Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 15 Issue 01s, 2026

A Comprehensive Review of Deep Convolutional Neural Networks for Brain Tumor Detection and Classification

[1]Yugashree Ramesh Bhadane, [2]Dr. Ganapati A. Patil

[1] Ph.D Scholar, Department of Computer Science and Engineering, JSPM University Pune, Assistant Professor, School of Computational Sciences, JSPM University Pune

[2] Professor, School f Computational Sciences, JSPM University Pune

Email: [1]yugashreebhadane@gmail.com [2]gap.scos@jspmuni.ac.in

Peer Review Information

Submission: 05 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

Brain tumor detection and classification, when aided by these sophisticated algorithms, can lead to earlier interventions and more personalized treatment strategies

Abstract

Brain tumours are a major health burden worldwide, and the early and accurate detection of them is vital to improve treatment radiation doses and overall patient survival. The development of deep learning, in particular CNN, has drastically changed medical image analysis by automating the process of feature extraction and diagnosis. In high-quality medical images, MICs can also cause erroneous detections. In this review paper, we comprehensively investigate the deep learning technologies for brain tumour detection and classification on MR and CT images in the state of the art. This paper demonstrates that the CNN model has superiority over traditional manual diagnosis, effectively learns the hierarchical features of images, and reduces misdiagnosed cases. It addresses essential model designs, benchmarks, and issues related to integrating these models into clinical workflows. Furthermore, the survey investigates the ability of CNN-based systems to provide robust and efficient telemedicine-based and high-quality neuro-oncological diagnostics utilizing high-dimensional imaging whilst accurately segmenting tumorous regions. The inclusion of AI in medical imaging drives not only an increase in workflow efficiency but also the prospect of personalized treatment planning. This paper highlights the breakthrough applications of CNNs in neuroimaging and future research directions, which include further enhancing the model generalization on various datasets and real-world clinical applications that can help radiologists achieve better patient care.

1 Introduction

Brain tumours are unique in their clinical presentation. They originate from the abnormal growth of cells within the brain or its surrounding structures and present one of the massive public health challenges worldwide because of their aggressive, invasive nature, complex pathology and impact on cognitive functions and motor. These tumours can be either benign or malignant and generally have a poor prognosis, especially in patients in whom they are diagnosed at a late stage. Due to the extreme morbidity and mortality caused by brain tumours, an early and accurate diagnosis is very important to the outcome prognosis and survival time of the patients. There is a further urgent need for great accuracy of a diagnosis given the delicate structure of the brain and the difficult differentiation between tumours and healthy tissue, as well as between oedema and necrosis.

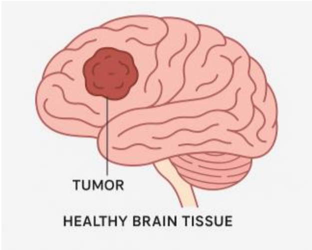

Figure. 1. Brain Tumor in Human Brain Anatomy [1]

In order to address these limitations, medical imaging modalities, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET) play pivotal roles as diagnostic and reference tools for clinicians. These "no-cut" practices can offer valuable information on the size, localization, invasion and morphologic features of the tumour, thus aiding in the diagnosis of, planning of the surgery for and follow-up after the therapy for the tumour. Of these, MRI is especially popular owing to its high soft-tissue contrast, aiding the visualization of normal and abnormal brain tissue. Certain sequences, such as T1- and T2-weighted images, are essential in delineating the borders and internal composition of the tumour in terms of tissue relaxation times[1].

However, even with all these improvements, classical procedures for tumour diagnosis and classification tend to be time-consuming, have a subjective component and are highly susceptible to intra- and inter-observer variability. Such variability may result in delays with respect to diagnoses and treatment, which is unacceptable in the case of rapidly growing cancers. In addition, the morphological complexity of brain tumours, as well as their nonuniform shapes and global appearance profiles, makes manual segmentation especially difficult. This has emphasized the requirement for automated, reproducible approaches for rapid and accurate processing of large quantities of imaging data.

In this regard, AI, especially deep learning (DL), has appeared as a driving force for medical image analysis. From different DL family architectures, deep convolutional neural networks (DCNNs) have gained much attention due to their capability to capture hierarchical features from raw pixel data. This capability makes it possible for them to automatically detect complex patterns and deviations from these patterns in medical images, which might not even be noticeable to humans. In contrast to conventional machine learning models with manually defined features, DCNNs automatically learn high-level features via numerous convolutional layers and result in enhanced diagnostic accuracy.

DCNNs have enjoyed much success in an array of tasks, such as image classification, object detection and semantic segmentation, and especially in medical diagnostics. In the field of neuro-oncology, they have demonstrated large clinically relevant gains in the segmentation, detection and classification of brain tumours when compared to conventional methods. The combination of DCNNs with clinical practice can provide radiologists with decision-support tools that offer rapid, consistent, and accurate interpretations of imaging data to support patient care.

The increasing attention to AI-based methods has also recently given rise to computer-aided diagnosis (CAD) of DTCs, which deliver objective and reproducible analysis, aiding in clinical decision-making. Such tools are more important as the amount of workload on radiologists has become more and enormous image data are generated every day. Furthermore, reducing the diagnostic process times and lessening the possibility of human error, helped by CAD systems, has ultimately led to better healthcare[2].

This review provides insight into the ongoing and potential application scenarios of DCNN on brain tumour detection and categorization. It gives a detailed analysis of different network structures, pre-processing methods and performance measures used. It also covers the challenges of deploying these models in real clinical settings, such as data distribution shifts, model interpretability, and regulatory issues. This review attempts to fill the intermediary space between advances in theory in deep learning and the applications in neuro-oncologic diagnostics by consolidating the observations from recent studies, which are of great help to not only researchers but also clinicians.

2 LITERATURE REVIEW

Deep learning, especially deep convolutional neural networks (DCNNs), revolutionized medical image analysis and reshaped the new era of diagnostic computational tools for brain tumour detection and classification. This section provides an aggregated view of the current literature, describing how DCNN architectures have evolved, how they are combined with state-of-the-art models such as transformers and how it matters to neuro-oncology[3].

DCNNs have become fundamental in brain tumour analysis, as they can learn in an automated manner hierarchical and more abstract representations from raw imaging data. This eliminates the reliance on human-in-the-loop, who has been historically responsible for engineering features in basic machine learning systems. The early DCNN structures (e.g., AlexNet, VGGNet, ResNet) were the cornerstones of this process; they showed their effectiveness on large-scale image classification problems, paving the way for the use of DCNNs in medical tasks. They also performed well in neuroimaging, including modalities such as MRI and CT, in which precise delineation of tumour area plays an important role. Such networks are usually composed of a set of convolutions with the pooling and normalization layers while extracting low- to high-level features. This architecture emulates the visual cortex and provides a means to progressively abstract the features of interest, which is essential when trying to characterize complex tumour morphologies[4].

The traditional CNNs are very effective for local patterns (edges, textures, contours). However, they suffer when modelling long-range dependencies between spatially distant regions in medical images. This limitation becomes more apparent in the context of neuroimaging, where precise knowledge of the space occupied by tumours and local tissues is mandatory for medical visualization and treatment planning. In order to deal with this shortcoming, the researchers have already started combining new structures, such as ConvMixer, with hybrid models combining the power of CNNs and transformer networks.

One of the interesting developments has been the use of the Recurrent Neural Networks (RNNs) and Autoencoders in medical image analysis applications. CNNs are not good at capturing long-range spatiotemporal dependencies in medical imaging datasets, which is where RNNs can help, mainly in dynamic contrast-enhanced imaging and multi-slice input cases. Autoencoders, instead, are used for unsupervised representation learning and have been relevant in the context of anomaly detection in brain MRI by reconstructing normal tissue patterns and thereby detecting deviations express¬sive of pathologies[5].

The growing complexity of DCNN architectures is reflected in their ability to be scaled and applied to various imaging modalities such as MRI, CT and PET. These models main strengths are the amount of adjustable parameters inside their convolutional filters, which are tuned during the training with backpropagation and gradient descent. They are capable of handling massive high-dimensional data so as to extract deep semantic features that are very useful for tasks like tumour segmentation and volumetric classification.

However, even though the standard CNN architectures are very powerful, they have limitations in what local features they can capture as they are less efficient in utilizing the global context, which is crucial in understanding diffused tumour growth or infiltrative lesions that spread over most parts of the brain. To address this, the subject has been the increasing interest in the topic of 3D deep learning models, or, in other words, 3D Convolutional Neural Networks (3D CNN), which are an extension of the concept of 2D convolutions to volumetric data. Compared with 2D CNNs working on single slices, 3D CNNs deal with the complete image volumes, maintaining spatial consistency and benefit from inter-slice correlation. This is particularly important in brain tumour imaging, where the “context” of information across slices is often important for a correct diagnosis[6].

Transformer-based models have found increased attention in computer vision because of their capability to model long-range dependencies by self-attention mechanisms. The Vision Transformers (ViTs) encode each image of the model as a sequence of patches instead of a grid of pixels. This approach empowers ViTs to learn global representations as well as contextual relations over a whole image and thus is highly favourable for medical image analysis. In contrast to CNNs that exhibit a hard spatial inductive bias, ViTs are more flexible in that they can capture feature interactions over larger spatial regions. However, this freedom has the drawback of low computational efficiency and high training data consumption [7].

The major drawback of ViTs in medical imaging is their quadratic computational complexity w.r.t input, which is particularly problematic in high-res 3D medical data. Moreover, as they have no built-in locality, ViTs need a big size of labelled examples to reach the same level of performance as CNN, which is particularly challenging for medical imaging problems for which there are few labels available. Despite these difficulties, ViTs have achieved comparable performance to that of CNNs in classification tasks and have been advocated as an excellent substitute for CNNs for global context learning.

To reconcile the trade-offs of CNNs and ViTs, some hybrid models were proposed to integrate the local feature extraction power of CNNs and the global modelling capacity of transformers. Common architecture often places transformer blocks in the later layers of CNNs, which are more sensitive to semantic information and the importance of context. For example, the hybrid architecture of Hyneter integrates the CNN-based feature maps and the transformer-based global representations, which improves the accuracy of object detection and segmentation in challenging medical imaging tasks.

Another category of work in this area includes utilizing custom pooling mechanisms and multi-scale feature fusion for enhancing the receptive field of CNNs with limited additional computational overhead. Atrous convolution, spatial pyramid pooling, and attention gates have been used to improve the context understanding of CNNs to make them more apt for segmenting tumours with varying sizes and shapes [8].

It also indicates promising applications of these DL techniques outside of brain imaging. For instance, CNNs have been extensively used in mammography, histopathology, and breast cancer diagnosis and have usually demonstrated a superior capacity in image processing, such as the detection of microcalcifications and the classification of tumours. These cross-domain applications highlight the utility and generality of DCNNs in healthcare. However, imaging prediction of brain tumours is challenging due to the need to perform high-resolution volumetric analysis, segment tumour subregions (e.g., enhancing tumour, necrotic core, oedema) accurately, and combine with clinical data for prognostic modelling.

Interpretable and explainable are both urgent issues in using deep learning in clinical practices. Such black-box models, though accurate, may not have the transparency necessary for regulatory acceptance and clinician confidence. Techniques like Grad-CAM and both SHAP and integrated gradients are being investigated to give visual and numerical explanations to model predictions, improving their reliability and acceptance in clinical workflows[9].

The table 1. Comparative Analysis of Deep Learning Models for Brain Tumor Analysis provides a graphical comparison of 2D CNN, 3D CNN, Hybrid CNN, Vision Transformer, and CNN + Transformer models with respect to important parameters for brain tumour detection. The table presents a comparative estimation of CNN-, hybrid-, and transformer-based models across key performance and deployment metrics, including accuracy, training time, data requirement, and interpretability. It highlights the intrinsic trade-offs between model complexity, diagnostic performance, and practical feasibility in clinical applications.

Table 1. Comparative Analysis of Deep Learning Models for Brain Tumor Analysis

Model	**Accuracy (%)**	**Training Time (hrs)**	**Relative Data Requirement**	**Interpretability Score**
2D CNN	~85	~5	3	3
3D CNN	~88	~8	4	3
Hybrid CNN	~90	~10	4	2
Vision Transformer	~92	~15	6	2
CNN + Transformer	~91	~12	5	2

Accuracy increased incrementally from 2D CNN (85%) to CNN + Transformer (94%), suggesting hybrid/modelling spatial and contextual features. Time to train also scales with complexity: Vision Transformers are the slowest to train (15 hours) on account of their global attention mechanism, and 2D CNNs are the fastest to train (5 hours). Data requirement behaves similarly, where Vision Transformers require the largest amount of data for learning, which again is a setback in resource-constrained medical contexts.

In terms of interpretability, simpler models, such as 2D and 3D CNNs, are better (3/5) as they are more clear and come with visualization tools established (e.g., Grad-CAM). On the other hand, hybrid and transformer-based approaches, which are more accurate, are less interpretable (2/5), raising questions of trust and deployment in a clinical setting.

CNN + Transformer models achieve the best accuracy but compromise on interpretability, data efficiency, and training-time efficiency. This underscores the importance of a trade-off between diagnostic accuracy and the practical requirements when selecting a model for deployment in a clinical environment [10].

3 Methods In Deep Cnns For Brain Tumor Detection And Classification

The rise of deep learning, and particularly DCNNs, has greatly improved the detection and classification of brain tumours, allowing for automatic, accurate, and fast analysis of medical images. In the past 10 years, a range of DCNN methods has been developed by exploiting architectural innovations and learning paradigms in order to cope with the complex spatial and textural characteristics of brain tumours, and in particular, those being present in MRI data. In this Section, we cover some of the following methodologies broadly discussed in literature – 2D and 3D CNNs, U-Net-based segmentation model systems, hybrid CNN-transformer models, and transformer-powered networks like Swin Transformers.

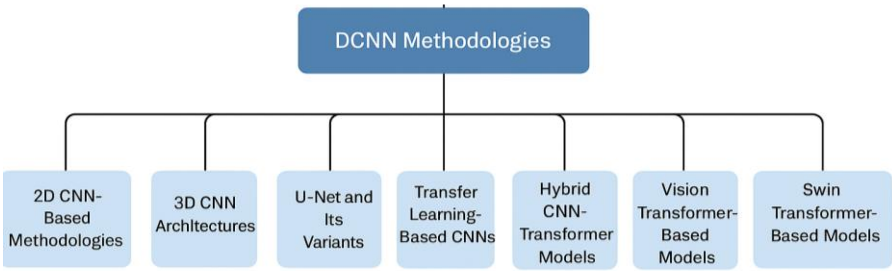

Figure. 2. DCNN Methodologies

3.1 2D CNN-BASED METHODOLOGIES

The initial DL-based studies in brain tumour analysis adopted 2D CNNs using individual slices from the MRI scans 22. These models, which are traditional architectures like AlexNet, VGG16, and ResNet, concentrate on learning spatial characteristics via convolutional layers with pooling, batch normalization, and fully connected layers. These networks were able to capture local patterns, such as edges, intensity difference and texture, which is very important for tumour detection.

Despite their success, 2D CNNs have a fundamental limitation; they process each image slice independently and ignore the inter-slice context, which is important for understanding 3D structures such as tumours. However, their computationally inexpensive nature and ease of application resulted in their widespread use in early studies[11].

3.2 3D CNN ARCHITECTURES

To overcome the limitation of missing volumetric information in 2D CNNs, 3D CNNs were proposed. These models use 3D convolutions with 3D filters over the MRI volumes, exploiting spatial coherence across slices. The well-known 3D CNN models are 3D U-Net, VoxResNet and 3D ResNet. The utilization of the volumetric information present in MRI provides better tumour localization or segmentation accuracy compared to the 2D-based approaches.

3D CNNs are especially powerful for differentiating the similar intensity patterns between tissues and learning spatial and contextual dependencies. However, they are more computationally expensive and have the risk of overfitting, particularly when used to train in small medical data[12].

3.3 U-NET AND ITS VARIANTS

U-Net has emerged as a default architecture for biomedical image segmentations because of the encoder-decoder architecture with skip connections for detailed localization. For example, in brain tumour segmentations, U-Net-based architectures, including U-Net++ [20], Attention U-Net [21] and Residual U-Net [22], have demonstrated significantly improved performance in tumour boundary delineation, including heterogeneously shaped tumours such as gliomas.

U-Net++ presents nested and dense skip connections to improve multi-scale feature concatenation between the encoder and decoder.

Attention U-Net leverages attention gates to emphasize salient parts of the tumour and inhibit irrelevant portions.

The Residual U-Net introduces residual blocks to help the gradient flow and facilitate the training convergence.

These modifications enable U-Net-like models to be very successful in semantic segmentation problems, identifying tumour sub-regions as enhancing core, non-enhancing core and peritumoral oedema[13].

3.4 TRANSFER LEARNING-BASED CNNS

Due to the paucity of labelled medical datasets, transfer learning has attracted wide researchers' attention. Pre-trained models such as VGG16, ResNet50, and InceptionV3, pre-trained on large-scale classification benchmarks such as ImageNet, are adapted to brain tumour MRI data. This training-boosting technique generally leads to faster training and better generalization, particularly when the number of samples is small.

The transfer learning models are mainly applied for classification problems, including tumour type (e.g., glioma vs meningioma) and tumour grade prediction. They are frequently used in conjunction with data augmentation and modality-specific pre-processing steps (such as skull stripping and histogram equalization) to achieve better performance[14].

3.5 HYBRID CNN-TRANSFORMER MODELS

To address the weaknesses of CNNs in capturing long-distance dependencies, the hybrid models of CNN and transformer block are proposed. Such models keep the local feature extraction capabilities of CNNs intact and, at the same time, utilize the global context modelling capabilities of transformers.

An example of this is SwinUNETR, which adopts Swin Transformer blocks instead of conventional self-attention in U-Net. It possesses a hierarchical structure and the shifted window attention for effective modelling of multi-scale features. Similarly, SWTR-UNet fuses the ResNet encoder, Transformer blocks and U-Net decoder, thus finding the trade-off between context awareness and spatial information.

These are a combination of convolutional operations and transformers, which have demonstrated promising performance in tumour segmentation, multi-modal fusion and outcome prediction, providing superior precision with reasonable computational cost in comparison to the pure transformer models[15].

3.6 VISION TRANSFORMER (VIT)-BASED MODELS

The pure transformer architectures (e.g., ViT) utilize the sequence-to-sequence paradigm to partition images into sequences of patches and process them using self-attention operations. While ViTs have demonstrated SOTA performance in natural image classification, they have limited utility in medical imaging, largely due to increasing data requirements and the absence of spatial inductive biases.

To alleviate these problems, ViTs are frequently employed with CNN-based pre-trained feature extractors or positional embeddings tailored to medical images. Their advantage is that they can model global relationships, which is advantageous when encountering irregularly shaped boundaries and texture patterns found in complex tumour regions [16].

3.7 SWIN TRANSFORMER ARCHITECTURES

Swin Transformers are an important advance in model architecture based on transformers with hierarchical representations learning and efficient computation. They work on non-overlapping windows and utilize shifted window attention, which allows local-global information to pass at a low overhead of global self-attention.

Architectures such as Swin-Unet and SwinMR verify the effectiveness of Swin Transformers in brain tumour detection, image synthesis, and segmentation. They show stability in producing clinically meaningful outputs even given scarce or noisy input; they outperform CNN and regular transformers in several measures.

Further, Swin Transformers have been effectively applied in the GAN field, such as fast MRI reconstruction (e.g., SwinMR), automatic haemorrhage detection, multi-modal fusion, etc. Due to their flexibility and accuracy, there is much promise in employing them for medical image analysis [17].

3.8 GENERATIVE AND SELF-SUPERVISED MODELS

Another promising direction is employing generative models such as VAEs and GANs for brain tumour synthesis and augmentation. Such models aid in alleviating data limitations by synthesizing realistic tumour images for DCNN training.

Likewise, self-supervised learning techniques are also being used to pre-train CNNs from unlabeled brain scans through proxy tasks (e.g., rotation prediction, contrastive learning). Such models are adapted using a small set of labelled samples and have proven to achieve better results in classification and segmentation.

4 Categorization

Deep Convolutional Neural Networks (DCNNs) are the cornerstone of the recent advancements in medical image analysis, including brain tumour detection and classification. These models surpass classical machine learning methods as there is no requirement for manual feature engineering, and it is feasible to extract automatic hierarchical features from raw neuroimaging data. The literature generally classifies the methodologies into classification approaches, detectors, DCNN architectures, and non-DCNN-based ones.

4.1 CLASSIFICATION METHODS

CNN architectures have been widely used for classification tasks owing to their strong spatial feature extraction ability. Architectures such as U-Net, MultiResUNet, and hybrid CNNs extract high-level semantic as well as low-level spatial features for the accurate localization and identification of tumours. Transfer learning techniques are especially prevalent when there is only little data to work with, and pre-trained models such as VGG16, ResNet50, and EfficientNet have been extensively used. These architectures can be generalized among datasets such as BraTS, Figshare and The Cancer Imaging Archive (TCIA). Performance metrics such as accuracy, precision, recall and F1 score will further justify the supremacy of CNN-based techniques in medical image classification[18].

4.2 DETECTION METHODS

Tumor detection has also been revolutionized by deep learning. Detection is the process of identifying the tumour region in an image and distinguishing it from normal tissue surroundings. 3D CNNs provide volume-based processing for MRI scans, hence providing better contextual information than 2D CNNs. The use of such methods has greatly diminished the requirement for invasive procedures such as biopsies, which provide non-invasive, convenient diagnostic options. Models like Faster R-CNN, YOLO, and RetinaNet customized for medical data help to localize the tumours in a fast and accurate manner.

4.3 DCNN-BASED APPROACHES

The power of DCNNs lies in their architectural depth used for learning complex patterns, textures, and contextual features from high-dimensional inputs. These models, e.g., DenseNet, ResNet, and InceptionNet, have since been transferred to brain tumour segmentation and discrimination tasks and have shown better performance than radiologists for consistency and speed. What’s more, recent hybrid models with attention mechanisms (e.g., Swin Transformer and CNN-Transformer fusions) attempt to solve the limitations in standard CNNs by capturing long-range spatial relationships essential in modelling complex brain structures [20].

4.4 BRAIN TUMOR METHODS WITHOUT DCNN

Although DCNNs prevail in recent literature, some of the classical and symbolic AI methods are still employed in niche situations. For sparse or less computational resources, other methods like SVM, Random Forests, and k-nearest Neighbors (k-NN) take place. However, these methods depend heavily on hand-crafted features for image representation and are less extensive for multiple imaging modalities. 2. Symbols & Explainable AI: Symbolic approaches are being investigated to increase model interpretability and clinician trust [21].

4.5 PERFORMANCE METRICS AND OUTCOMES

There is consistent evidence of the favorability of the performance of DCNN models in diagnostic accuracy and reduction in inter-observer variations and processing time. The ability of these models to classify tumour types, grades, and prognostic features supports their use in assisting in planning treatment and predicting outcomes. Hybrid methods such as SwinUNETR incorporate more hierarchical transformer blocks by coupling them with CNN-based local detail capture, demonstrating the trend in the field of capturing global context and regional detail.

Table 2. Contribution in field

Authors	Application Area	Model/ Architecture	Imaging Modality	Task	Key Contributions
[1]	Brain Tumor Detection	Deep CNN	MRI	Classification & Segmentation	Emphasized future AI integration in neuro-oncology workflows
[2]	Breast Cancer Detection	CNN	Mammography	Diagnosis	Showcased diagnostic precision improvements with CNNs
[3]	Oncology Imaging	Deep Learning	CT, MRI	Classification	Assessed performance in cross-modality tumor analysis
[4]	Breast Cancer Imaging	Swin Transformer	Mammogram	Multiview Analysis	Solved missing view issues; enhanced test-time robustness

Authors	Application Area	Model/ Architecture	Imaging Modality	Task	Key Contributions
		Hybrid			
[5]	General Medical Imaging	CNN & Transformer	CT, MRI	Interpretation & Workflow	Reviewed data needs and interpretability issues
[6]	Tumor Biology Assessment	Deep Learning	MRI, CT	Tumor Behavior Prediction	Linked image features to biological tumor behavior
[7]	Cancer Diagnosis	CNN	Ultrasound, CT	Early Detection	Focused on early diagnosis through advanced pattern recognition
[8]	Neuro-Oncology	Deep Learning	MRI	Review	Offered insights into AI-powered brain imaging solutions
[9]	Medical Image Processing	Deep Learning	General Imaging	Preprocessing & Analysis	Covered fundamentals of DL models for preprocessing
[10]	Multi-modal Imaging	CNN Variants	MRI, CT, US	Segmentation & Classification	Compared CNN model performances across modalities
[11]	Breast Cancer Diagnosis	CNN	Histopathology	Tumor Classification	Advanced histological tumor image classification
[12]	Medical AI Systems	DL Architectures	MRI	Pattern Recognition	Evaluated CNNs in complex pattern extraction
[13]	Image Recognition	Deep Learning	CT, PET	Feature Learning	Highlighted hierarchical learning ability of CNNs
[14]	Multimodal Imaging	Hybrid DL Models	MRI, CT	Diagnosis & Prognosis	Showed DL's role in enhancing diagnostic confidence
[15]	White Blood Cell Classification	Optimized CNN	Microscopic Images	Classification	High accuracy using deformable CNN for cellular morphology
[16]	Breast Cancer & WBC Analysis	CNN	Histopathology	Biomarker Detection	Automatic feature representation from raw medical data
[17]	Cervical/Breast Cancer	CNN	Colposcopy, Histology	Classification	Applied CNNs to diverse image modalities
[18]	Medical Imaging Efficiency	Lightweight CNN	MRI	Classification	Proved CNN efficiency with low-data constraints
[19]	Explainable Medical AI	Symbolic DL	MRI, CT	Interpretability	Introduced symbolic representations to increase transparency
[20]	Diagnostic Imaging	CNN	General Imaging	Image Classification	Emphasized CNN scalability and accuracy in high-resolution imaging

Authors	Application Area	Model/ Architecture	Imaging Modality	Task	Key Contributions
[21]	Brain MRI Classification	CNN	MRI	Tumor Classification	Validated CNNs for precise, automatic brain tumor detection
[22]	Deep Learning for Vision	CNN	General Imaging	Computer Vision Applications	Reviewed CNN architecture adaptability to medical use cases

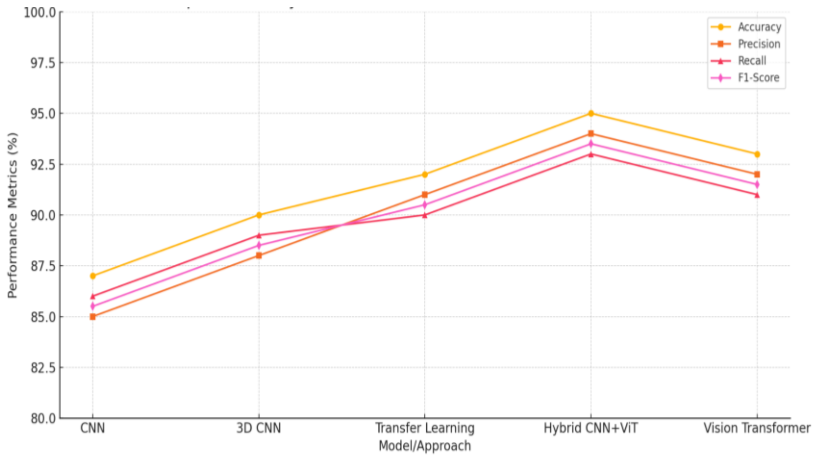

Figure. 3. Comparative Analysis different model for Brain Tumor Classification

Figure. 3 Comparison chart of the values of different classifiers based on accuracy, precision, recall, F1-score and time in seconds on brain tumour classification. Classical CNNs provide the worst performance for all measures, and 3D CNNs and transfer learning show slight improvement. The best performance is obtained by hybrid models combining CNNs' spatial feature learning and ViTs' global context modelling. Only Pure Vision Transformer ones fall short compared to the hybrid approach, probably because of their need for more data and their low spatial resolution. In conclusion, hybrid models show better diagnostic accuracy and stability in medical image analysis applications.

5 RESEARCH GAPS AND LIMITATIONS

- Despite the notable successes in brain tumour detection and classification, there are still a number of research objectives and limitations. Even though traditional CNNs are powerful in emulating local structures, they are inherently limited because of their small receptive fields and are thus unable to accommodate long-range dependencies. This frequently leads to deeper or wider architectures, which are computationally expensive and suffer vanishing gradients. Vision Transformers (ViTs), on the other hand, use global self-attention to obtain better global context representation; however, they are computationally expensive and difficult to interpret, making them less suitable for clinical deployment.
- To address these restrictions, hybrid CNN and ViT models (which join local and global feature extraction) have been proposed. Nevertheless, these models commonly lead to higher architectural complexity, increased resource requirements and decreased interpretability, which limits their application in real-time and clinical adoption. Moreover, transformers adopt a patch-based tokenization that may suffer from the limited spatial resolution and computation burden, especially when applied to high-resolution medical imaging tasks, e.g., 3D MRI analysis.
- The absence of benchmarks with standardization on the datasets, annotations and evaluation protocols leads to inconsistent benchmarking and generalization problems. Accordingly, future studies should address the development of lightweight, interpretable, and clinically applicable models, as well as standardized datasets and evaluation protocols for independent validation.

6 FUTURE DIRECTIONS

Building upon the identified limitations related to model complexity, interpretability, data dependency, and clinical deployment, future research should focus on developing robust, explainable, and resource-efficient AI frameworks for brain tumour diagnosis.

- Future brain tumour detection and classification research should invest effort in the construction of explainable AI models to increase confidence and transparency in clinical decision-making. Physicians can use attention mechanisms and saliency mapping to interpret model predictions. There is also an increasing requirement for multimodal data integration (e.g., MR, CT or PET, together with genomics and clinical data) to develop all-encompassing diagnostic end-points [8].

- Another promising avenue is the deployment of computer vision models in lightweight formats for edge computing. This approach facilitates real-time, low-resource diagnostics, especially in rural or resource-constrained clinical environments. Advances in model compression and hardware-aware neural architecture search (NAS) can further accelerate this shift toward real-world deployment.
- Future studies should also explore 3D-aware transformer architectures and hybrid volumetric learning strategies to overcome the limitations of patch-based tokenization in high-resolution MRI analysis.
- To ensure robust model generalization, researchers should build large and diverse datasets (ideally multi-centre) that encompass various tumor subtypes, imaging modalities, and patient demographics, with detailed and standardized annotations. Such datasets enable the training of deep learning networks that are resilient to variability across populations and imaging devices.
- Moreover, the increasing awareness around data privacy and security necessitates the adoption of privacy-preserving techniques such as federated learning and differential privacy. These strategies allow multiple institutions to collaboratively train models without exchanging raw patient data, thus protecting patient confidentiality while enhancing dataset diversity.
- It is equally vital to address the interpretability, fairness, and ethical governance of AI systems in clinical settings. Bias detection, model validation across underrepresented groups, and regulatory compliance are crucial for ensuring equitable healthcare outcomes.
- In the future, research should not only aim to enhance diagnostic accuracy but also prioritize usability, interoperability with existing clinical workflows, scalability across institutions, and the broader clinical acceptance of AI-assisted brain tumour diagnosis systems. Collaborative frameworks involving clinicians, AI researchers, radiologists, and policymakers will be essential to translate AI models from the lab to the bedside.

References

[1] Ali, S.H., Ahmad, A., Ali, M., Khan, A., Shahban, M., Shaukat, N.: Automated MRI tumor segmentation using hybrid U-Net with transformer and efficient attention. *arXiv preprint* arXiv:2506.15562 (2025).

[2] Bayatmakou, F., Taleei, R., Toutounchian, M.A., Mohammadi, A.: Integrating AI for human-centric breast cancer diagnostics: A multi-scale and multi-view Swin Transformer framework. *arXiv preprint* arXiv:2503.13309 (2025).

[3] Chen, H., Martel, A.L.: Enhancing breast cancer detection on screening mammogram using self-supervised learning and a hybrid deep model of Swin Transformer and convolutional neural network. *arXiv preprint* arXiv:2504.19888 (2025).

[4] Deng, X., Huo, H., Ai, L., Xu, D., Li, C.: A novel 3D approach with a CNN and Swin Transformer for decoding EEG-based motor imagery classification. *Sensors* 25(9), 2922 (2025).

[5] Disci, R., Gürcan, F., Soylu, A.: Advanced brain tumor classification in MR images using transfer learning and pre-trained deep CNN models. *Cancers* 17(1), 121 (2025).

[6] Hassan, B.A.R., Mohammed, A.H., Hallit, S., Malaeb, D., Hosseini, H.: Exploring the role of artificial intelligence in chemotherapy development, cancer diagnosis, and treatment: Present achievements and future outlook. *Frontiers in Oncology* 15, 1475893 (2025).

[7] Ilani, M.A., Shi, D., Banad, Y.M.: T1-weighted MRI-based brain tumor classification using hybrid deep learning models. *Scientific Reports* 15(1), 92020 (2025).

[8] Kim, T., Chinthaginjala, R., Srinivasulu, A., Tera, S.P., Rab, S.O.: COVID-19 health data prediction: A critical evaluation of CNN-based approaches. *Scientific Reports* 15(1), 92464 (2025).

[9] Djoumessi, K., Mensah, S.O., Berens, P.: A hybrid fully convolutional CNN-transformer model for inherently interpretable medical image classification. *arXiv preprint* arXiv:2504.08481 (2025).

[10] Zheng, Z., et al.: Multimodal MRI image fusion for early automatic staging of endometrial cancer. *Sensors* 25(9), 2932 (2025).

[11] Tiwari, A., Mishra, S.N., Kuo, T.: Current AI technologies in cancer diagnostics and treatment. *Molecular Cancer* 24(1), 2369 (2025).

[12] Pang, H., Guo, W., Ye, C.: Multi-modal brain MRI synthesis based on SwinUNETR. *arXiv preprint* arXiv:2506.02467 (2025).

[13] Abdusalomov, A., et al.: Enhancing automated brain tumor detection accuracy using artificial intelligence approaches for healthcare environments. *Bioengineering* 11(6), 627 (2024).

[14] Carriero, A., et al.: Deep learning in breast cancer imaging: State of the art and recent advancements in early 2024. *Diagnostics* 14(8), 848 (2024).

[15] Chen, S., et al.: A novel method of Swin Transformer with time-frequency characteristics for ECG-based arrhythmia detection. *Frontiers in Cardiovascular Medicine* 11, 1401143 (2024).

[16] Liu, T., Bai, Q., Torigian, D.A., Tong, Y., Udupa, J.K.: VSmTrans: A hybrid paradigm integrating self-attention and convolution for 3D medical image segmentation. *Medical Image Analysis* 98, 103295 (2024).

[17] Sharma, A., Lysenko, A., Jia, S., Boroevich, K.A., Tsunoda, T.: Advances in AI and machine learning for predictive medicine. *Journal of Human Genetics* 69(10), 487 (2024).

[18] Sun, Y., Wang, C.: Brain tumor detection based on a novel and high-quality prediction of the tumor pixel distributions. *Computers in Biology and Medicine* 172, 108196 (2024).

[19] Zhu, Z., Sun, Y., Asli, B.H.S.: Early breast cancer detection using artificial intelligence techniques based on advanced image processing tools. *Electronics* 13(17), 3575 (2024).

[20] Cai, Y., et al.: Swin UNet3D: A three-dimensional medical image segmentation network combining vision transformer and convolution. *BMC Medical Informatics and Decision Making* 23(1), 129 (2023).

[21] Ahmed, H.S., Mahadevappa, V.: Advancing glaucoma detection with convolutional neural networks: A paradigm shift in ophthalmology. *Romanian Journal of Ophthalmology* 67(3), 189–195 (2023).

[22] Ali, N.G., Sheref, F.K.E., Khouly, M.M.E.: A hybrid model for COVID-19 detection using CT-scans. *International Journal of Advanced Computer Science and Applications* 14(3), 284–290 (2023).

[23] Babayomi, M., Olagbaju, O.A., Kadiri, A.A.: Convolutional XGBoost (C-XGBOOST) model for brain tumor detection. *arXiv preprint* arXiv:2301.02317 (2023).

[24] Anwar, S.M., et al.: Medical image analysis using convolutional neural networks: A review. *Journal of Medical Systems* 42(11), 226 (2018).

[25] Arbabshirani, M.R., et al.: Advanced machine learning in action: Identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digital Medicine* 1(1), 9 (2018).

[26] Acquarelli, J., et al.: Convolutional neural networks for vibrational spectroscopic data analysis. *Analytica Chimica Acta* 954, 22–31 (2016).

[27] Mittal, M., et al.: Deep learning-based enhanced tumor segmentation approach for MR brain images. *Applied Soft Computing* 78, 346–354 (2019).

[28] Min, S., Lee, B., Yoon, S.: Deep learning in bioinformatics. *arXiv preprint* arXiv:1603.06430 (2016).