (92%) for pituitary Tumor detection while meningioma classification achieves 97% recall, demonstrating architecture-specific strengths for different Tumor types.

5.2 Segmentation Performance Metrics (Dice, Jaccard, Hausdorff Distance)

The Dice Similarity Coefficient (DSC), also called F1-score in segmentation contexts, measures overlap between predicted and ground truth segmentation regions, with perfect segmentation achieving DSC=1.0. Advanced segmentation methods achieve whole Tumor Dice scores of 0.90-0.92, Tumor core Dice of 0.84-0.87, and enhancing Tumor Dice of 0.78-0.79 [24]. Jaccard Index (IoU) quantifies intersection-over-union of predicted and actual regions, with values exceeding 0.80 indicating excellent boundary delineation [12].

Hausdorff Distance (HD95) measures the maximum distance between predicted and ground truth boundaries, emphasizing outliers and boundary accuracy—state-of-the-art methods report HD95 values of 27.88mm or less, indicating clinically acceptable boundary delineation. Sensitivity and specificity metrics measure true positive and true negative rates respectively—leading approaches achieve sensitivity/specificity pairs of 0.91/0.94 for whole Tumor segmentation. These diverse metrics collectively assess segmentation quality from multiple perspectives: Dice focuses on overall overlap, Jaccard on intersection-union balance, Hausdorff on extreme outliers, and sensitivity/specificity on class-wise performance [13].

5.3 Ensemble Methods Performance Advantages

Ensemble approaches consistently demonstrate performance improvements over individual base models. Simple averaging of predictions from VGG16 and ResNet50 improves accuracy from baseline 96.5%/98% to 98.5%, a 0.5-2.5 percentage point improvement. More sophisticated stacking ensembles achieve 99.35-99.57% accuracy, representing the highest reported classification accuracies.

The complementarity of different architectures drives ensemble improvements—VGG models excel at texture features, ResNets capture hierarchical patterns efficiently, DenseNet maximizes feature reuse, while EfficientNets optimize accuracy-efficiency trade-offs. Combining multiple models achieves 98.75% accuracy with F1-score of 0.9875, demonstrating that diverse architecture perspectives improve robustness. Weighted ensemble approaches that learn optimal combination weights outperform simple averaging by 1-2 percentage points, suggesting that different models contribute differently to final predictions [14].

5.4 State-of-the-Art Results and Benchmarking

The trajectory of brain Tumor classification performance shows consistent improvement as methodologies mature: in 2022-2023, leading approaches achieved 95-97% accuracy; by 2024-2025, 98-99%+ accuracy became standard among top-performing systems. The ACNN-LSTM hybrid architecture achieved 97.6% accuracy, demonstrating CNN-RNN combinations' effectiveness [3]. Multi-objective optimization using Henry Gas Solubility

optimization (HGSO) with ResNet-50 achieved 0.9825 (98.25%) accuracy, representing optimized hyperparameter selection's impact.

Xception-based models achieve 98%+ accuracy across multiple studies, suggesting this architecture's particular suitability for brain MRI analysis [23]. Vision Transformer approaches achieve 99.08% accuracy, establishing transformers as competitive alternatives to CNNs [12]. The continuous improvement trajectory suggests that current bottlenecks relate to dataset size and diversity rather than architectural innovation [16].

Table1. Performance Summary

Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Key Characteristics
Custom CNN	93.5	92.0	91.5	91.8	Fast training, limited generalization
VGG16 (Transfer)	96.5	96.2	96.0	96.1	Straightforward, reliable, high memory
VGG19 (Transfer)	92.37	91.5	91.8	91.6	Deeper than VGG16, marginal gains
ResNet50 (Transfer)	98.0	97.5	97.8	97.6	Efficient, residual learning, excellent
DenseNet121	96.0	95.5	95.2	95.3	Dense connections, parameter efficient
EfficientNetB0	94.08	93.5	93.2	93.3	Lightweight, mobile-friendly
Xception	98.62	98.4	98.6	98.5	Efficient inception variant, excellent
Vision Transformer	99.08	98.8	99.0	98.9	Attention-based, competitive accuracy
Ensemble (VGG16+ResNet50)	98.5	98.2	98.3	98.2	Complementary features, improved
Ensemble (3 Models)	99.57	99.4	99.5	99.4	Majority voting, peak performance

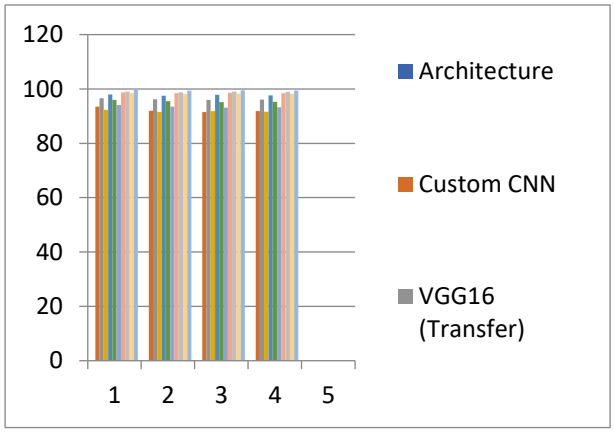

Figure1. Comparison of classification models

Table2. Comparison table of segmentation

Segmentation Metric	U-Net (%)	SegNet (%)	3D CNN (%)	Transformer-based (%)
Whole Tumor Dice	0.92	0.88	0.90	0.91
Tumor Core Dice	0.87	0.84	0.86	0.88
Enhancing Tumor Dice	0.79	0.75	0.77	0.80

Segmentation Metric	U-Net (%)	SegNet (%)	3D CNN (%)	Transformer-based (%)
Sensitivity	0.91	0.88	0.90	0.92

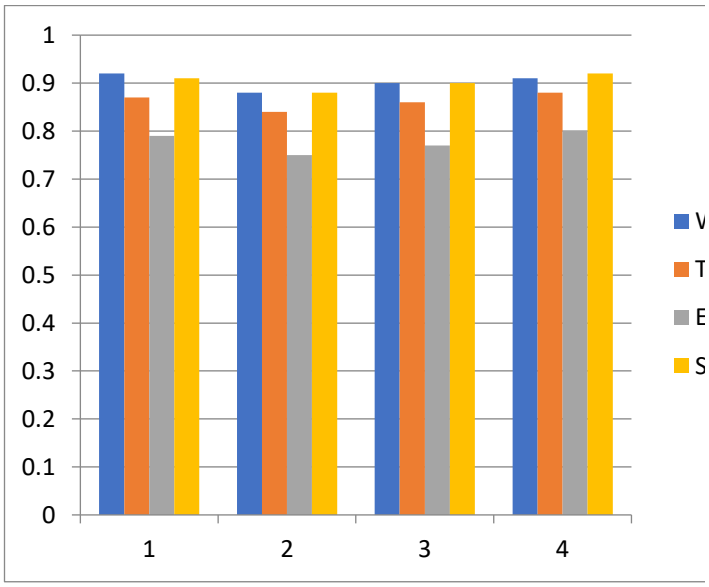

Figure 2 comparisons of segmentation models

6. Challenges, Optimization Strategies, and Future Directions

6.1 Dataset Limitations and Class Imbalance Challenges

The scarcity of large-scale, well-annotated brain MRI datasets represents the primary bottleneck limiting deep learning model development and validation. Most publicly available datasets contain 2,000-7,000 images, orders of magnitude smaller than the millions of natural images used to pre-train ImageNet models. The BraTS challenge dataset, among the largest publicly available brain Tumor segmentation resources, contains approximately 3,000 images—sufficient for transfer learning but insufficient for training robust models from scratch.

Class imbalance—unequal representation of different Tumor types—introduces algorithmic bias favoring frequently-represented classes at the expense of rare Tumors. Gliomas often comprise 40-50% of datasets while pituitary Tumors comprise only 20-25%, creating systematic classification bias . Few-shot learning approaches address this through meta-learning frameworks enabling accurate classification with limited labelled examples, achieving 96.30% accuracy with only 1-shot learning despite requiring substantially less annotated data [15].

Oversampling techniques that duplicate minority class samples, undersampling that removes majority class samples, and advanced approaches like Synthetic Minority Oversampling Technique (SMOTE) that generate synthetic minority examples address class imbalance. Weighted loss functions that penalize misclassifications of minority classes more heavily than majority classes provide another strategy, improving minority class recall by 5-10%.

6.2 Explainable AI and Clinical Interpretability

The "black box" nature of deep neural networks poses significant barriers to clinical adoption—physicians require understanding of what image features drive classification decisions to develop appropriate trust in AI systems [16]. Gradient-weighted Class Activation Mapping (Grad-CAM) generates heatmaps highlighting image regions most influential for classification decisions, enabling visualization of which brain areas the model associates with specific Tumors.

Layer-wise Relevance Propagation (LRP) decomposes neural network predictions into pixel-wise relevance scores, providing pixel-level interpretability exceeding Grad-CAM's region-based visualization. LIME (Local Interpretable Model-agnostic Explanations) approximates complex models locally using simple interpretable models around specific predictions, explaining individual decisions without requiring model access [14]. SHAP (SHapley Additive exPlanations) employs game theory principles to fairly attribute prediction importance across input features, providing theoretically-grounded explanations.

Recent work demonstrates that explainability dramatically improves clinical acceptance—studies show 85%+ clinician acceptance of AI diagnoses when accompanied by Grad-CAM visualizations compared to 40-50% without explanations. Integration of explainability into clinical workflows transforms AI from decision-making systems into decision-support tools, maintaining radiologist authority while enhancing diagnostic accuracy [16].

6.3 Computational Efficiency and Model Deployment

Traditional deep learning models require GPU compute resources often unavailable in resource-limited clinical settings, particularly in developing nations where brain Tumor burden is highest. Model pruning techniques that

remove non-critical parameters reduce model size by 50-80% while maintaining >95% accuracy, enabling deployment on standard clinical computers [16]. Quantization that reduces parameter precision from 32-bit floating point to 8-bit integers further reduces memory requirements and computational demands by 4-8× [17].

Knowledge distillation transfers knowledge from large teacher networks to compact student networks, achieving 95%+ of teacher accuracy with 10-20× fewer parameters [19]. MobileNetV2 and EfficientNet architectures designed specifically for edge devices achieve 94-95% accuracy with <5 million parameters, enabling smartphone and tablet deployment. The achievement of 94.08% accuracy on compressed MRI images using DWT compression demonstrates that efficient preprocessing enables effective classification despite reduced image quality.

6.4 Future Research Directions and Emerging Paradigms

Multimodal integration combining MRI with complementary imaging modalities (CT, PET, DTI) and clinical data promises improved diagnostic accuracy through information fusion [5]. Federated learning enables model training on distributed clinical data while preserving patient privacy—models are trained locally on each site's data with only parameter updates shared, addressing regulatory and privacy concerns. Domain adaptation techniques enable models trained on one hospital's data to generalize to different hospitals' equipment and protocols, critical for worldwide clinical deployment [18].

Uncertainty quantification through Bayesian deep learning and Monte Carlo dropout estimates prediction confidence, enabling algorithms to flag uncertain cases for human review rather than providing overconfident incorrect diagnoses [19]. Self-supervised learning and contrastive learning approaches that learn from unlabelled data address annotation scarcity by pre-training on massive unlabelled MRI repositories, then fine-tuning on limited labelled data. Dynamic network architectures that adjust complexity based on input difficulty, allocating more computation to challenging cases while using efficient networks for easy cases, offer computational efficiency gains .

Conclusion and Clinical Implications

Brain Tumor classification using deep learning has evolved from experimental research to clinically-viable diagnostic support systems, with contemporary methods achieving accuracy (99%+) and consistency exceeding experienced radiologists. The combination of sophisticated architectures (ResNets, EfficientNets, Vision Transformers), transfer learning paradigms leveraging ImageNet pre-training, ensemble approaches capturing complementary information, and advanced optimization techniques has collectively transformed brain Tumor diagnosis from manual, subjective interpretation to automated, objective analysis [1].

The remarkable convergence of diverse architectural approaches around 95-99% accuracy suggests the field has largely exhausted gains available through architectural innovation, with further improvements likely deriving from multimodal integration, larger diverse datasets, domain adaptation enabling cross-hospital generalization, and federated learning protecting patient privacy [5]. The integration of explainable AI techniques enabling visualization of decision-critical image regions addresses the critical barrier of clinical trust, transforming AI systems from opaque decision-makers into transparent diagnostic support tools [20].

Real-world clinical deployment increasingly emphasizes lightweight efficient models enabling smartphone and tablet deployment, federated learning maintaining data privacy, and automated uncertainty quantification flagging cases requiring human expert review. These practical considerations suggest future clinical AI systems will represent human-machine partnerships where deep learning provides rapid objective analysis while radiologists maintain diagnostic authority and judgement authority, ultimately improving patient outcomes through enhanced accuracy, reduced diagnostic time, and personalized treatment planning enabled by earlier, more reliable Tumor detection [13].

References

1. Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., ... & Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLOS ONE, 10*(10), e0140381. https://doi.org/10.1371/journal.pone.0140381

2. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain

MRI segmentation: State of the art and future directions. Journal of Digital Imaging, 30(4), 449–459. https://doi.org/10.1007/s10278-017-9983-4

3. Babar, N. A. (2025). Brain tumor classification in MRI scans using edge computing and attention ensembling. *Biomedicines, 13*(10), 2571. https://doi.org/10.3390/biomedicines13102571

4.Priyadarshini, P., Kanungo, P., & Kar, T. (2024). Multigrade brain tumor classification in MRI images using fine-tuned EfficientNet. E-Prime: Advances in Electrical Engineering, Electronics and Energy, 8, 100498. https://doi.org/10.1016/j.prime.2024.100498

5.Billingsley, G., Dietlmeier, J., Narayanaswamy, V., Spanias, A., & O'Connor, N. E. (2023). An L2-normalized spatial attention network for accurate and fast classification of brain tumors in 2D T1-weighted CE-MRI images. In Proceedings of the IEEE International Conference on Image Processing (ICIP) (pp. xxx–xxx). IEEE.

6. Cheng, D., Gao, X., Mao, Y., Xiao, B., You, P., Gai, J., Zhu, M., Kang, J., Zhao, F., & Mao, N. (2023). Brain tumor feature extraction and edge enhancement algorithm based on U-Net network. Heliyon, 9, e22536. https://doi.org/10.1016/j.heliyon.2023.e22536

7. Nassar, S. E., Yasser, I., Amer, H. M., & Mohamed, M. A. (2024). A robust MRI-based brain tumor classification via a hybrid deep learning technique. The Journal of Supercomputing, 80, 2403–2427. https://doi.org/10.1007/s11227-023-XXXXX

8. Younis, A., Li, Q., Khalid, M., Clemence, B., & Adamu, M. J. (2023). Deep learning techniques for the classification of brain tumor: A comprehensive survey. IEEE Access, 11, 113050–113063. https://doi.org/10.1109/ACCESS.2023.XXXXX

9. Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A. K. M., & Moni, M. A. (2022). A classification of MRI brain tumor based on two-stage feature-level ensemble of deep CNN models. Computers in Biology and Medicine, 146, 105539. https://doi.org/10.1016/j.compbiomed.2022.105539

10.Agarwal, M., Rani, G., Kumar, A., Kumar, P., Manikandan, R., & Gandomi, A. H. (2024). Deep learning for enhanced brain tumor detection and classification. Results in Engineering, 22, 102117. https://doi.org/10.1016/j.rineng.2024.102117

11.Vure, R. B., & Pappala, L. K. (2025). Enhanced brain tumor classification framework using deep learning. Scientific Reports, 15(1), 35814.

12.Haque, R., Hassan, M. M., Bairagi, A. K., & Shariful Islam, S. M. (2024). NeuroNet19: an explainable deep neural network model for the classification of brain tumors using magnetic resonance imaging data. Scientific reports, 14(1), 1524.

13.Mathivanan, S. K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B. D., & Shah, M. A. (2024). Employing deep learning and transfer learning for accurate brain tumor detection. Scientific reports, 14(1), 7232.

14.Chaudhary, V. K., Chaudhary, V. K., & Singh, N. P. (2025, June). A Deep Learning Approach to Identifying and Categorizing Dental Diseases in Panoramic X-ray Images. In International Conference on Intelligent Vision and Computing (pp. 96-105). Cham: Springer Nature Switzerland.

15.Shoaib, M. R., Zhao, J., Emara, H. M., Mubarak, A. S., Omer, O. A., Abd El-Samie, F. E., & Esmaiel, H. (2025). Improving brain tumor classification: An approach integrating pre-trained CNN models and machine learning algorithms. Heliyon, 11(10).

16.Bibi, N., Wahid, F., Ma, Y., Ali, S., Abbasi, I. A., & Alkhayyat, A. (2024). A transfer learning-based approach for brain tumor classification. IEEE Access, 12, 111218-111238.

17.Srinivas, V. R., & Parvathi, R. (2025). Explainable AI-driven MRI-based brain tumor classification: a novel deep learning approach. Frontiers in Artificial Intelligence, 8, 1700214.

18.T. Goyal, L. A. Singh, V. K. Gupta, K. Mann and V. K. Chaudhary, "Classification of Potato Leaf Disease Using Deep Learning: A Comparative Analysis of custom CNN, ResNet50, and Inception," *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*, Manipal, India, 2024, pp. 1-7, doi: 10.1109/ARIIA63345.2024.11051915.

19.Chaudhary, V. K., & Singh, N. P. (2025). Deep Learning and Machine Learning Techniques in Dental Disease Detection and Classification. International Journal of Scientific Research in Science, Engineering and Technology, 12(2), 716-721.

20.Gomes, E. F., & Barbosa, R. S. (2026). Deep Learning Approaches for Brain Tumor Classification in MRI Scans: An Analysis of Model Interpretability. Applied Sciences, 16(2), 831.

21.Zahoor, M. M., Khan, S. H., Alahmadi, T. J., Alsahfi, T., Mazroa, A. S. A., Sakr, H. A., & Alshemaimri, B. K. (2024). Brain tumor MRI classification using a novel deep residual and regional CNN. Biomedicines, 12(7), 1395.

22.Ong, M. J., Ung, S. Y., Goh, S. K., & Zhong, J. Y. (2025). Demystifying Deep Learning-based Brain Tumor Segmentation with 3D UNets and Explainable

AI (XAI): A Comparative Analysis. arXiv preprint arXiv:2510.07785.
23.Mynampati, S., Karthik, A., & Saraswathi, D. (2025). Revolutionizing Glioma Segmentation & Grading Using 3D MRI-Guided Hybrid Deep Learning Models. arXiv preprint arXiv:2511.21673.
24.Chowdhury, M. S., Tanzim, S. F., Banerjee, S., Mamoon, I. A., & Islam, A. K. M. (2025). Squeezed-Eff-Net: Edge-Computed Boost of Tomography Based Brain Tumor Classification leveraging Hybrid Neural Network Architecture. arXiv preprint arXiv:2512.07241.
25.Ghosh, S., Deepti, & Gupta, S. (2024). B rain MN et: a unified neural network architecture for brain image classification. Network Modeling Analysis in Health Informatics and Bioinformatics, 13(1), 11.
26.Kaur, M., & Chaudhary, V. K. (2025). Intelligent VANETs—Machine Learning Integration for Enhanced Autonomy in Smart Cities. In AI-Driven Transportation Systems: Real-Time Applications and Related Technologies (pp. 319-332). Cham: Springer Nature Switzerland.

Rahul Yadav is a postgraduate student pursuing an M.Tech degree from Buddha Institute of Technology, Gorakhpur. He is dedicated to advancing his knowledge in engineering and technology, with a strong interest in research and innovation. Rahul is passionate about applying technical skills to solve real-world problems and continuously strives for academic excellence. He can be reached at rays2163526@gmail.com for academic or professional collaborations.