Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:50
https://doi.org/10.28919/cmbn/9888
ISSN: 2052-2541

AN EXPLAINABLE KNOWLEDGE-DRIVEN HYBRID CNN–TRANSFORMER AND RADIOMICS FRAMEWORK FOR MULTI-MODAL MRI-BASED BRAIN TUMOR ANALYSIS

MONICA LUTHRA[1,2,*], SELLAPPAN PALANIAPPAN[3], DANIEL AROCKIAM[4]

[1]Post doc fellow, Faculty of Computing and Digital Technology, HELP University, Kuala Lumpur 50490, Malaysia

[2]Apex Institute of Technology (CSE), Chandigarh University, Mohali, Punjab 140413, India

[3]Faculty of Computing and Digital Technology, HELP University, Kuala Lumpur 50490, Malaysia

[4]School of Engineering and IT, Manipal Academy of Higher Education, Dubai Campus, UAE

Abstract: The increasing complexity of medical data in modern healthcare environments necessitates intelligent frameworks that not only deliver accurate predictions but also support effective knowledge generation and decision-making. In the context of brain tumor analysis using multi-modal magnetic resonance imaging (MRI), challenges such as tumor heterogeneity, variability in imaging protocols, and limited interpretability of deep learning models hinder their integration into clinical knowledge workflows. This study proposes an explainable, knowledge-driven hybrid CNN–Transformer–Radiomics (HCTR) framework designed to facilitate both predictive performance and clinical knowledge extraction. The framework integrates convolutional neural networks for localized feature learning, transformer-based architectures for global contextual understanding, and radiomic descriptors for structured, domain-relevant feature representation. A cross-attention-based fusion mechanism is employed to combine these heterogeneous knowledge sources into a unified representation. Beyond detection and analysis, the proposed system emphasizes interpretability through Grad-CAM visualizations and feature attribution methods, enabling the transformation of model outputs into clinically meaningful insights. This supports enhanced transparency, trust, and

---

[*] Corresponding author

E-mail address: monica.l@helplive.edu.my

Received March 23, 2026

knowledge dissemination within clinical decision-making processes. The proposed framework contributes to the development of intelligent decision support systems by bridging data-driven modeling with knowledge-centric interpretation. It provides a scalable approach for integrating explainable AI into healthcare knowledge management environments, facilitating improved diagnostic reasoning and informed clinical decisions.

Keywords: brain tumor detection; hybrid deep learning; CNN-transformer architecture; radiomics; explainable AI; multi-modal MRI; medical image analysis.

2020 AMS Subject Classification: 68T07, 68T10, 92C55.

1. INTRODUCTION

Brain tumors are a major global health challenge, particularly in terms of diagnosis, treatment, and prognosis. The Global Cancer Observatory (GLOBOCAN) reports that more than 300,000 new cases are recorded each year [1], and the mortality rate remains high despite advances in neuro-oncology. The characteristics of tumors of the brain and their growth can significantly impact the functioning of the nerve system because they are directly related to the central nervous system. This the reason that these usually cause defects in the neurological system, cognitive decline, and worsened quality of life. The burden is especially pronounced in low- and middle-income nations in which access to high-tech imaging, and specialized neuro-oncological care is restricted. In addition, heterogeneity of brain tumors in terms of size, shape, location and sub-type of histology serves as a complicating factor in the detection and classification of brain tumors making early and accurate diagnosis highly challenging.

The existing brain tumor classification methods are mainly based on radiological investigations, Magnetic Resonance Investigation (MRI), Computed Tomography (CT) and in certain instances, Positron Emission Research (PET). Under such imaging techniques, the radiologists or neurologist perform the interpretation and this is a subjective process and therefore subject to intra and inter observer variability. In addition, the histopathological analysis which is regarded as the gold standard in final diagnosis is invasive and slow. Traditional workflows are often ineffective when analyzing large numbers of tumor images that show subtle morphological changes. Recent deep learning-based approaches in cancer image analysis have demonstrated improved diagnostic performance and robustness [2]. This limitation is particularly relevant for younger patients or individuals with recurrent tumors. Such constraints can cause delays, misclassification and poor treatment planning in high-risk clinical situations. This has had an important effect on the direction of research in brain tumor detection because of the availability of benchmark datasets and the increasing adoption of transfer learning in medical imaging applications [3]. The frequency of the usage of most of the datasets that are often used in recent studies is illustrated in Figure 1. This distribution shows the dominance of benchmark datasets in DL-based brain tumor research.

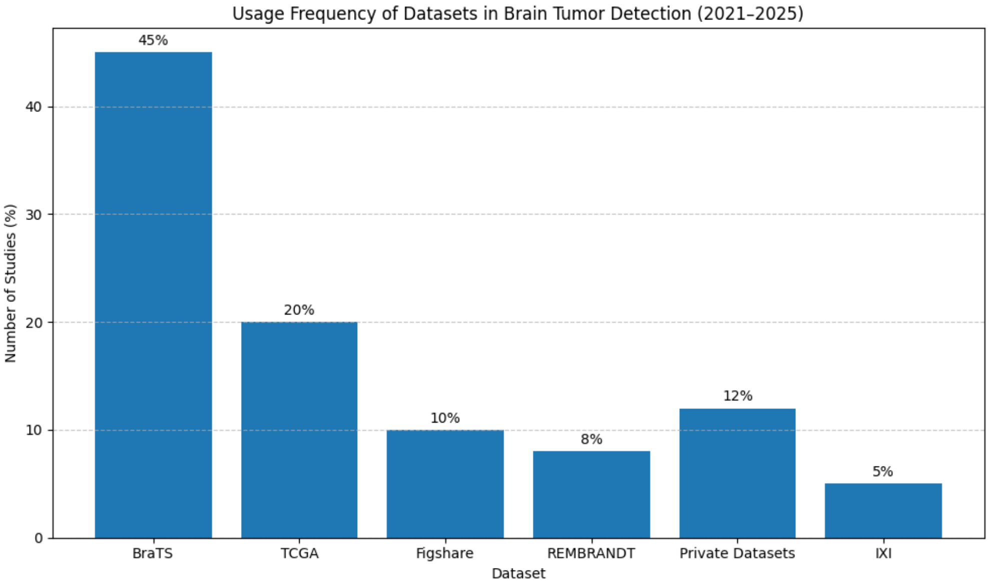

Figure 1: Usage Frequency of Datasets in Brain Tumor Detection (2021–2025)

The limitations of the manual interpretation and application of traditional imaging strategies have led to an increase in the demand of automated, quantitative, and effective diagnostic methods. In that respect, the use of artificial intelligence (AI), and, in particular, deep learning (DL) has become an effective tool to use in the sphere of medical image analysis [4]. Deep learning is a branch of machine learning that uses multi-layer neural networks. It has shown strong performance in many computer vision tasks, including object detection, classification, and segmentation. Convolutional Neural Networks (CNNs) are an example of DL models that when trained with medical imaging can learn without using handcrafted features to learn complex hierarchies of imaging data. Recent works show that the rate of adoption of deep learning architectures can vary in brain tumor detection. Figure 2 shows the relative frequency of the most commonly used models.

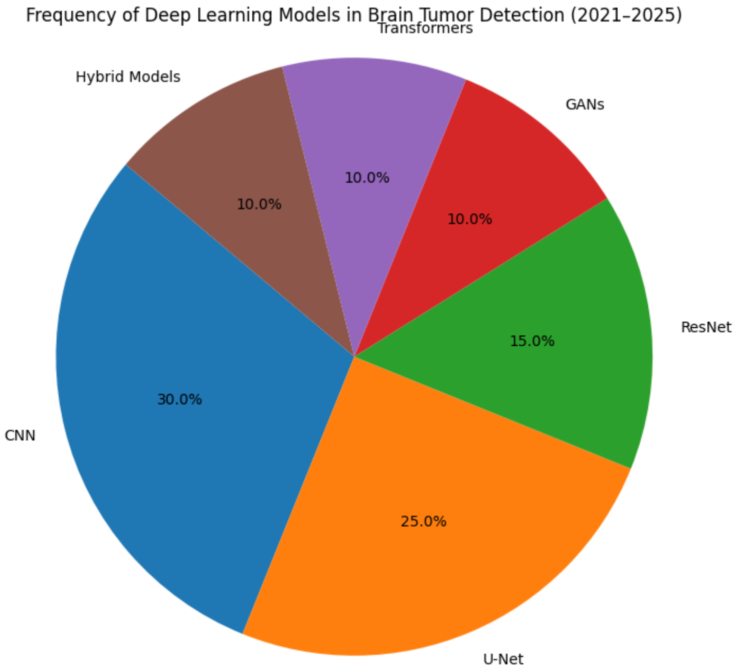

Figure 2: Frequency of Deep Learning Models in Brain Tumor Detection (2021–2025)

The distribution show that CNN-based models are still dominant, whereas hybrid and transformer-based models are gradually gaining attention.

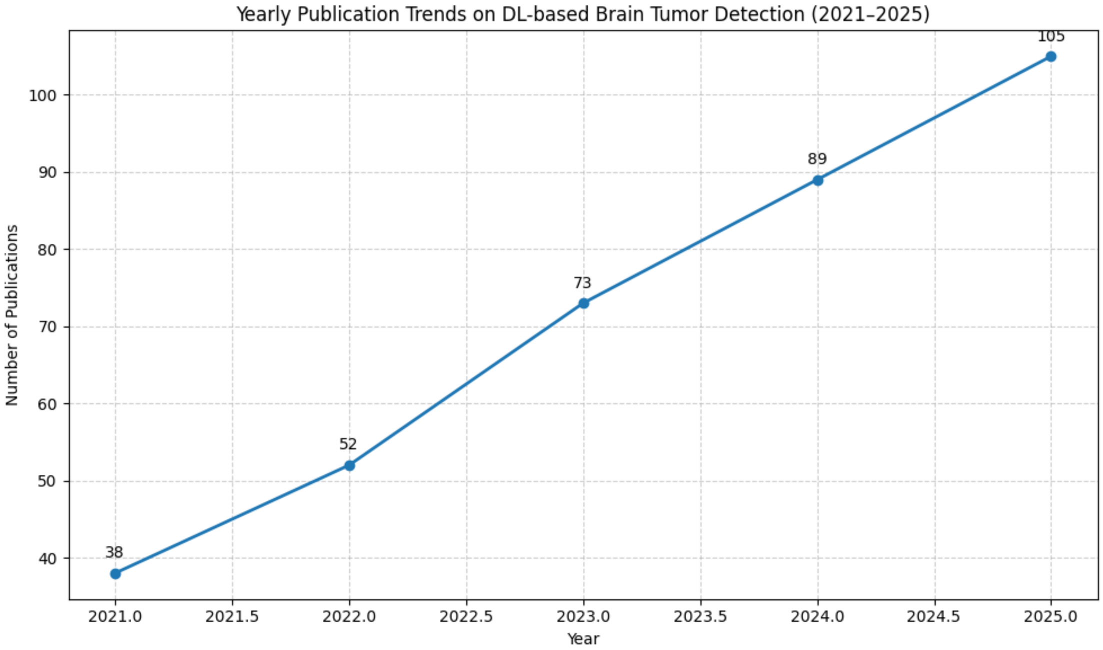

Figure 3: Yearly Publication Trends on DL-based Brain Tumor Detection (2021–2025)

The gradual rise in the number of publications indicates the increase in the research interest and the swift development of DL algorithms to detect brain tumors.

Even with the promising research outcomes, there are a number of challenges remain. Patient-centric diagnostics is another opportunity that is connected with the application of deep learning in the smart healthcare context. AI has the potential to enable tailored medicine by computing longitudinal data of image and clinical data to forecast disease progression and therapy response. Moreover, interpretability of diagnostic models via explainable artificial intelligence (XAI) algorithms including Grad-CAM, LIME, and SHAP are also being applied in clarifying the explanation of the estimated values to the clinicians and thus boosting clinician trust and model interpretability. These advances are a paradigm shift on intelligent and adaptive diagnostic systems that can enhance patient outcomes and decrease diagnostic disparities. Even a narrower examination of the recent years (2021-2025) more clearly demonstrates the accelerated growth of DL-based studies, as shown in Figure 3.

Out of these motivations, the current paper proposes a Hybrid CNN-Transformer-Radiomics (HCTR) system to detect brain tumors. The proposed solution combines convolutional feature learning to obtain the local spatial representation, transformer-based attention models to obtain the global contextual model, and hand-crafted radiomic characterizations to obtain the quantitative description of tumor properties. The framework aims to optimize the use of the complementary feature representations in a single architectural design to improve the diagnostic performance and also aid interpretability in brain tumor detection.

2. OVERVIEW OF BRAIN TUMOR IMAGING MODALITIES

The early and precise diagnosis of brain tumors is generally very reliant on the successful application of the medical imaging devices that enable visualization the structure, function, and pathology of brain tissues. The number of imaging modalities that are utilized in clinical practice is immense; Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET) and histopathological imaging are some of the most common imaging modalities used in neuro-oncology. Modes have advantages and limitations which include resolution, contrast, biological understanding and availability. The processes of brain tumor detection and classification based on deep learning methods highly dependent on the data produced on the basis of these imaging modalities. Therefore, the insights about the nature of these data are of great importance to the creation of successful AI models. Histopathology and PET imaging are experiencing a continual rise in research focus [5]. The distribution of the most frequently used datasets in recent brain tumor studies is shown in Figure 4.

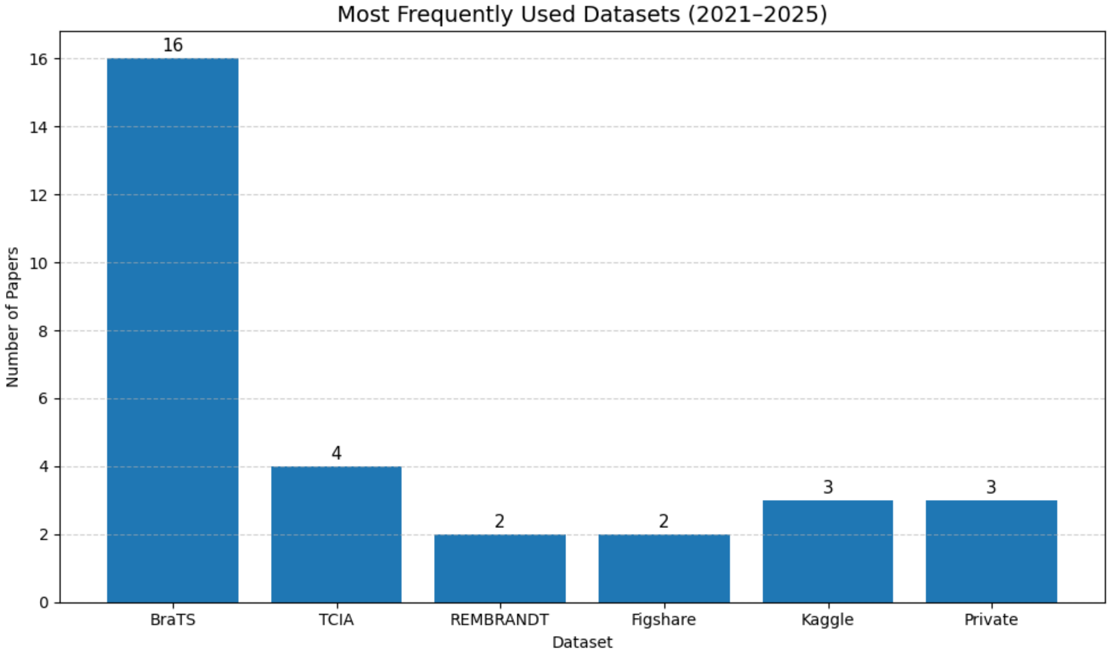

Figure 4: Most Frequently Used Datasets (2021–2025)

2.1. Magnetic Resonance Imaging (MRI)

MRI is the most popular imaging modality in the diagnosis of brain tumor because it has good soft-tissue contrast, non-invasive, and the ability to provide multi-parametric information. This is in contrast to CT which uses ionizing radiations and hence the use of CT is best on repeat scanning conditions especially in the cases, when dealing with children and follow-up in specific cases. MRI is based on the powerful magnetic field which acts on the body and orders the hydrogen protons and determines the energy spawned as the protons restore their balance. It is a signal that is dependent on the tissue, and hence giving a fine resolution visualization of the view of different brain structures.

The various sequences of MRI give complementary information concerning the tumor. Important MRI image types include T1-weighted images, which provide detailed anatomical information. T2-weighted and FLAIR (Fluid-Attenuated Inversion Recovery) sequences highlight edema and water content associated with tumors [6]. The contrast-enhanced T1-weighted is particularly effective, as much as it is even able to determine blood-brain barrier ruptures associated with high-grade tumors. There are modern imaging modalities like Diffusion Weighted Imaging, Perfusion-Weighted Imaging and Magnetic Resonance Spectroscopy which provide data on the functioning and metabolism that are useful during tumor grading and treatment planning.

Although it may sound favorable, MRI has a couple of challenges. It is motion sensitive and tends to suffer from inhomogeneity because of magnetic field fluctuations. In addition, the long acquisition time may be unpleasant to patients. The difference in scanner type, protocol and sequence parameter between institutions also contributes to the difficulty in developing generalized deep learning models. Annotations of MRI images also require expert radiologists, thereby restricting the supply of high-quality large-sized labeled databases.

2.2. Computed Tomography (CT)

CT scans use X-ray beams to produce cross-sectional images of the brain and are widely available in emergency settings. CT is useful for identifying hemorrhage, calcification, and bone involvement [7]. It is sometimes employed when MRI is contraindicated due to metallic implants or other medical conditions.

However, CT provides lower soft-tissue contrast compared to MRI and exposes patients to ionizing radiation. From a deep learning perspective, CT datasets are less frequently used for brain tumor segmentation due to reduced contrast between tumor and surrounding tissues. Nevertheless, CT may contribute to multimodal learning systems when combined with other imaging modalities.

2.3. Positron Emission Tomography (PET)

PET imaging is also providing metabolic and functional information related to the brain tumors, which cannot be provided alone by structural scanning. It involves the involvement of radioactive tracers like fluorodeoxyglucose (FDG), which gathers in regions of high metabolism. Malignant tumors typically exhibit increased glucose metabolism and thus they can be perceived on PET examinations as hypermetabolic regions [8].

PET is frequently applied in clinical practice in combination with CT or MRI in the cases of a better localization and characterization of tumors, mainly in doubtful cases. An example is the option of PET-MRI fusion to distinguish between tumor recurrence and radiation necrosis which borders are hard to tell by use of conventional imaging. More recently novel tracers, with potential improved specificity (e.g., tumor markers: amino acid transporters (e.g., FET, MET) or proliferation marker (e.g., FLT)) are also under investigation.

Nevertheless, PET imaging is costly, not available, and it uses radiation. PET has a comparatively low spatial resolution as well, and interpretation can generally be demanding without being combined with anatomical imaging. Being in a sense of deep learning, the issue is that there are not many large-scale datasets of PET done and therefore the heterogeneity of the tracer uptake patterns cannot be ignored. However, PET imaging has potential towards creating multi-modal AI systems that are able to allow integration of functional and structural data to enhance clinical decision-making.

2.4. Histopathological Imaging

Histopathological evaluation constitutes the gold standard by which tumors are definitively identified, graded, and practically characterized. This is a modality that deals with the analysis of tissue samples using a microscope either in biopsy or surgical excision. Digital pathology, the process of scanning glass slides into high-resolution whole-slide images (WSIs), therefore makes it possible to apply AI and deep learning methods to detect tumors, segment them, and classify them.

Rich morphological and cellular data are encoded in histopathological images, and could be used to differentiate between sub-types of tumors like glioblastoma, astrocytoma and oligodendroglioma. Incorporation of deep learning models on WSIs can produce proficient skills capable of matching the tumor classification as well as tumor grading of experts. Due to the high spatial resolution and intricate texture of histology images, CNNs would particularly be suitable in the processing of such images.

Nonetheless, image analysis of histological pictures is not devoid of issues. Noise and artifacts can be significantly depending on the variability of the staining procedures, the type of scanner and tissue preparation. Moreover, WSIs are extremely large (often several thousand pixels in dimension) that processing them needs huge amounts of computational resources. It is also a labor-intensive process that necessitates the assistance of expert pathologists and the inter-observer variability is an issue.

2.5. Imaging-Related Challenges in Brain Tumor Detection

Although each imaging modality has its own advantages, the modalities come with several limitations that interfere directly with the performance and generalizability of deep learning models. Image noise represents one of the key problems as it might occur due to limited hardware, patient movement, or parameters of acquisition. The noise level lowers the ratio between signal and noise and may cover even very small structural details that will be needed to get the precise tumors boundaries.

The other potentially devastating issue is low contrast between tumor tissue and healthy tissue in non-enhancing lesion or low-grade gliomas. This complicates the process of both human and AI model-based delineation of tumors. Another widespread problem with brain tumor datasets is class imbalance, in which normal tissues are orders of magnitude more abundant than abnormal ones, causing biased model training.

In addition, tumor appearance varies substantially depending on inter-patient as well as intra-patient variability caused by tumor type, grade, location, and size. This diversity makes it hard to come up with effective deep learning models that can be used to make generalizations across populations. Multi-center data are further affected by domain shift in the form of incongruence in scanners, imaging protocols, and patient demographic.

The problems could be solved by using preprocessing methods like intensity normalization, noise removal, and data augmentation. Besides, standardized datasets such as BraTS have had a major impact in the field by offering multimodal MRI scans of cancer patients that have been preprocessed with expert annotations.

A comparative accuracy analysis across major imaging modalities from recent studies is presented in Figure 5.

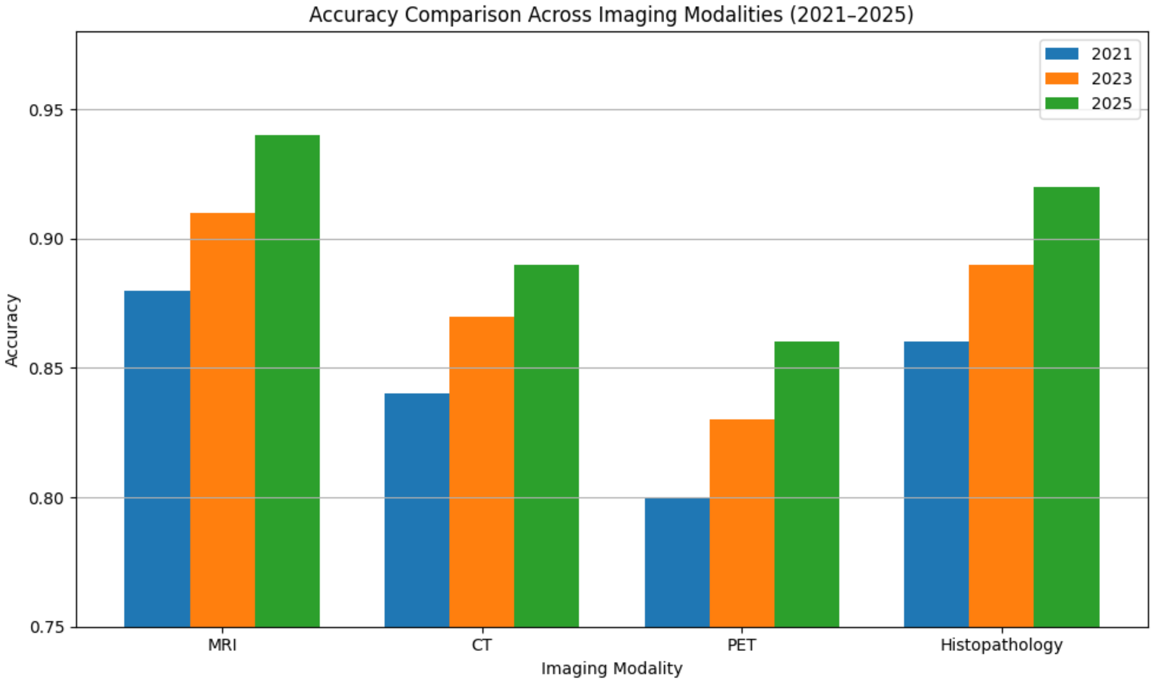

Figure 5: Accuracy Comparison Across Imaging Modalities (2021–2025)

3. RELATED WORK

Recent years have indicated rapid growth in deep learning (DL) approaches for brain tumor detection using MRI and multimodal imaging. Most studies focus on classification, segmentation, or hybrid diagnostic frameworks and report competitive performance. This section categorizes the existing work into major research directions. The proportion of deep learning model types adopted in recent studies is illustrated in Figure 6.

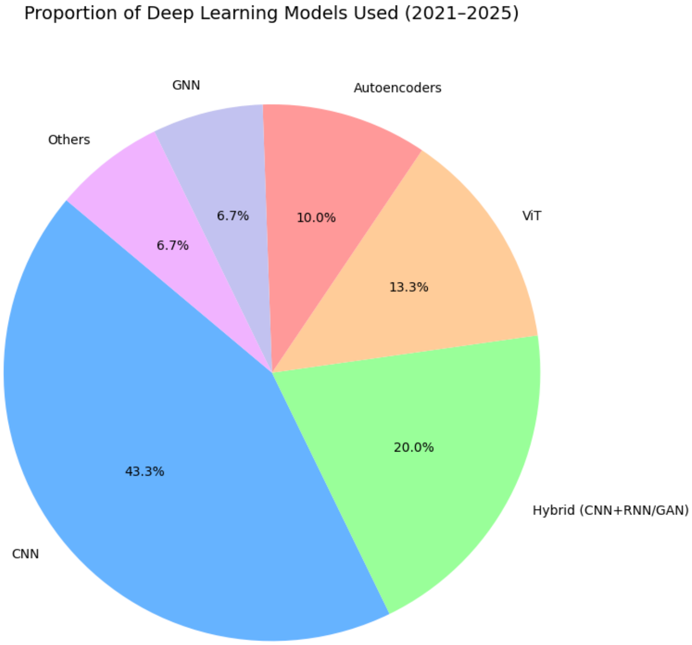

Figure 6: Proportion of Deep Learning Models Used (2021–2025)

3.1. CNN-Based Architectures for Brain Tumor Detection

Convolutional Neural Networks (CNNs) are widely adapted for tumor detection using MRI. Abdusalomov et al. [9] optimized YOLOv7 with CBAM and BiFPN modules. Their approach achieved efficient accuracy on 10,288 MRI images, although sensitivity for small tumors remained limited. Shah et al. [10] optimized EfficientNet-B0 and achieved an accuracy of 98.87%, which is higher than the results with VGG16 and ResNet50. Similarly, Ibrahim et al. [11] proposed a modified Xception model with efficient accuracy and strong sensitivity and specificity values.

Mahmud et al. [13] compared custom CNN with ResNet-50 and VGG16, and they had the highest accuracy (93.3%) and better AUC (98.43%). Hussien et al. [14] further optimized Xception architecture and reported efficient results on the Sartaj dataset. Preetha et al. [15] employed EfficientNet-B4 with Bayesian hyperparameter tuning and K-fold validation, achieving 99.33% accuracy. The scalability of EfficientNetB0 was shown by Raj et al. [21] with a validation accuracy of 97.5% and presented the suitability of the model in low-resource environments. Kumar and Mohanty [25] compared the DenseNet-121, ResNet-101, and MobileNetV2 where DenseNet-121 performed at 99% classification accuracy. Musallam et al. [27] proposed a lightweight DCNN achieving 98.22% accuracy on 3394 MRI images.

Though CNN-based methods show effective performance, but most of the studies use small-scale datasets or 2D slices or are not externally clinically validated.

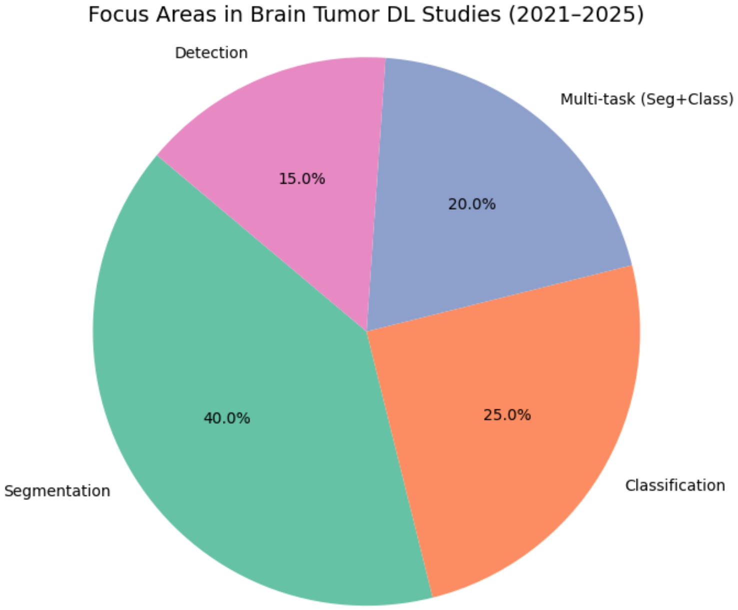

Figure 7: Focus Areas in Brain Tumor DL Studies (2021–2025)

Figure 7 summarizes the primary focus areas of brain tumor DL research, including classification, segmentation, and hybrid frameworks.

3.2. Lightweight and Transfer Learning-Based Models

To address computational constraints in smart healthcare environments, there are several works which focus on lightweight architectures and transfer learning.

Kumar et al. [22] created a model based on MobileNetV2 for edge deployment. The authors stated to obtain greater accuracy with better preprocessing. Nasar et al. [23] compared MobileNet and EfficientNet with variations on small dataset (300 MRI images). The authors highlighted the hyperparameter optimization in order to detect it earlier. Khan et al. [28] compared DenseNet, EfficientNet, and MobileNet, highlighting MobileNet's computational efficiency despite slightly lower accuracy. These studies demonstrate feasibility in low-resource environments but often suffer from small sample sizes and limited generalization.

3.3. Hybrid Deep Learning Frameworks

The studies related to hybrid models combining deep learning with classical ML or optimization techniques have also been explored. Saeedi et al. [12] proposed a CNN-based hybrid model which outperforms classical ML methods (KNN, MLP). The authored have achieved 96.47% training accuracy. Princiba et al. [16] integrated EfficientNetB3V2 with a Maximum Entropy classifier and achieved efficient validation accuracy.

Ahmad and Choudhury [17] evaluated 35 hybrid CNN-ML combinations. The author reported the significant accuracy results using transfer learning with ensemble classifiers. Dhakshnamurthy et al. [24] combined VGG16 and ResNet-50. These authors have also achieved more than 99% accuracy in multiclass classification.Li et al. [26] integrated Firefly Optimization with parallel CNNs, reporting 98.6% accuracy on BraTS 2018. Boopathy and Kavitha [30] proposed VGG16-ResNet hybrid fusion and achieved more than 99% performance results.

Hybrid methods often report higher accuracy, but there is a concern on the rise in architectural complexity and the lack of real-world testing.

3.4. Multimodal, Security-Aware, and Smart Healthcare-Oriented Approaches

Recent research also emphasizes integration with smart healthcare ecosystems. Banik et al. [19] introduced BrainTumorNet incorporating blockchain for secure data management. The authors achieved the accuracy of upto 98.66% across datasets. Jajoo and Musande [20] proposed a multimodal MRI-CT framework using optimized ResNet50 and Xception architectures. Pawar and Shrivastava [29] presented a survey highlighting segmentation limitations, over-reliance on BraTS, and lack of real-world validation. It recommended hybrid CNN-based architectures and federated learning directions.

These studies highlight the importance of privacy, multimodal learning, and deployment feasibility but still lack large-scale clinical integration.

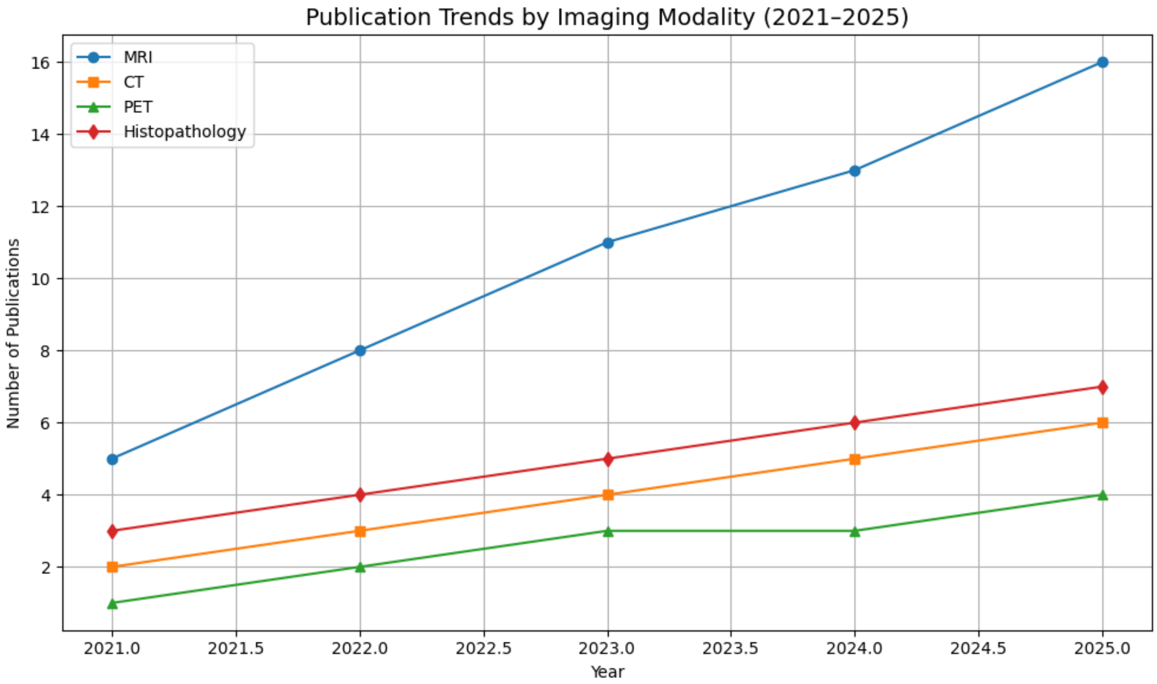

Figure 8: Publication Trends by Imaging Modality (2021-2025)

Publication trends categorized by imaging modality are presented in Figure 8.

3.5. Comparative Analysis

Table 1 summarizes representative works from 2021-2025, which includes model type, dataset, performance, strengths, and limitations.

The majority of models report higher accuracy using CNN or transfer learning frameworks. However, common limitations include: small or custom datasets, lack of external validation, limited interpretability, high computational cost, and domain shift issues.

Table 1: Comparative Analysis: Brain Tumor Detection (2021–2025)

Ref	**Author (Year)**	**Model / Architecture**	**Key Techniques**	**Dataset**	**Strengths**	**Limitations**
[9]	Abdusalomov et al. (2023)	YOLOv7 + CBAM + BiFPN	Multi-scale fusion, attention, augmentation	10,288 MRI images	Real-time detection, strong feature fusion	Low sensitivity for small tumors
[10]	Shah et al. (2022)	EfficientNet-B0	Fine-tuning, augmentation	MRI dataset	Lightweight, real-time feasible	Sensitive to data quality
[11]	Ibrahim et al. (2024)	Modified Xception	Depth-wise separable conv	Masoud Nickparvar dataset	High sensitivity & F1-score	Needs diverse validation
[12]	Saeedi et al. (2023)	2D CNN vs Autoencoder	DL + Classical ML comparison	3,264 MRI scans	Statistically validated improvement	Limited complexity modeling
[13]	Mahmud et al. (2023)	Custom CNN	Multi-model comparison	3,264 MRI images	Comparative framework	Slight performance gap vs SOTA
[14]	Hussien et al. (2025)	Optimized Xception	Transfer learning	Sartaj dataset	Robust, suitable for low-resource	Needs cross-dataset testing
[15]	Preetha et al. (2024)	EfficientNet-B4	Bayesian tuning, K-fold CV	Kaggle 2020 dataset	High generalization, scalable	Dataset-specific validation
[16]	Princiba et al. (2025)	EfficientNetB3 V2 + MaxEnt	CNN + probabilistic fusion	MRI dataset	Handles overlapping classes	No real-world validation

[17]	Ahmad & Choudhury (2022)	CNN + ML ensemble	Transfer learning + ensemble	2D MRI	Hybrid benchmarking	Only 2D data
[18]	Asif et al. (2022)	Xception + Adam	Transfer learning	Two MRI datasets	Optimizer comparison	No interpretability
[19]	Banik et al. (2023)	BrainTumorNet + Blockchain	CLAHE + secure storage	3 datasets	Privacy-preserving design	Domain shift sensitivity
[20]	Jajoo & Musande (2025)	ResNet50 + Xception	MRI + CT fusion	Multi-modal dataset	Multi-modal integration	Limited deployment validation
[21]	Raj et al. (2025)	EfficientNet-B0	Compound scaling	MRI dataset	Scalable, efficient	Moderate accuracy
[22]	Kumar et al. (2024)	MobileNetV2	Lightweight design	MRI dataset	Edge deployment friendly	Dataset limitation
[23]	Nasar et al. (2023)	EfficientNetV2	Hyperparameter tuning	300 MRI images	Effective with small data	Very small dataset
[24]	Dhakshnamurthy et al. (2024)	VGG16 + ResNet50	Hybrid feature extraction	MRI dataset	Very high performance	No explainability
[25]	Kumar & Mohanty (2024)	DenseNet-121	Feature reuse	MRI dataset	Strong gradient flow	Dataset details missing
[26]	Li et al. (2024)	CNN + Firefly Optimization	ROI segmentation + optimization	BraTS 2018	Bio-inspired hybrid approach	No clinical validation
[27]	Musallam et al. (2022)	Lightweight DCNN	Batch normalization	3,394 MRI images	Computational efficiency	Limited benchmarking
[28]	Khan et al. (2024)	DenseNet, EfficientNet, MobileNet	Multi-architecture comparison	MRI dataset	Resource-aware evaluation	No hyperparameter tuning
[29]	Pawar & Shrivastava (2024)	Survey + Proposed CNN	Gap analysis	Review-based	Strategic research direction	No experimental validation
[30]	Boopathy & Kavitha (2025)	VGG16 + ResNet50	Transfer learning hybrid	MRI dataset	High generalization	Limited deployment details

4. IDENTIFIED GAPS AND RESEARCH MOTIVATION

While recent studies demonstrate promising performance metrics, practical deployment in clinical settings remains limited. Several practical and structural limitations continue to restrict the real-world adoption of deep learning-based brain tumor detection systems.

Nonetheless, recent breakthroughs in deep learning for brain tumor detection, several issues still restrict practical deployment in smart healthcare environments. The most important ones are technical, clinical, ethical and infrastructural. The single largest shortcoming is poor generalization. Curated datasets such as BraTS or TCIA may have limited utility when models are deployed on real-world clinical data because of domain shifts caused by variability in scanners or imaging protocols, or demographics. Indicatively, the models developed with 3T MRI scans of Western populations can fail to perform well when applied to 1.5T scans or non-Western populations (unless they are suitably adapted). Another obstacle is the lack of high-quality annotated data. Time consuming, expensive and usually inconsistent, annotation of tumors is performed manually by experts. Definitions of tumor subregions are also variable, which decreases the reliability of training. Despite leading to reduced reliance on labels, semisupervised and unsupervised approaches still lag compared to the supervised ones, particularly when it comes to the fine-grained segmentation. Active learning is beneficial but needs to be incorporated in the clinical setting and professional opinion. There is also data imbalance and it skews model performance towards the more frequent tumor types. Another major issue is model robustness. Upon introducing the slightest distortions to an MRI, or even adversarial inputs, the DL models prove ineffective, may not affect human interpretation. These weaknesses establish safety concerns of real time clinical environments. There is a work in progress to come up with stronger and uncertainty-sensitive models of detecting such anomalies. The cost and infrastructure requirements in computations are limiting its real-world adoption. DL models are computationally intensive and they are not necessarily common in low resource places. It is somehow diverse when it comes to lightweight analogs such as through pruning or quantization but they are likely to compromise the accuracy. Deployment is further complicated by ethical issues like being responsible in situations of misdiagnosis. It must also possess adequate models of responsibility so that it can be in a position to trust AI systems. Lastly, it is not integrated with clinical workflows. The large number of high-performing research models fail to make it to practice because of the impossibility to be connected with the hospital systems or because of the research clinicians not being trained. In brief, the main challenges of the use of the DL in neuro-oncology are generalization, data quality, strength, ethical accountability and system integration that should be systematically tackled to enable the stable use of the DL. Only after addressing these challenges can DL systems be reliably integrated into routine clinical practice [32].

5. PROPOSED HCTR FRAMEWORK

This study proposes a Hybrid CNN-Transformer-Radiomics (HCTR) network, which combines deep learning and handcrafted extracted feature into a unified diagnostic framework. The design focuses on three main goals: accuracy, interpretability, and robustness, which are essential for clinical applicability of smart healthcare settings. The architectural block diagram is presented in Figure 9.

5.1. System Overview

The framework is conceptually designed to be trained and validated on multi-modal MRI data, such as BraTS 2019, BraTS 2021, and TCGA-GBM.

In a practical implementation, each MRI volume would undergo a preprocessing pipeline, which includes skull stripping, N4 bias field correction, histogram matching, and z-score normalization. This ensures intensity consistency across scanners and institutions. The z-score normalization can be calculated as:

$$X_{norm} = \frac{X-\mu}{\sigma} \tag{1}$$

where μ and σ represent the mean and standard deviation of voxel intensities within the MRI volume.

Tumor-aware patch extraction improves learning focus on regions with pathological content while preserving relevant contextual information.

Data augmentation strategies such as rotation, scaling, flipping, and elastic deformation are applied to improve geometric diversity and reduce class imbalance. Where necessary, generative approaches can be incorporated to enhance feature diversity and domain generalization. The Augmentation transformation can be described as:

$$X' = T(X) \tag{2}$$

where $T(\cdot)$ represents geometric transformation functions applied during augmentation.

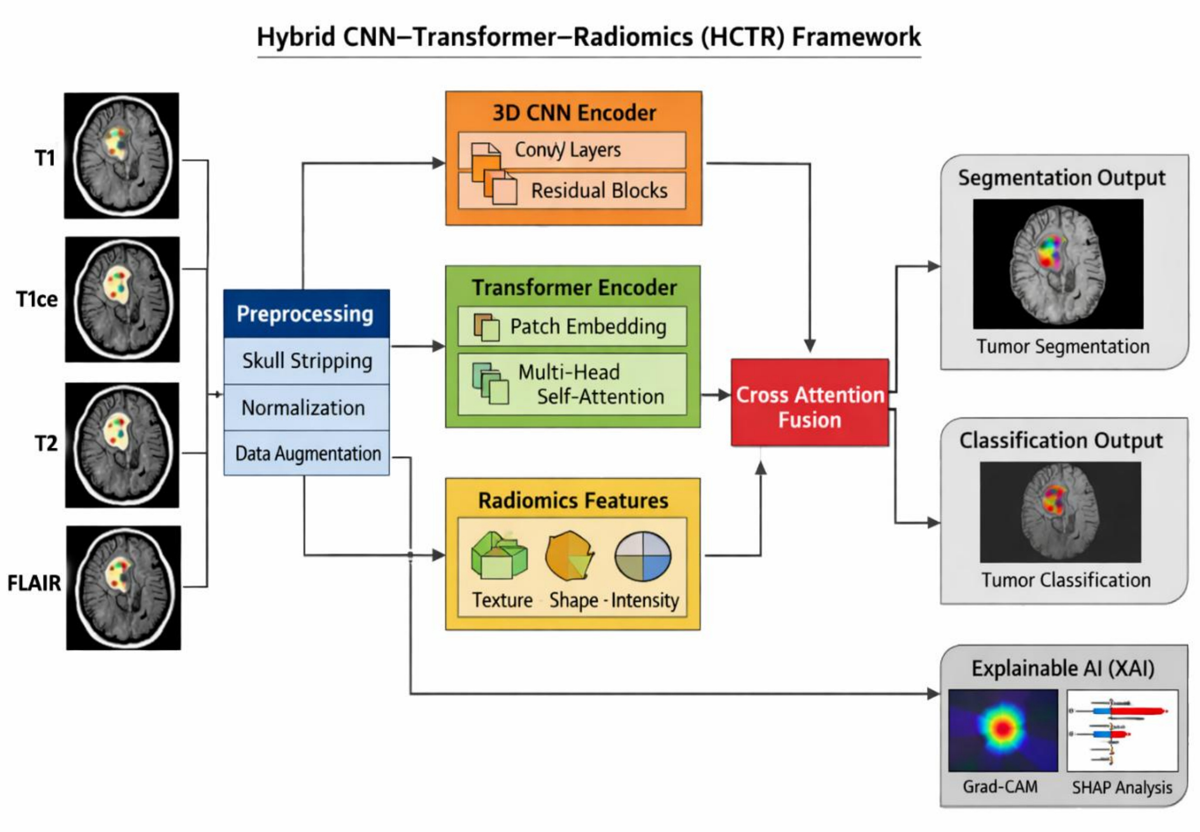

Figure 9: Architecture block diagram of HCTR

5.2. Hybrid Feature Extraction

The HCTR framework consists of the following components:

- **3D CNN Encoder:** Extracts local spatial and textural features using residual connections to facilitate gradient flow and boundary preservation. The residual learning can be performed using Equation:

$$F_{res}(x) = F(x) + x \tag{3}$$

  where $F(x)$ represents convolutional transformation and $x$is the identity mapping.

- **Vision Transformer Encoder:** Processes 3D patch-wise MRI representations using multi-head self-attention to capture long-range contextual dependencies. The Self-Attention can be expressed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

  where Q, K, and V denote query, key, and value matrices, and $d_k$is the dimensionality scaling factor.

- **Radiomics Branch:** Extracts handcrafted features (intensity, texture, shape, wavelet, and statistical descriptors) using PyRadiomics. These features are projected into a learnable latent space via a feed-forward network. The Radiomic Feature Vector can be described as:

  $$R = [r_1, r_2, \dots, r_n] \tag{5}$$

  Where, R represents the extracted radiomic feature vector projected into a latent space via:

  $$R' = \phi(W_r R + b_r) \tag{6}$$

  with learnable parameters W_r, b_r, and activation function $\phi(\cdot)$.
- **Cross-Attention Fusion Module:** Integrates deep embeddings and radiomic representations to produce a unified semantic-radiomic feature space. The Fusion function can be performed as:

  $$F_{fusion} = \text{Attention}(D, R', R') \tag{7}$$

  This produces a unified semantic–radiomic embedding.
- **Prediction and Explanation Layer:** Dual output heads perform segmentation and classification. Optimization is carried out using a compound loss function incorporating Dice loss, Cross-Entropy loss, and Focal loss. The Compound Loss can be determined as:

  $$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{CE} + \lambda_3 \mathcal{L}_{Focal} \tag{8}$$

  Where:

  Dice Loss:

  $$\mathcal{L}_{Dice} = 1 - \frac{2|P \cap G|}{|P| + |G|} \tag{9}$$

  Cross-Entropy Loss:

  $$\mathcal{L}_{CE} = -\sum_i y_i \log(\hat{y}_i) \tag{10}$$

  Focal Loss:

  $$\mathcal{L}_{Focal} = -(1 - \hat{y})^{\gamma} \log(\hat{y}) \tag{11}$$

  Such a combination achieves a tradeoff between segmentation accuracy and class imbalance. The interpretation of the model is supported with the help of Grad-CAM heatmaps, visualization of attention, and feature contributions analysis with SHAP.

5.3. Training and Optimization

Training can be performed using the AdamW optimizer with cosine annealing learning rate scheduling and early stopping. The AdamW Update Rule is given by:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{m_t}{\sqrt{v_t} + \epsilon} + \lambda \theta_t \right) \tag{12}$$

where:

- m_t and v_t are moment estimates,
- η is learning rate,
- λ is weight decay.

Cosine annealing scheduling:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + \cos\left(\frac{t}{T}\pi\right)\right) \tag{13}$$

The overfitting is reduced using the regularization strategies, such as dropout, attention normalization, and weight decay. The mixed precision training may be considered to augment the computational efficiency.

5.4. Deployment and Privacy Integration

The framework is designed for hybrid cloud-edge deployment. Inference can be performed locally on the imaging devices, and the global update of the models can be supported with the help of federated learning. The architecture aids in adhering to privacy regulations like the HIPAA, GDPR, and is designed to interoperate with hospital information systems and electronic health records. The modular architecture can be incorporated into the telemedicine and smart healthcare diagnostic processes. The Federated Averaging is defined as:

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_t^{(k)} \tag{14}$$

where:

- $w_t^{(k)}$ represents local model weights,
- n_k is dataset size of client k,
- n is total sample size.

This ensures decentralized training while preserving privacy.

6. HYBRID ARCHITECTURES AND MODEL INTERPRETABILITY

With the trend of continuously increasing the complexity of deep learning models to detect brain tumors, the possibility of integrating hybrid architectures and interpretability mechanisms has taken on significant research interest. Hybrid deep learning systems aim to enhance the diagnostic accuracy and generalization by using a combination of various techniques. This is usually obtained through addition of conventional machine learning methods with more advanced deep learning models. Meanwhile, interpretability or explainability has also been emerged, particularly in the medical community, where black-box models present several ethical, legal and practical issues. The section will be devoted to the hybrid architectures that combine the handcrafted and learned features, interpretable AI methods, such as Grad-CAM and SHAP, and analyses the underlying trade-off between the high degree of accuracy and the retained interpretability in the brain tumor classification [31].

6.1. Feature-Level Fusion: Traditional + Deep Learning Approaches

In medical image analysis, state-of-the-art performance has been achieved by combining handcrafted features such as texture, intensity, and shape with deep features. This approach has shown strong robustness, particularly in brain tumor detection tasks that use CNNs and related models. Whereas the traditional radiomic features can be used to capture pattern of interest in a field of knowledge by use of clinical knowledge, the deep learning features automatically generate a high-level abstract representation of data.

Hybrid approaches generally follow two strategies:

- *Intermediate-Layer Fusion*: Traditional features such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), or Gray Level Co-occurrence Matrix (GLCM) features are fused with deep features at intermediate layers.
- *Multi-Branch Architectures*: Separate branches process handcrafted and deep features independently before fusion at higher levels.

Studies indicate that such hybrid approaches frequently achieve higher accuracy particularly when dealing with smaller datasets. For example: CNNs with texture measures based on GLCM have been demonstrated to enhance the sensitivity and specificity of prediction of brain tumors, particularly of the high-grade gliomas, which have complex interior texture. Despite all the benefits of the feature-level fusion, it includes more complexity and may require more attention with preprocessing, feature normalization, and tuning purposes because of such issues as feature redundancy or overfitting. Figure 10 shows the mean accuracy of different DL models; hybrid models are reported to achieve the highest average accuracy (94.2%). This suggests that combining different architectures or types of features is generally more effective in the task of detection. Figure 11 summarizes reported performance metrics (Accuracy, F1-Score and Specificity) of different types of DL models. The bars in the stack indicate that even though all the types of models can be highly performing, each of them has a variance in the balance between these critical measures.

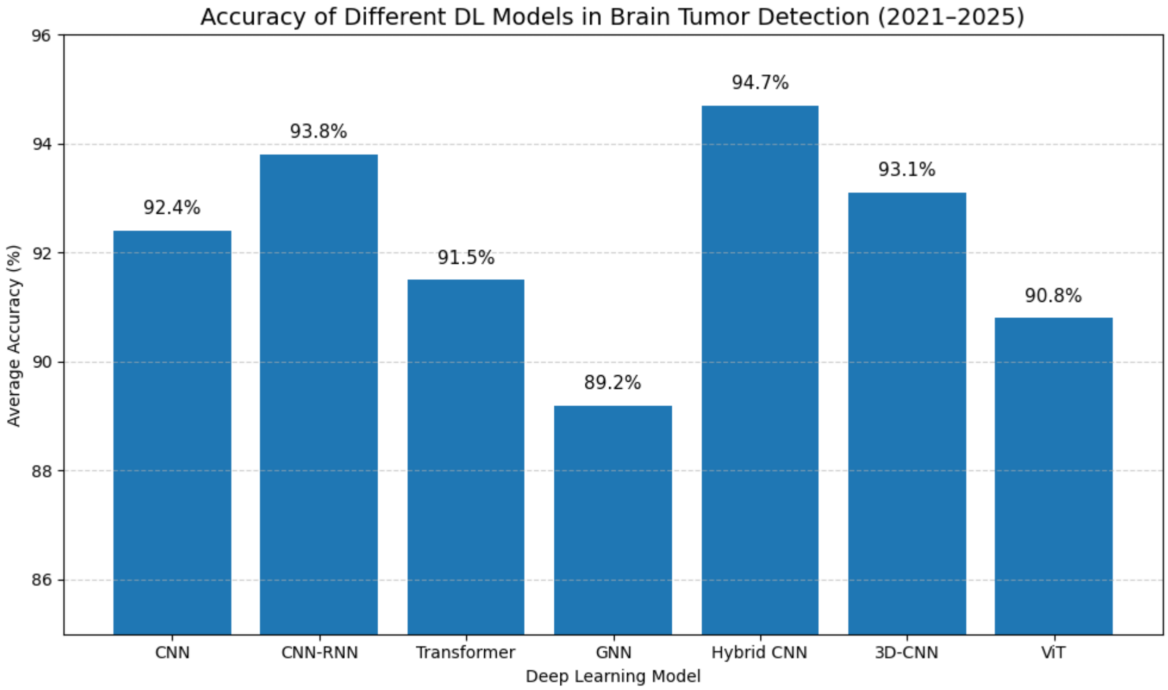

Figure 10: Accuracy of Different DL Models in Brain Tumor Detection (2021–2025)

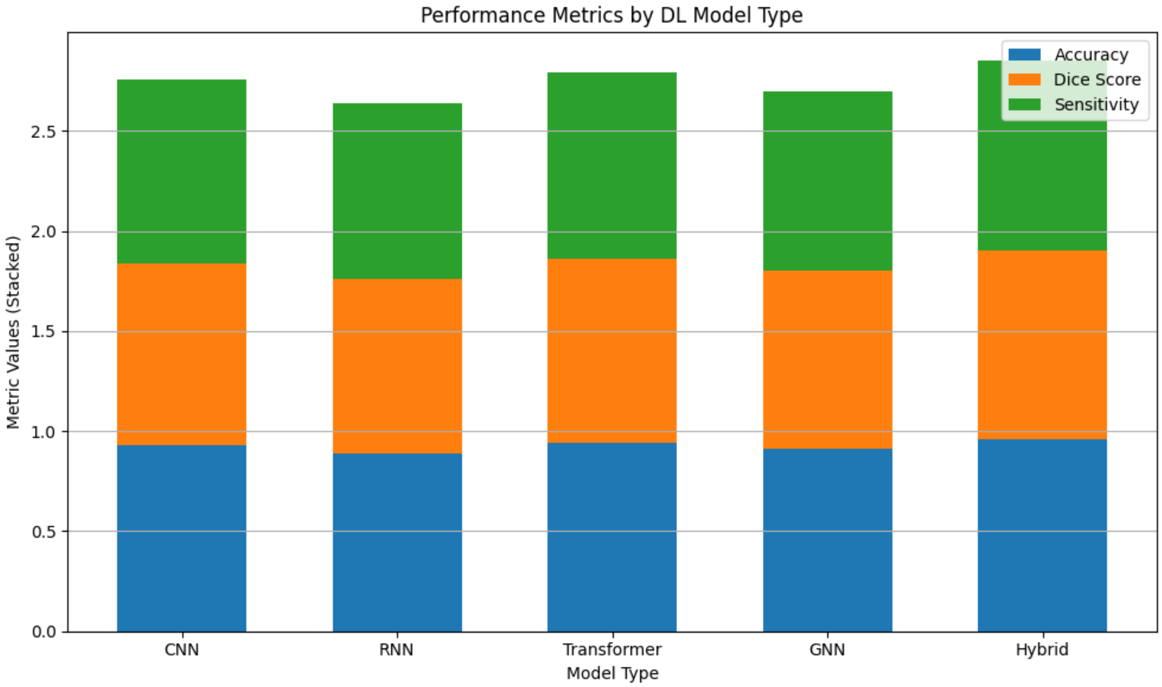

Figure 11: Performance Metrics by DL Model Type

6.2. Interpretability Through Explainable AI (XAI): Grad-CAM, SHAP, and Others

Among the most urgent constraints of deep learning in clinical diagnostics, most models are considered black-box systems. In contrast to traditional decision trees as used, or rule-based systems, deep neural networks consist of thousands or millions of parameters, making their underlying decision logic opaque. This lack of transparency that can inhibit trust, particularly in high stakes in medical functions which include the diagnosis of brain tumor. In order to mitigate this, researchers adopt Explainable AI (XAI) methods which offer either visual or semantic interpretations to the predictions of a model.

Two commonly used techniques in medical imaging are:

- *Grad-CAM (Gradient-weighted Class Activation Mapping)*: Grad-CAM visualizes parts of an original image that the model used to make the most significant contributions to the decision. In brain MRI classification or segmentation, Grad-CAM heatmaps may be used to indicate tumor regions responding to a certain output (e.g., glioma vs. meningioma) and thus guide radiologists to interpret whether the model is concentrating on clinically significant areas.
- *SHAP (SHapley Additive exPlanations)*: SHAP is a method that attributes a score to each feature that is based on its contribution to the prediction. Whereas Grad-CAM is primarily applied to image data, SHAP can be used on tabular or imaging data. In hybrid models and features with radiomics and deep features, SHAP assists in the interpretation of the extent to which some features contributed to the overall classification results.

Additional techniques include LIME, attention visualization, and Integrated Gradients. These approaches improve alignment between AI outputs and clinical reasoning.

6.3. Multi-Modal and Multi-Task Learning

Multi-modal learning is another hybrid approach in which information provided by various modalities in imaging (MRI, PET, CT, histopathology) is also used in order to enhance the learning. As an example, a hybrid system can be built where CNNs are used to extract MRI and histopathology images, followed by their combination in making a final diagnosis. These models take advantage of complementary data e.g. structural data from MRI and cellular data in histopathology to more accurately and comprehensively evaluate. Another example of such hybrid setup is multi-task learning where one model is trained to accomplish several inter-related tasks, e.g., segmentation, and classification. These models are typically more efficient during training and inference compared to single-task networks and from learning shared representations, these models even significantly outperform single-task networks on multiple tasks. As an illustration, tumor segmentation can be implemented in U-Net like architecture and tumor grading can be implemented in classification head. This shared learning paradigm enhances generalization, reducing the risk of overfitting, by adding any supplementary learning goals.

6.4. Trade-off Between Accuracy and Interpretability

Although hybrid architectures and deep models are often more accurate, in general, they lack interpretability. State-of-the-art models such as transformers, deep CNNs or ensemble networks can be trained to perform better, but are more difficult to interpret, due to their complex internal mechanisms. Alternatively, models or hybrid systems with handcrafted features are less predictive of the future and more easily explained, assuming models with a lower predictive power are more comprehensible. This trade-off is a focus of clinical adoption. In practice, having less accurate but more transparent model can be even desirable as compared to a very accurate black-box model. Regulatory demands like the GDPR in Europe or the HIPAA in the US also outline the need to make the medical AI tools interpretable, especially in cases where AI recommendations affect patient care decisions.

The recent trend of research is on an optimal balance by:

- *Post-hoc Interpretability*: how to explain post-hoc explanation of trained models with tools such as Grad-CAM, SHAP.
- *Intrinsic Interpretability*: Architectures that allow models to be interpreted at the design level (e.g. prototype learning).
- *Human-in-the-Loop System*: Making use of models with predictions and human response during training and inference.

The yearly growth in average accuracy across model types is illustrated in Figure 12.

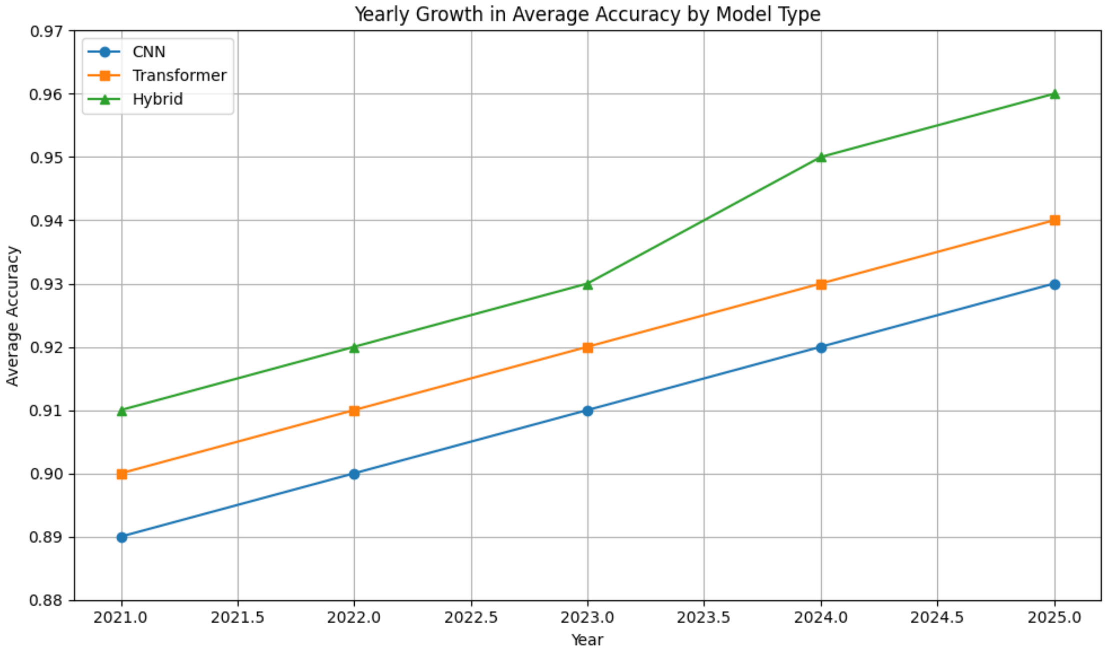

Figure 12: Yearly Growth in Average Accuracy by Model Type

7. SMART HEALTHCARE INTEGRATION

Smart healthcare and deep learning represent a major innovation in the field of diagnosis and management of brain Tumors. Traditional diagnostic techniques critically depend on labor-intensive, subjective and highly variable manual interpretation of radiology imaging. In comparison, smart healthcare uses such technologies as artificial intelligence (AI), the Internet of Things (IoT), real-time analytics, cloud/edge computing to provide quality care faster, more customized and more predictable. One of the most significant growths is associated with the creation of AI tools that could process medical images in a matter of seconds, helping clinicians analyze the situation in such aspects as the segmentation of tumors, their classification, and monitoring of progression. The systems include deep learning models trained in advance to process MRI and CT scans within seconds producing probability maps, labels, and clinical insights. Being incorporated into hospital PACS, they facilitate workflow and exclude manual data manipulations those are the necessary processes on the way to timely planning of neurosurgery and diagnosis. Cloud and edge computing structures are usually applied in implementing smart healthcare systems. Cloud computing and storage is also scalable and applied to train large networks and work with complex multi-modal brain images. Ready-made machines can be deployed remotely and reduce the amount of hardware in the vicinity. Furthermore, cloud systems allow collaboration in diagnostics such that a remote expert can see and review the same patient information in real time.

Despite these advancements, challenges related to latency, privacy, interoperability, and clinician training must be systematically addressed to enable sustainable clinical adoption.

8. CONCLUSION AND FUTURE DIRECTIONS

This study presents a Hybrid CNN-Transformer-Radiomics (HCTR) framework for brain tumor detection. The framework integrates convolutional feature extraction, transformer-based contextual modeling, and radiomic feature analysis within a unified architecture. The proposed framework addresses the key challenges in neuro-oncological imaging. It includes limited interpretability, generalization constraints, and deployment feasibility in real clinical environments. By integrating CNN-based spatial learning, transformer-based global reasoning, and handcrafted radiomic descriptors, the HCTR model seeks to achieve strong performance, robustness, and transparency.

The model incorporates explainability techniques such as Grad-CAM and SHAP. These methods provide spatial and feature-level insights into model predictions and improve clinical interpretability. In addition, the architecture is scalable to be used in a hybrid cloud-edge and privacy-conscious learning schemes, which can be used in smart healthcare ecosystems. Overall, the HCTR framework provides a systematic approach for developing reliable, explainable and clinically adaptable AI-based systems for brain tumor diagnosis. It can lead to the development of intelligent neuro-oncology systems.

Although deep learning has significantly improved the detection of brain tumors, but there are still challenges linked to generalization, robustness, and clinical implementation. Future research should emphasize federated learning for privacy-preserving multi-institutional training, domain adaptation to reduce scanner and demographic bias, and lightweight model optimization for the edge deployment.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, et al., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA: Cancer J. Clin. 71 (2021), 209-249. https://doi.org/10.3322/caac.21660.

[2] I. Intan, B.J. Nelwan, M.M. Henry, A.S. Karnyoto, R.E. Puspitasari, B. Pardamean, Deep Ensemble Transfer Learning for Detecting Breast Cancer in Histopathological Images, Commun. Math. Biol. Neurosci. 2024 (2024), 101. https://doi.org/10.28919/cmbn/8823.

[3] H.H. Muljo, B. Pardamean, G.N. Elwirehardja, A.A. Hidayat, D. Sudigyo, et al. Handling Severe Data Imbalance in Chest X-Ray Image Classification with Transfer Learning Using SwAV Self-Supervised Pre-Training, Commun. Math. Biol. Neurosci. 2023 (2023), 13. https://doi.org/10.28919/cmbn/7526.

[4] M. Li, Y. Jiang, Y. Zhang, H. Zhu, Medical Image Analysis Using Deep Learning Algorithms, Front. Public Health 11 (2023), 1273253. https://doi.org/10.3389/fpubh.2023.1273253.

[5] D. Hussain, N. Abbas, J. Khan, Recent Breakthroughs in Pet-Ct Multimodality Imaging: Innovations and Clinical Impact, Bioengineering 11 (2024), 1213. https://doi.org/10.3390/bioengineering11121213.

[6] E.K. Lee, E.J. Lee, S. Kim, Y.S. Lee, Importance of Contrast-Enhanced Fluid-Attenuated Inversion Recovery Magnetic Resonance Imaging in Various Intracranial Pathologic Conditions, Korean J. Radiol. 17 (2016), 127. https://doi.org/10.3348/kjr.2016.17.1.127.

[7] D.Y.F. Chung, D. Mondal, E.J. Holmes, R. Misra, Emergency Cross-sectional Radiology, Cambridge University Press, Cambridge, 2012.

[8] J.R. Fink, M. Muzi, M. Peck, K.A. Krohn, Multimodality Brain Tumor Imaging: MR Imaging, PET, and PET/MR Imaging, J. Nucl. Med. 56 (2015), 1554-1561. https://doi.org/10.2967/jnumed.113.131516.

[9] A.B. Abdusalomov, M. Mukhiddinov, T.K. Whangbo, Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging, Cancers 15 (2023), 4172. https://doi.org/10.3390/cancers15164172.

[10] H.A. Shah, F. Saeed, S. Yun, J.H. Park, A. Paul, et al., A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned EfficientNet, IEEE Access 10 (2022), 65426-65438. https://doi.org/10.1109/ACCESS.2022.3184113.

[11] E.H. Ibrahim, S.E. Abo-Youssef, K. El-Bahnasy, K.A.M. Fathy, A New Approach for Brain Tumor Detection Using Machine Learning, Dubai Med. J. 7 (2024), 160-176. https://doi.org/10.18502/dmj.v7i3.17732.

[12] S. Saeedi, S. Rezayi, H. Keshavarz, S. R. Niakan Kalhori, MRI-Based Brain Tumor Detection Using Convolutional Deep Learning Methods and Chosen Machine Learning Techniques, BMC Med. Inform. Decis. Mak. 23 (2023), 16. https://doi.org/10.1186/s12911-023-02114-6.

[13] M.I. Mahmud, M. Mamun, A. Abdelgawad, A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks, Algorithms 16 (2023), 176. https://doi.org/10.3390/a16040176.

[14] E.H. Ibrahim, S.E. Abo-Youssef, K. El-Bahnasy, et al. Early Diagnosis of Brain Tumor Using Machine Learning Approaches, J. Inf. Syst. Eng. Manag. 10 (2025), 1089-1100. https://doi.org/10.52783/jisem.v10i34s.5913.

[15] R. Preetha, M.J.P. Priyadarsini, J.S. Nisha, Automated Brain Tumor Detection from Magnetic Resonance Images Using Fine-Tuned Efficientnet-B4 Convolutional Neural Network, IEEE Access 12 (2024), 112181-112195. https://doi.org/10.1109/ACCESS.2024.3442979.

[16] D. Beaulah Princiba, P. Ezhilarasi, S. Rajeshkannan, Integrated EfficientNetB3V2 Fused MaxEnt Classifier Model for Brain Tumor Classification in MR Images, J. Chin. Inst. Eng. 48 (2025), 254-267. https://doi.org/10.1080/02533839.2025.2464551.

[17] S. Ahmad, P.K. Choudhury, On the Performance of Deep Transfer Learning Networks for Brain Tumor Detection Using MR Images, IEEE Access 10 (2022), 59099-59114. https://doi.org/10.1109/ACCESS.2022.3179376.

[18] S. Asif, W. Yi, Q.U. Ain, J. Hou, T. Yi, et al., Improving Effectiveness of Different Deep Transfer Learning-Based Models for Detecting Brain Tumors from MR Images, IEEE Access 10 (2022), 34716-34730. https://doi.org/10.1109/ACCESS.2022.3153306.

[19] S. Banik, N.G. Barai, F.M.J.M. Shamrat, Blockchain Integrated Neural Networks: A New Frontier in MRI-based

Brain Tumor Detection, Int. J. Adv. Comput. Sci. Appl. 14 (2023), 954-964.

[20] Amita Jajoo, Implementing Deep Learning Models for Accurate Brain Abnormality Detection and Positioning Using MRI and CT, J. Inf. Syst. Eng. Manag. 10 (2025), 717-736. https://doi.org/10.52783/jisem.v10i11s.1671.

[21] A. Raj, D. Rawat, D. Verma, S. Pandey, D. Rana, et al., Comparative Study of Different Models on Mri-Based Brain Tumor Detection, in: 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI), IEEE, 2025, pp. 1292-1297. https://doi.org/10.1109/IC3ECSBHI63591.2025.10991161.

[22] S. Kumar, A. Kumar, A. Jaiswal, A Low Complexity MobileNetV2 Based CNN Model for Brain Tumor Detection in MRI Images, in: 2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), IEEE, 2024, pp. 1-7. https://doi.org/10.1109/IAICT62357.2024.10617450.

[23] N. Nasar, M.M. Thiruthuvanathan, A. Pauls, Brain Tumor Detection Using Hyper Parameter Tuning and Transfer Learning, in: 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), IEEE, 2023, pp. 1-6. https://doi.org/10.1109/InC457730.2023.10263136.

[24] V.K. Dhakshnamurthy, M. Govindan, K. Sreerangan, et al. A. Thomas, Brain Tumor Detection and Classification Using Transfer Learning Models, Eng. Proc., 62 (2024), 1. https://doi.org/10.3390/engproc2024062001.

[25] G.D. Kumar, S.N. Mohanty, Brain Tumor Detection Based on Multiple Deep Learning Models for MRI Images, EAI Endorsed Trans. Pervasive Health Technol. 10 (2024), 1-8. https://doi.org/10.4108/eetpht.10.5499.

[26] C. Li, F. Zhang, Y. Du, H. Li, Classification of Brain Tumor Types Through MRIs Using Parallel CNNs and Firefly Optimization, Sci. Rep. 14 (2024), 15057. https://doi.org/10.1038/s41598-024-65714-w.

[27] A.S. Musallam, A.S. Sherif, M.K. Hussein, A New Convolutional Neural Network Architecture for Automatic Detection of Brain Tumors in Magnetic Resonance Imaging Images, IEEE Access 10 (2022), 2775-2782. https://doi.org/10.1109/ACCESS.2022.3140289.

[28] M.Z. Khan, F.U. Zaman, P. Bhatti, A.A. Khan, S.A. Sultan, et al., A Comparative Analysis of Deep Learning Architectures for Efficient Brain Tumor Detection, in: 2024 International Conference on Sustainable Technology and Engineering (i-COSTE), IEEE, 2024, pp. 01-06. https://doi.org/10.1109/i-COSTE63786.2024.11024836.

[29] P.K. Ramtekkar, A. Pandey, M.K. Pawar, A Comprehensive Review of Brain Tumour Detection Mechanisms, Comput. J. 67 (2023), 1126-1152. https://doi.org/10.1093/comjnl/bxad047.

[30] T. Boopathy, M. Kavitha, Hybrid Deep Learning for Brain Tumor Detection Using Transfer Learning Models, in: 2025 International Conference on Next Generation Communication and Information Processing (INCIP), IEEE, 2025, pp. 239-242. https://doi.org/10.1109/INCIP64058.2025.11019932.

[31] M. Aamir, Z. Rahman, Z.A. Dayo, W.A. Abro, M.I. Uddin, et al., A Deep Learning Approach for Brain Tumor

Classification Using MRI Images, Comput. Electr. Eng. 101 (2022), 108105. https://doi.org/10.1016/j.compeleceng.2022.108105.

[32] F. Pérez-García, R. Sparks, S. Ourselin, TorchIO: A Python Library for Efficient Loading, Preprocessing, Augmentation and Patch-Based Sampling of Medical Images in Deep Learning, Comput. Methods Programs Biomed. 208 (2021), 106236. https://doi.org/10.1016/j.cmpb.2021.106236.