

Deep Learning-Based Multimodal Brain Tumor Analysis: A Comprehensive Review of Classification, Segmentation, Detection, and Prognosis

Asif Raza^{1*}, Muhammad Javed Iqbal¹, Zayaan Alvi¹, Muhammad Basit Umair², Altaf Khan³, and Muhammad Rmzan Shahid⁴

¹Department of Computer Science, University of Engineering and Technology Taxila, Taxila, Pakistan.

²Department of CS and IT, Thal University, Bhakkar, Pakistan.

³Department of Software Engineering, University of Mianwali, Mianwali, Pakistan.

⁴Department of Computer Science, NAMAL University Mianwali, Mianwali, Pakistan.

*Corresponding Author: Asif Raza. Email: asifraza@umw.edu.pk

Received: June 12, 2026 Accepted: August 30, 2026

Abstract: Brain tumors are still one of the most difficult diseases in neuro-oncology due to their diverse morphology, their varying grades, and the lack of clarity in demarcating tumor and normal tissues. Structural and functional information for diagnosis, treatment planning, and prognosis is provided by magnetic resonance imaging (MRI), which may be combined with computed tomography (CT), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). Over the past decade, deep learning has revolutionized brain tumour analysis by shifting the paradigm from handcrafted features and manual delineation to automated convolutional neural networks (CNNs), Vision Transformers (ViTs), and hybrid architectures that learn directly from imaging data. This paper summarizes recent advances across four related challenges: tumour classification, tumour segmentation, disease forecasting, and object detection. The literature is organized by methodological family, covering transfer learning, attention mechanisms, capsule networks, graph networks, transformer-based encoders, generative and semi-supervised learning methods, and federated and privacy-preserving approaches. We also summaries the data types, imaging modalities, and evaluation metrics commonly used to benchmark these systems, and compare reported performance trends from 2019 to 2025. Lastly, we consider persistent clinical challenges such as domain shift between institutions, missing-modality robustness, limited interpretability, and regulatory readiness, as well as emerging trends like foundation models, self-supervised pretraining, and federated diffusion-based data synthesis, that will impact the future of clinically deployed systems.

Keywords: Brain Tumor; Machine Learning; Computed Tomography; Deep Learning; Computer-Aided Diagnostic and Detection

1. Introduction

Brain tumors are abnormal growths of cells in the brain or central nervous system that are either primary (from cells normally found in the brain) or secondary (from cancer that has spread from another part of the body). Commonly investigated primary tumour types include gliomas, meningiomas, and pituitary adenomas, and the World Health Organization grading system categorizes tumours from grade I to IV based on aggressiveness and prognostic indicators of life expectancy [1]. The treatment options and prognosis vary significantly depending on the type, grade, and location of the tumor and hence a timely and accurate diagnosis is of utmost importance; however, manual interpretation of medical images by radiologists is time-

consuming, highly dependent on the individual skills of the interpreter and increasingly time-consuming as the number of medical images to be analyzed increases [1], [2].

Magnetic resonance imaging (MRI) is the most important modality in neuro-oncology because it provides excellent soft-tissue contrast resolution and can obtain complementary sequences, such as T1-weighted, contrast-enhanced T1 (T1ce), T2-weighted and Fluid-Attenuated Inversion Recovery (FLAIR), each of which emphasizes different features of the tumor anatomy and/or peritumoral edema [3]. Unimodal approaches, which depend on a single sequence or imaging technique, have limitations: for example, MRI cannot directly image metabolic activity as PET can, and no single modality is immune to noise, motion artefacts, or incomplete acquisitions [1]. Combining information from multiple MRI sequences, or from MRI, CT, PET, and SPECT, has therefore become a more reliable approach to characterizing tumour heterogeneity and increasing diagnostic confidence.

The three recent and complementary sources of evidence for this review are a wide survey of multimodal deep learning approaches to classification, segmentation, forecasting and object detection tasks, a detailed methodological review of classification and segmentation methods published in the past five years, with a large-scale protocol-aware comparison of datasets and validation strategies [15] and an empirical study that proposes a lightweight and optimized CNN for four-class brain tumor classification, directly comparing it to ResNet50, VGG19, and DenseNet121 [31]. This review provides a global overview of the state-of-the-art from a broader taxonomic perspective, identifies promising architectural families, and highlights the remaining obstacles between research prototypes and clinically viable systems.

2. Imaging Modalities and a Taxonomy of Deep Learning Tasks

MRI sequences provide complementary diagnostic information. T1 and T1ce sequences are better suited to anatomical detail and to detect blood-brain barrier disruption and areas of contrast enhancement; T2 and FLAIR sequences are better at detecting areas of oedema; and non-enhancing margins of infiltrative tumour are more easily detected with FLAIR sequences, particularly near the ventricles [15]. DWI/DTI provide data on tumour cellularity and white matter tract involvement; PWI and MRS yield data on vascularity and metabolic profile, and SWI is sensitive to microhemorrhage and calcification [15] [41][42]. Because no single sequence captures most clinically relevant tumour properties, most state-of-the-art pipelines use two or more sequences, and many combine MRI with CT, PET, or SPECT to better disambiguate tumour subregions [1].

Brain imaging deep learning research can be grouped into four general task families based on the reviewed literature: classification is used to assign categorical labels like tumor type or grade; segmentation is the delineation of the spatial extent of the tumor and its subregions; forecasting involves predicting disease progression, genotype, or patient outcome; and object detection involves locating tumors within an image using a bounding box, which are typically built on top of the architectures developed for natural-image detection, like the YOLO family [1]. The four families are not mutually exclusive, but complementary: segmentation results are used in classification, and detection is often a pre-processing step for finer-grained segmentation or classification.

Fig. 1 summarizes the principal imaging modalities, analytical tasks, model families, and clinical outputs reviewed in this article, based on the taxonomies reported in [1] and [15]. Fig. 2 illustrates an end-to-end multimodal brain tumour analysis workflow, from image acquisition and preprocessing to model inference, evaluation, and clinical outcomes, based on methodological evidence summarized in [1], [15], and [31].

3. Deep Learning for Brain Tumor Classification

Studies on brain tumour classification have evolved through several iterations, starting with classical machine learning classifiers on handcrafted features, then moving to transfer learning, hybrid feature fusion, attention-based architectures, transformer and graph-based models, and finally multimodal radiogenomic frameworks [15].

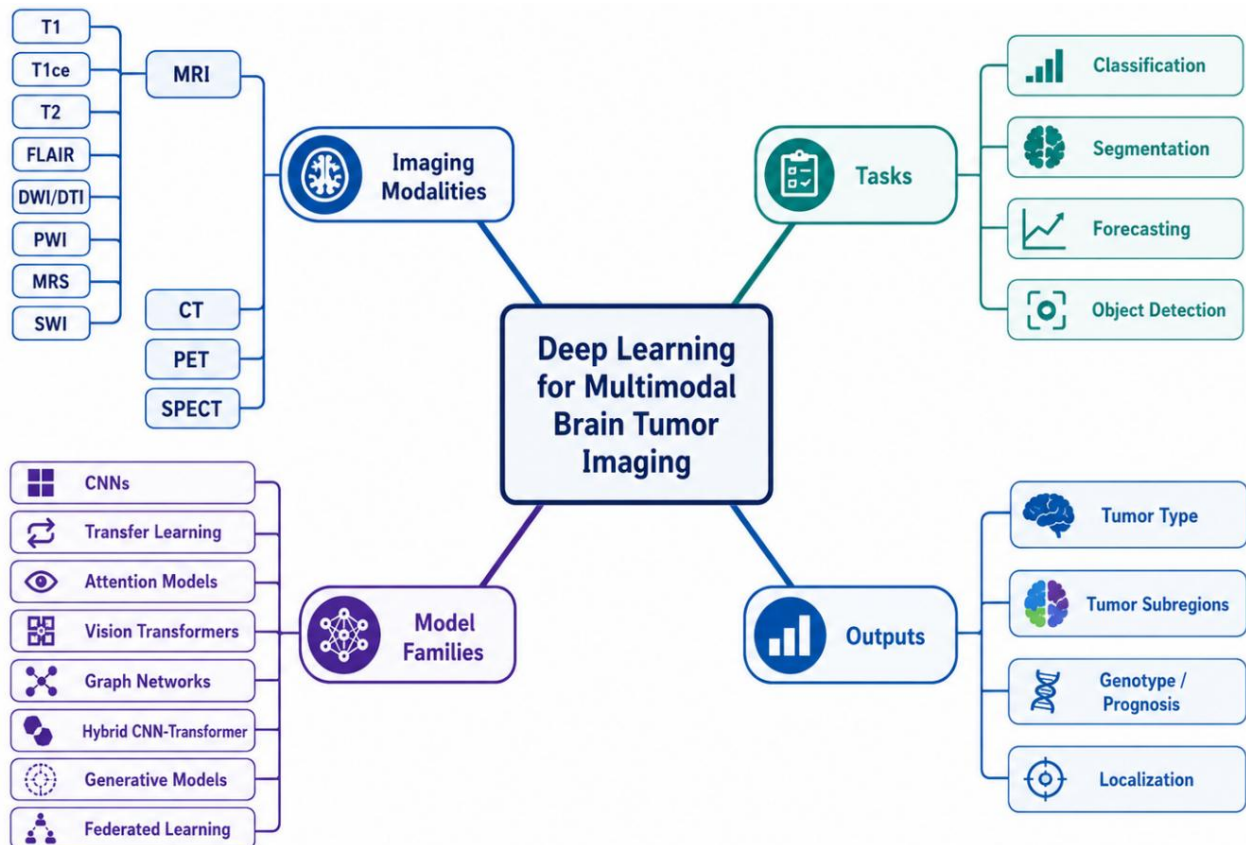


Figure 1. Taxonomy of Deep Learning for Multimodal Brain Tumor Imaging

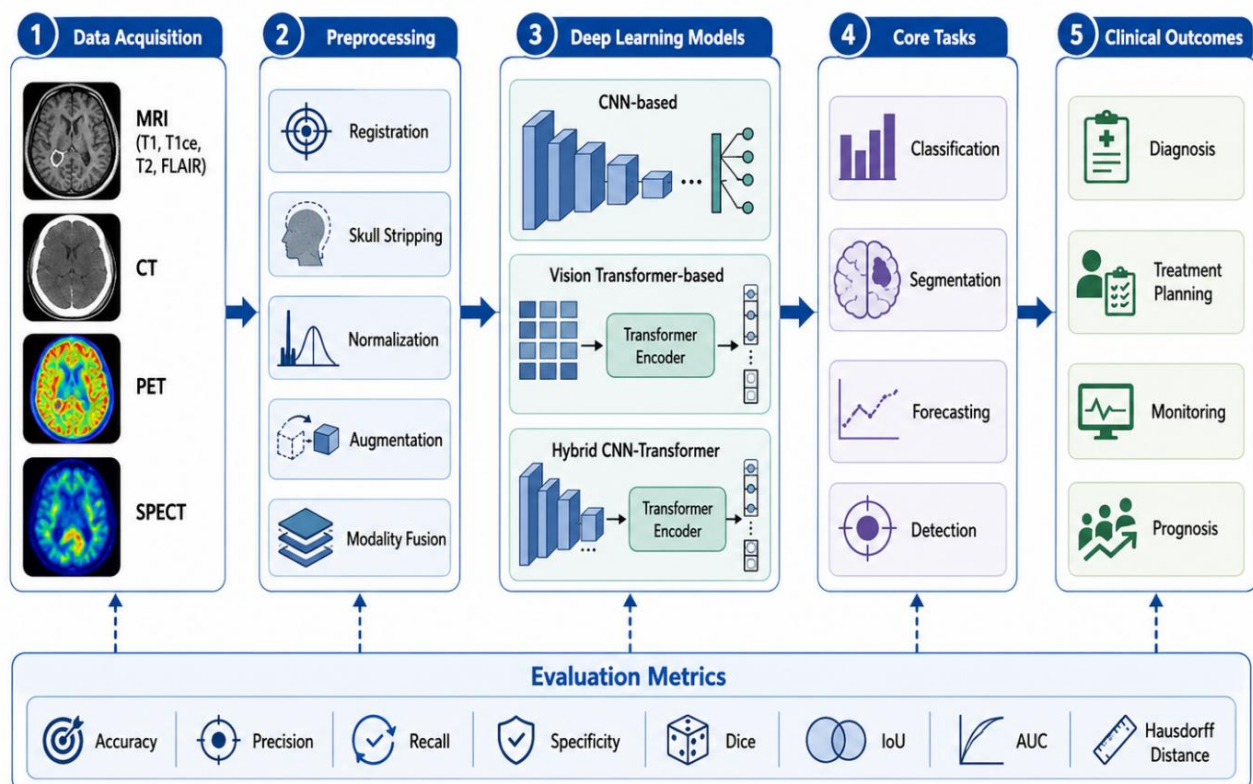


Figure 2. End-to-End Workflow for Brain Tumor Analysis

3.1. Classical Machine Learning and CNN-LSTM Hybrids

Early techniques have included handcrafted texture classifiers, including local binary patterns, gray-level co-occurrence matrices, or discrete wavelet transform (DWT) coefficient classifiers, and classifiers like support vector machines (SVM) or ensemble voting schemes, which have been used to classify glioma, meningioma, pituitary adenoma, and multiple sclerosis lesions [22]. Several studies combined CNNs with recurrent architectures to exploit sequential or temporal dependencies in feature representations, such as a cascaded multiscale deep CNN based on DWT features to classify brain tumour types, with 89.5% precision and 98.5% accuracy [32]. CNN-LSTM hybrids were further extended by incorporating an IoT system for deployment and hill-climbing hyperparameter tuning [15] to detect brain abnormalities across multiple classes. These hybrids preserve interesting temporal or sequential information. Still, they are also computationally expensive and require a large annotated dataset, which is why there is a shift toward transfer learning.

3.2. Transfer Learning with Pretrained CNNs

Transfer learning, which adapts a network by fine-tuning it or using it as a fixed feature extractor from a large natural-image dataset like ImageNet, has been a key approach to overcome limited annotated medical data [15]. Examples are the use of ResNet50 features with an SVM classifier [23] for multimodal image retrieval and classification in very deep networks; progressive block-wise fine-tuning with VGG-19 features [24] for multimodal image retrieval and classification; and the EFF_D_SVM framework, which features EfficientNetB0 features and an SVM classifier [25] for multimodal image retrieval and classification. In this direction, a new EfficientNetV2 architecture with a Global Attention Mechanism and an Efficient Channel Attention module further improved MRI-based classification accuracy using an interpretable method visualized with Grad-CAM [26]. A comparison between SVM and k-nearest neighbor classifiers and pretrained GoogLeNet features yielded an average accuracy of ~98% with five-fold cross-validation on a popular Figshare dataset [1]. However, while transfer learning can significantly save time and increase accuracy in low-data regimes, the domain gap between natural image and medical scan datasets may introduce biases that reduce cross-institution generalization [15].

3.3. Hybrid Deep Learning and Feature Fusion

Hybrid pipelines fuse deep features with classical machine learning classifiers, generative augmentation, or domain-guided architectural modifications to improve representational power and enable interpretable or data-efficient designs. BrainNet, for instance, combines deep learning and evolutionary (Grey Wolf) optimization to classify four MRI modalities (T1, T1ce, T2, and FLAIR) and uses a fusion-assisted framework composed of residual blocks cascaded with autoencoders [2]. BrainGAN directly addresses data scarcity by using a generative adversarial network and a CNN to generate brain MRI images, and reported 99.09% accuracy [3] for its generative adversarial network with ResNet152V2. A near-perfect glioma-meningioma-pituitary classification accuracy of 99% was obtained by a three-class glioma-meningioma-pituitary task on the Dual Convolutional Tumor Network (DCTN), a network that was built based on a VGG16 backbone and a custom CNN [4]. Later, DeepTumorNet, a modified GoogLeNet architecture with 15 additional layers and leaky ReLU activation, was optimized using a quantum-inspired genetic algorithm [5]. A dual-input VGG16 model that simultaneously processes original MRI images and texture-feature images extracted from the MRI using GLCM was tested and achieved an average accuracy of 96.5% [33]. In [35], a custom nineteen-layer CNN achieved 97.1% accuracy in classifying glioma, meningioma, and pituitary tumour, with only minimal image preprocessing. A similar “fused” CNN-deep and classical optimization-based decision architecture yielded exceptionally high multiclass accuracy, aligning with the broader trend of fusing CNN deep feature extractors with more efficient decision layers [21]. At this level, federated learning has also emerged, with techniques like FedAvg and FedProx that allow decentralized CNN training on brain MRI data without compromising patient privacy [15]. Such hybrid methods often incur higher computational overhead and more complex, multi-stage pipelines that are harder to deploy in real time, though they can improve accuracy [15].

3.4. Attention Mechanisms, Transformers, and Graph-Based Models

Attention mechanisms help focus network resources on clinically relevant tumor regions, improving boundary localization and subregion discrimination. Representative examples include the Multi-level Attention Network, which combines spatial and channel attention [27], and the hierarchical deformable attention module, along with saliency mapping, to model irregular tumour shapes for joint classification and segmentation [30]. Related work includes VSA-GCNN, which uses Variational Spatial Attention (VSA) with Graph Convolutional Networks (GCN), and AlexNet features with a Bi-GRU classifier for both segmentation and classification, evaluated on several BraTS releases [6]. In addition to attention-augmented CNNs, a few capsule networks, including BayesCap, added Bayesian uncertainty to tumour classification [28]. A network whose topology was randomly generated from complex graphs (CNNBCN) achieved 95.49% accuracy for tumor classification, beating manually designed ResNet, DenseNet and MobileNet baselines under matched conditions [34]. Other designs, such as Vision Transformers [30] and CNN-Transformer hybrid architectures, have also improved global context modelling, and researchers have explored graph-based methods with multi-scale graph neural networks and spatio-temporal transformer attention for tumour grading [29]. These architectures work well for spatial and relational modeling but are resource-intensive and still face bandwidth issues due to limited datasets and class imbalance [15].

3.5. Multimodal and Radiogenomic Models

Multimodal classification platforms combine MRI, histopathology, clinical metadata, or genomic data to enhance diagnostic accuracy. Examples of the trend toward integrated tumor characterization beyond a single image source include dual-modality CNNs that combine MRI and histopathology images and radiogenomic reviews that associate imaging features with molecular markers such as IDH mutation status [15]. Semi-supervised approaches that use unlabeled MRI data, together with a small labeled sample, have also been suggested to reduce annotation burden and, in imbalanced datasets, improve generalization [15]. The improved optimization-based framework, based on an Extended Sea-Horse-optimized Multi-Layer Perceptron on the BraTS-2021 dataset, achieved 98.5% accuracy, matching that later reported for the lightweight CNN (Section 3.6) [20] [36] [37] [39] [40].

4. Deep Learning for Brain Tumor Segmentation

The U-Net family and its variants have driven this field of segmentation research, along with 3D CNNs and, more recently, transformer-based and generative models that harness long-range spatial context across MRI volumes [15].

4.1. Self-Adaptive and General-Purpose Architectures

Self-configuring frameworks like nnU-Net automatically tune preprocessing, network architecture, and learning approach to the specific dataset, and have been frequently used as a baseline throughout the BraTS challenge years [16]. By combining the complementary strengths of the two networks, an ensemble of DeepSeg and nnU-Net further improved delineation of glioblastoma subregions [19]. Related work has proposed several other approaches to improve generalization, including autoencoder-based regularization [14, 15], edge-processing-based and semantic-based decoder branches [16,17], and self-adaptive weighted fusion modules that dynamically determine modality-specific fusion weights [18, 19]. These adaptive designs often perform well on example data but can increase computational complexity and reduce interpretability, which may not suit resource-limited clinical settings [15].

4.2. Enhanced U-Net Variants and Hybrid Encoders

Encoder-decoder symmetry and skip connections still make U-Net the network backbone. Dilated convolutions and spatial pyramid pooling have been added to the basic model; group cross-channel attention is added for more precise boundaries; squeeze-and-excitation blocks are added for channel-wise discrimination; one-encoder-three-decoder architectures are used to predict subregions simultaneously; and encoders based on EfficientNet are used with U-Net decoders for enhanced boundary detection with limited

annotations [15]. Strong results have also been obtained on various BraTS releases with the DEMD-ResUNet architecture, which leverages approximate brain symmetry by operating on both the original image and a mirrored version [10]. These improvements typically yield better boundary delineation but require more memory and are susceptible to imaging artefacts [15].

4.3. Transformer-Based and Attention-Guided U-Nets

Transformer-augmented segmentation networks can model long-range spatial dependencies that are difficult to model using convolutional receptive fields. In addition to self-attention layers, the popular TransUNet-based designs also use a U-Net architecture to integrate self-supervised pre-training [17]; the hierarchical Swin Transformer design used in SwinBTS [11] applies to 3D MRI volumes at multiple scales; and the dilated attention convolutions found in the popular hybrid CNN-Transformer designs expand the receptive field without sacrificing spatial resolution [15]. A generalized Wasserstein Dice loss combined with transformer layers and test-time augmentation boosted boundary calibration in the BraTS 2021 challenge [18], and lightweight sequence-modelling blocks inspired by Mamba have been introduced as an efficient alternative to self-attention [15]. Together, these transformer-based U-Nets enable scalable, anatomy-specific segmentation, but they remain computationally expensive and sensitive to training data.

4.4. Cascaded, Multimodal, and Uncertainty-Aware Networks

Many architectures explicitly account for modality variance by modelling it with multi-pathway or cascaded designs, because subregion visibility differs across MRI sequences. They include multimodal, multiscale double-pathway 3D residual CNNs with grey-level normalization and CRF post-processing; multi-pathway U-Nets with shared encoders and modality-specific decoders that decompose the segmentation task into hierarchical binary subproblems; and uncertainty-aware mutual learning frameworks that exchange reliable soft labels between 2D, 2.5D, and 3D models. Building on this, uncertainty-guided transformers use stochastic dropout masks, and semi-supervised methods use the Segment Anything Model to produce pseudo-labels when annotation is limited. These methods are robust to modality variance, but they typically require all multi-sequence inputs and are sensitive if one or more sequences are missing [15].

5. Forecasting and Object-Detection-Based Models

Beyond classification and segmentation, deep learning is now used for a variety of tasks, including genotype prediction, disease progression, and patient outcome prediction. In glioma, a multi-encoder architecture has been used to predict isocitrate dehydrogenase (IDH) mutation from limited diffusion MRI (dMRI), with self-supervised brain-network construction to address the scarcity of dMRI [12]. A similar dual-stream network estimated post-menstrual age in neonates using structural and diffusion MRI images, learning modality-specific features separately and then combining them in a self-attention module [13]. A multimodal approach using EHR and imaging was also proposed to predict LOH and death together [1]. Most object detection methods, usually adapted from the YOLO family for natural images, localize tumors with bounding boxes before or alongside segmentation. RCS-YOLO extends the YOLO model to increase inference speed and achieve higher accuracy than previous YOLO versions, using reparametrized convolution and one-shot aggregation [7]. BGF-YOLO introduces a fourth detection head, generalized feature pyramid networks, and bi-level routing attention, improving mean average precision over YOLOv8x [8]. On independent test datasets, Enhanced TumorNet, a hybrid pipeline that integrates YOLOv8s and U-Net, achieved precision, accuracy, and recall above 95% for combined tumor detection, classification, and localization [9]. In addition to separation, detection and segmentation are often performed simultaneously, as in meningioma firmness classification with YOLO [1].

6. Datasets and Evaluation Metrics

Large, well-annotated, publicly available datasets are particularly important for progress in this area. The BraTS challenge series was introduced as a standardized benchmark and multimodal brain tumour segmentation challenge [14] and remains the most widely used benchmark for this task, with multi-

institutional, multimodal, expert-annotated MRI (T1, T1ce, T2, FLAIR) provided across multiple challenge years [1,15]. Kaggle also hosts several classification research datasets, including the Figshare brain tumour dataset with 3,064 T1-weighted images with contrast enhancement from 233 patient cases [1,31] and a larger aggregated set with more than 7,000 images across four classes: glioma, meningioma, pituitary tumour, and healthy [1, 31]. TCGA-LGG and TCGA-GBM collections provide imaging, molecular, and clinical metadata and support radio genomics and outcome prediction; developmental neuroimaging studies use resources such as the Developing Human Connectome Project [1]. Table 1 summarizes some dataset families across the three source reviews.

Table 1. Summary of commonly used brain tumor imaging datasets.

Dataset family	Primary use	Approx. size	Modalities
BraTS (2012-2023)	Segmentation / classification / survival	~2,000 glioma patients	T1, T1ce, T2, FLAIR
Figshare / CJDATA	Classification, segmentation	3,064 images / 233 patients	T1-weighted contrast-enhanced
Kaggle Brain Tumor MRI	4-class classification	3,264-7,023 images	T1-weighted MRI
TCGA-LGG / TCGA-GBM	Classification, radiogenomics	~350 subjects	T1, T1ce, T2, FLAIR + genomics
UCSF-PDGM (TCIA)	Segmentation, radiogenomics	495 patients	T1, T1ce, T2, FLAIR, DWI, ASL
Developing Human Connectome Project	Forecasting (brain age)	887 datasets / 783 infants	sMRI, dMRI, rs-fMRI
Br35H	Object detection	701 images	MRI

Evaluation practices differ between classification and segmentation studies. Classification performance is typically reported using accuracy, precision, and recall (sensitivity). In contrast, segmentation performance is reported using accuracy, sensitivity, and specificity, and overlap-based measures such as the Dice coefficient and Jaccard index (IoU), with Hausdorff distance as a boundary-sensitive measure [15]. Each metric has known limitations: background tissue can dominate the image and be misclassified as tumour; Dice and IoU are highly sensitive to small enhancing subregions; and Hausdorff distance can be unstable near noisy or disconnected tumour boundaries [15]. Thus, overlap-based, boundary-based, and classification metrics are typically recommended to assess the situation from a clinically meaningful perspective [15]. Another methodological point frequently discussed in the literature is validation: many cross-hospital studies validate at the slice level rather than the patient level, which generally yields an overly optimistic accuracy metric compared with what would be observed in cross-patient or cross-institution testing [15].

7. Comparative Analysis of State-of-the-Art Methods

Performance in both classification and segmentation has steadily improved from 2020 to 2025. Across dozens of studies, the mean accuracy of the classifiers has increased from around 90% in 2020 to over 99% by 2025, with the shift from only CNN-based pipelines (roughly 60-65% of the studies) to transformer-based (15-20%) and hybrid CNN-Transformer (10-15%) designs. In segmentation, mean Dice scores increased from about 0.78 in 2020 to nearly 0.94 in 2025, with a shift from CNN-dominant architectures (65-70% of studies) to transformer-based architectures (30-35%) [15]. FLAIR and T1ce sequences are the most commonly used sequences for both tasks, as they complement each other in their sensitivity to infiltrative margins and contrast-enhancing tumor cores, respectively [15]. Table 2 presents representative results from the three source reviews,

showing the variation in performance reported in each method family. Fig. 3 visualises the approximate evolution of classification accuracy, segmentation Dice scores, and dominant model families between 2020 and 2025; the values represent literature-level trends rather than a single controlled benchmark [15], [31].

Table 2. Comparative performance of representative deep learning methods.

Method family	Representative approach	Task	Reported performance
Classical ML / hybrid CNN-SVM	GoogLeNet + SVM, CNN-HOG fusion	4-class classification	~97-99% accuracy
Transfer learning (single backbone)	ResNet50, VGG19, DenseNet121	4-class classification	75-93% accuracy [31]
Lightweight optimized CNN	4-block custom CNN (Adam, cross-entropy)	4-class classification	98.5% accuracy [31]
Attention / multiscale CNN	MANet, ARM-Net	Classification	~96-99% accuracy
Transformer / hybrid CNN-Transformer	RanMerFormer, DeiT + Firefly	Classification	~98-99% accuracy
U-Net and variants	nnU-Net, GCAUNet, SEResU-Net	Segmentation (BraTS)	Dice ~0.78-0.91
Transformer-based U-Net	SwinBTS, TransUNet, DAC-Former	Segmentation (BraTS)	Dice ~0.83-0.93
Generative / diffusion	BTSegDiff, TumorGAN	Segmentation	Dice ~0.87-0.93
YOLO-based detection	RCS-YOLO, BGF-YOLO	Object detection	mAP gains of 1-4.7% over prior YOLO

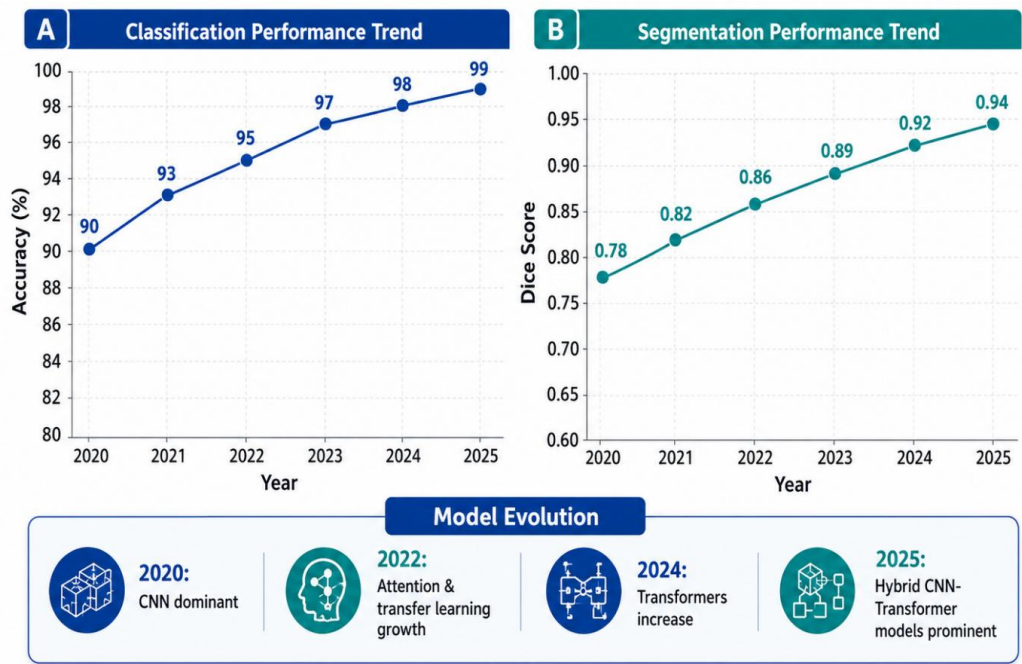


Figure 3. Trends in Brain Tumor Deep Learning Research (2020–2025)

From this comparison, two overall observations emerge. First, no single family of architectures dominates over all evaluation axes: CNN and U-Net variants are still efficient and reliable baseline models; hybrid CNN-classical ML architectures increase robustness at the cost of increased complexity; transformer-based models offer improved long-range context modelling capabilities at the expense of additional training data and computational power, and generative or semi-supervised methods overcome the challenge of annotation scarcity but suffer from the risk of synthetic artefacts [15]. Second, by looking at the empirical case study in Section 3.6, it becomes clear that moderately sized architectures can outperform much larger transfer learning backbones when applied to a fixed dataset, further reinforcing the general conclusion of the survey: the choice of the model should be based on the clinical setting of interest, instead of its size [15, 31].

8. Discussion: Clinical Limitations and Real-World Readiness

Despite consistently high benchmark performance, several limitations currently prevent deep learning brain tumor systems from translating readily to the clinic. Most studies evaluate models on one dataset, usually either BraTS, or Figshare, and do not validate on independent institutional data; models trained on BraTS have been reported to have a decrease of 10-20% in Dice score when tested on real hospital data, highlighting the challenge of domain shift between curated research datasets and the daily clinical acquisition [15]. Interpretability is also limited: only a handful of studies use explanations like Grad-CAM, layer-wise relevance propagation, or attention-map visualization to show that the model's decisions correspond to clinically relevant tumour regions, and criteria for assessing the interpretability of model results remain inconsistent [15]. While robustness to missing modalities, noise, and motion artefacts matters in routine clinical practice, it remains poorly studied, and even state-of-the-art multimodal models often perform poorly when one or more MRI sequences are missing [15]. This fragility is evident: ResNet50V2 and DenseNet121 were compared for meningioma detection, showing near-perfect accuracy on clean images (100% and 99.86%, respectively), while accuracy dropped to around 55% on images with added Gaussian noise. Grad-CAM confirmed that the networks localized the correct tumour areas on clean images [38]. These results suggest that benchmark accuracy is not a reliable indicator of real-world performance.

Regulatory and workflow limitations exacerbate these technical barriers. Few segmentation/classification systems have been evaluated in prospective clinical settings, and deployments have yielded mixed results, including success in incorporating the system into radiology workflows and challenges with model calibration and interoperability with picture archiving and communication systems [15]. Taken together, these insights indicate that most current deep learning systems for brain tumour analysis remain research tools, and this gap should be addressed through improved external validation, standardized reporting of tool interpretability, modality independence, and stronger connections to the clinical regulatory environment [15].

9. Future Research Directions

Multiple trends will significantly shape the next generation of multimodal brain tumour imaging systems. To move beyond concatenating simple modalities, researchers are investigating attention-based and hypernetwork methods that combine imaging with clinical or genomic data to enhance content fusion [1]. Efforts to address this need are emerging through self-supervised and few-shot learning, which have already shown promise for predicting neurodevelopmental diseases and Alzheimer's disease using only multimodal scans without ground-truth annotations [1]. Furthermore, foundation models trained on large collections of MR scans using contrastive learning or masked image modeling are likely to reduce the need for voxel-level labels and improve generalization across scanners and institutions [15]. In the context of data augmentation and missing-modality reconstruction, generative methods such as diffusion-based MRI synthesis and GAN-based modality completion are likely to be used more widely, with the key challenge being to ensure anatomically realistic outputs while avoiding unrealistic values through domain-aligned priors and uncertainty-aware synthesis [15]. Lastly, federated and privacy-preserving learning is expected to increasingly integrate with the self-supervised and generative pipelines, allowing for secure multi-institutional training while also meeting the demands of interpretability, auditability and real-time processing needed for eventual

clinical use, including new applications that combine imaging with wearable physiological monitoring for ongoing neurological assessment [1, 15]. Fig. 4 summarizes the relationship between unresolved clinical and technical challenges and the most promising future research directions, following the priorities identified in [1] and [15].

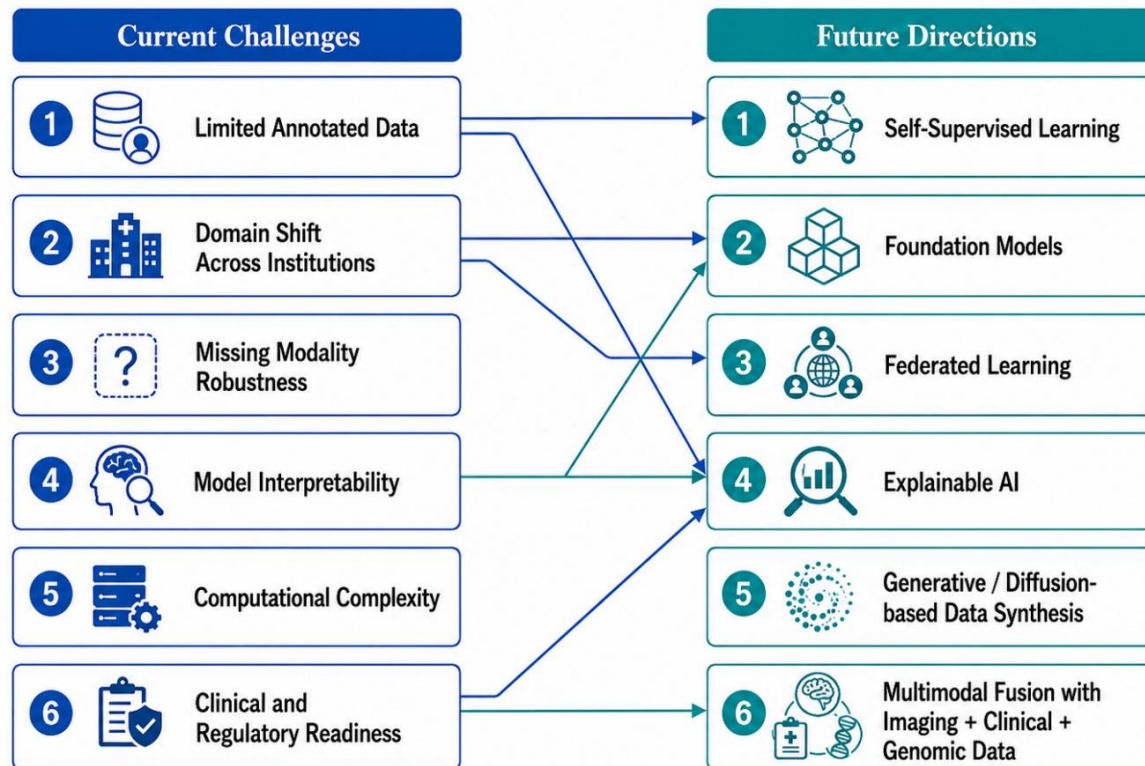


Figure 4. Challenges and Future Directions in Brain Tumor Imaging AI

10. Conclusion

Deep learning has revolutionized brain tumour imaging research, replacing traditional manual feature-engineering pipelines for classification, segmentation, forecasting, and object detection with CNN, transformer, and hybrid architectures on multimodal MRI and other imaging data. Overall, classification and segmentation accuracy (Dice scores) have significantly increased since 2020, with advanced attention mechanisms, transformer-based encoders, and hybrid CNN-Transformer architectures having a major impact on both tasks, while transfer learning and lightweight, custom-designed architectures provide strong baselines at minimal computational cost. Meanwhile, limitations in dataset diversity, cross-institution generalization, missing-modality robustness, interpretability, and regulatory validation suggest that the shift from benchmark performance to routine clinical use is far from complete. In the future, these impressive performance improvements on curated benchmarks will likely rely on the integration of self-supervised and foundation-model pretraining, generative data synthesis, federated and privacy-preserving training, and greater emphasis on external validation and interpretability for clinicians to ensure that the significant progress made on these benchmarks equates to reliable patient outcomes in actual clinical practice within the neuro-oncology space.

References

1. S. M. Balsabti et al., "Advances in deep learning for multimodal brain imaging: A comprehensive survey," *Neuroscience Informatics*, vol. 6, Art. no. 100252, 2026, doi: 10.1016/j.neuri.2025.100252.
2. M. S. Ullah et al., "BrainNet: A fusion assisted novel optimal framework of residual blocks and stacked autoencoders for multimodal brain tumor classification," *Scientific Reports*, vol. 14, Art. no. 5895, 2024, doi: 10.1038/s41598-024-56657-3.
3. H. H. N. Alrashedy, A. F. Almansour, D. M. Ibrahim, and M. A. A. Hammoudeh, "BrainGAN: Brain MRI image generation and classification framework using GAN architectures and CNN models," *Sensors*, vol. 22, no. 11, Art. no. 4297, 2022, doi: 10.3390/s22114297.
4. A. M. Al-Zoghby, E. M. K. Al-Awadly, A. Moawad, N. Yehia, and A. I. Ebada, "Dual deep CNN for tumor brain classification," *Diagnostics*, vol. 13, no. 12, Art. no. 2050, 2023, doi: 10.3390/diagnostics13122050.
5. A. Raza et al., "A hybrid deep learning-based approach for brain tumor classification," *Electronics*, vol. 11, no. 7, Art. no. 1146, 2022, doi: 10.3390/electronics11071146; K. Gencer and G. Gencer, "Hybrid deep learning approach for brain tumor classification using EfficientNetB0 and novel quantum genetic algorithm," *PeerJ Computer Science*, vol. 11, Art. no. e2556, 2025, doi: 10.7717/peerj-cs.2556.
6. K. Pratap Joshi, V. B. Gowda, P. Bidare Divakarachari, P. Siddappa Parameshwarappa, and R. K. Patra, "VSA-GCNN: Attention guided graph neural networks for brain tumor segmentation and classification," *Big Data and Cognitive Computing*, vol. 9, no. 2, Art. no. 29, 2025, doi: 10.3390/bdcc9020029.
7. M. Kang, C. M. Ting, F. F. Ting, and R. C.-W. Phan, "RCS-YOLO: A fast and high-accuracy object detector for brain tumor detection," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2023, Lecture Notes in Computer Science*, vol. 14223, pp. 600-610, 2023, doi: 10.1007/978-3-031-43901-8_57.
8. M. Kang, C. M. Ting, F. F. Ting, and R. C.-W. Phan, "BGF-YOLO: Enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2024*, pp. 35-45, 2024, doi: 10.1007/978-3-031-72111-3_4.
9. W. Zafar et al., "Enhanced TumorNet: Leveraging YOLOv8s and U-Net for superior brain tumor detection and segmentation utilizing MRI scans," *Results in Engineering*, vol. 24, Art. no. 102994, 2024, doi: 10.1016/j.rineng.2024.102994.
10. Q. Xing, Z. Li, Y. Jing, and X. Chen, "A 3D dual encoder mirror difference ResU-Net for multimodal brain tumor segmentation," *IEEE Access*, vol. 13, pp. 1621-1635, 2025, doi: 10.1109/ACCESS.2024.3522682.
11. Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A method for 3D multimodal brain tumor segmentation using Swin Transformer," *Brain Sciences*, vol. 12, no. 6, Art. no. 797, 2022, doi: 10.3390/brainsci12060797.
12. Y. Wei, X. Chen, L. Zhu, L. Zhang, C.-B. Schonlieb, S. Price, and C. Li, "Multi-modal learning for predicting the genotype of glioma," *IEEE Transactions on Medical Imaging*, vol. 42, no. 11, pp. 3167-3178, 2023, doi: 10.1109/TMI.2023.3244038.
13. H. Zhao, H. Cai, and M. Liu, "Transformer-based multi-modal MRI fusion for prediction of post-menstrual age and neonatal brain development analysis," *Medical Image Analysis*, vol. 94, Art. no. 103140, 2024, doi: 10.1016/j.media.2024.103140.
14. B. H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993-2024, 2015, doi: 10.1109/TMI.2014.2377694.
15. A. C. Yadav and M. H. Kolekar, "Deep feature-based approaches for brain tumor classification and segmentation in medical imaging," *Biomedical Signal Processing and Control*, vol. 117, Art. no. 109603, 2026, doi: 10.1016/j.bspc.2026.109603.
16. F. Isensee, P. F. Jager, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-Net for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Lecture Notes in Computer Science*, vol. 12659, pp. 118-132, 2021, doi: 10.1007/978-3-030-72087-2_11.
17. M. Dobko, D.-I. Kolinko, O. Viniavskyi, and Y. Yelisieiev, "Combining CNNs with transformer for multimodal 3D MRI brain tumor segmentation with self-supervised pretraining," *arXiv preprint arXiv:2110.07919*, 2021.

18. L. Fidon, S. Shit, I. Ezhov, J. C. Paetzold, S. Ourselin, and T. Vercauteren, "Generalized Wasserstein Dice loss, test-time augmentation, and transformers for the BraTS 2021 challenge," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Lecture Notes in Computer Science, vol. 12963, pp. 173-186, 2022, doi: 10.1007/978-3-031-09002-8_17.
19. R. A. Zeineldin, M. E. Karar, F. Mathis-Ullrich, and O. Burgert, "Ensemble CNN networks for GBM tumors segmentation using multi-parametric MRI," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Lecture Notes in Computer Science, vol. 12962, pp. 473-483, 2022, doi: 10.1007/978-3-031-08999-2_41.
20. S. Nalawade et al., "Federated learning for brain tumor segmentation using MRI and transformers," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Lecture Notes in Computer Science, vol. 12963, pp. 444-454, 2022, doi: 10.1007/978-3-031-09002-8_39.
21. M. Rasool, N. A. Ismail, W. Boulila, A. Ammar, H. Samma, W. M. S. Yafooz, and A.-H. M. Emara, "A hybrid deep learning model for brain tumour classification," *Entropy*, vol. 24, no. 6, Art. no. 799, 2022, doi: 10.3390/e24060799.
22. A. Shafi, M. B. Rahman, T. Anwar, R. S. Halder, and H. E. Kays, "Classification of brain tumors and auto-immune disease using ensemble learning," *Informatics in Medicine Unlocked*, vol. 24, Art. no. 100608, 2021, doi: 10.1016/j.imu.2021.100608.
23. S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine*, vol. 111, Art. no. 103345, 2019, doi: 10.1016/j.compbimed.2019.103345.
24. Z. N. K. Swati, Q. Zhao, M. Kabir, F. Ali, Z. Ali, and J. Lu, "Brain tumor classification for MR images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34-46, 2019, doi: 10.1016/j.compmedimag.2019.05.001.
25. J. Zhang, X. Tan, W. Chen, G. Du, Q. Fu, H. Zhang, and H. Jiang, "EFF_D_SVM: A robust multi-type brain tumor classification system," *Frontiers in Neuroscience*, vol. 17, Art. no. 1269100, 2023, doi: 10.3389/fnins.2023.1269100.
26. I. Pacal, O. Celik, B. Bayram, and A. Cunha, "Enhancing EfficientNetV2 with global and efficient channel attention mechanisms for accurate MRI-based brain tumor classification," *Cluster Computing*, vol. 27, pp. 11187-11212, 2024, doi: 10.1007/s10586-024-04532-1.
27. N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: Application to brain tumor classification," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 817-824, 2022, doi: 10.1007/s11760-021-02022-0.
28. P. Afshar, A. Mohammadi, and K. N. Plataniotis, "BayesCap: A Bayesian approach to brain tumor classification using capsule networks," *IEEE Signal Processing Letters*, vol. 27, pp. 2024-2028, 2020, doi: 10.1109/LSP.2020.3034858.
29. S. Srivastava, P. Jain, S. K. Pandey, G. Dubey, and N. N. Das, "Automated brain tumor classification and grading using multi-scale graph neural network with spatio-temporal transformer attention through MRI scans," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 18, no. 1, pp. 122-150, 2026, doi: 10.1007/s12539-025-00718-2.
30. E. Zarenia, A. A. Far, and K. Rezaee, "Automated multi-class MRI brain tumor classification and segmentation using deformable attention and saliency mapping," *Scientific Reports*, vol. 15, Art. no. 8114, 2025, doi: 10.1038/s41598-025-92776-1.
31. V. Anand, A. Khajuria, R. K. Pachauri, and V. Gupta, "Multi-class classification of brain tumors using optimized CNN and transfer learning techniques," *Scientific Reports*, vol. 16, Art. no. 4709, 2026, doi: 10.1038/s41598-025-34806-6.
32. A. Annadurai et al., "Discrete wavelet transform based multiscale deep CNN cascaded LSTM model for classification of brain tumor," *In Silico Research in Biomedicine*, vol. 2, Art. no. 100239, 2026, doi: 10.1016/j.ins.2026.100239.
33. O. N. Belaid and M. Loudini, "Classification of brain tumor by combination of pre-trained VGG16 CNN," *Journal of Information Technology Management*, vol. 12, no. 2, pp. 13-25, 2020.
34. Z. Huang, X. Du, L. Chen, Y. Li, M. Liu, Y. Chou, and L. Jin, "Convolutional neural network based on complex networks for brain tumor image classification with a modified activation function," *IEEE Access*, vol. 8, pp. 89281-89290, 2020, doi: 10.1109/ACCESS.2020.2993618.
35. Z. A. Sejuti and M. S. Islam, "An efficient method to classify brain tumor using CNN and SVM," in *Proc. 2nd Int. Conf. Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021, pp. 644-648, doi: 10.1109/ICREST51555.2021.9331060.
36. S. Rajendran et al., "Automated segmentation of brain tumor MRI images using deep learning," *IEEE Access*, vol. 11, pp. 64758-64768, 2023, doi: 10.1109/ACCESS.2023.3288017.

37. S. L. Jemina, T. Tamilvizhi, and W. B. Sherine, "Brain tumor detection using extended optimization based multi-layer perceptron," in Proc. 2025 Int. Conf. Computing Technologies and Data Communication (ICCTDC), 2025, pp. 1-6, doi: 10.1109/ICCTDC64446.2025.11158990.
38. G. Femila Irene and R. Surendran, "Deep learning on brain MRI: ResNet50V2 vs. DenseNet121 meningioma detection," in Proc. 4th Int. Conf. Innovative Mechanisms for Industry Applications (ICIMIA), 2025, pp. 1412-1417, doi: 10.1109/ICIMIA67127.2025.11200655.
39. R. Jayalakshmi, T. Tamilvizhi, and P. Ramya, "Comprehensive evaluation of federated learning based models for disease detection in healthcare," in Proc. 2024 Int. Conf. Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2024, pp. 643-650, doi: 10.1109/3ICT64318.2024.10824463.
40. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
41. Balsabti, S. M., Al-Gburi, R. M., Mustafa, A., Ahmed, S. K., Issa, A. M., Al-Naimi, T. M., ... & Elhenidy, A. M. (2026). Advances in deep learning for multimodal brain imaging: A comprehensive survey. *Neuroscience Informatics*, 6(1), 100252.
42. S. Ünlüleblebici, R. Nur Kaçmaz and C. Öztürk, "A Systematic Review of Machine Learning and Deep Learning Techniques for Brain Tumor Classification, and Grading," in *IEEE Access*, vol. 14, pp. 18075-18098, 2026, doi: 10.1109/ACCESS.2026.3659641.