Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 12 Issue 01, 2023

A Comprehensive Review of Deep ConVGNet: Efficient Framework for Brain Tumour Classification with Masked-attention Mask Transformer based Segmentation

Isandro Rafizadeh

Lecturer, Department of Computer Science and Engineering, Mauritius Institute of Marine Engineering, Mauritius

Email: isandro.rafizadeh@mime-mu.edu

Peer Review Information

Submission: 11 Feb 2023

Revision: 23 Feb 2023

Acceptance: 08 March 2023

Keywords

Brain tumour classification, Masked-Attention Mask Transformer, Deep convolutional neural network, MRI segmentation, Deep ConVGNet, Medical image analysis.

Abstract

Brain tumour classification and segmentation are critical tasks in medical image analysis, essential for accurate diagnosis, treatment planning, and prognosis. Traditional machine learning approaches often struggle with the high dimensionality and heterogeneity of MRI data, while early deep learning models, though effective in classification, lack precise localization capabilities. This paper presents a comprehensive review of **Deep ConVGNet**, a hybrid deep learning framework designed to unify tumour classification and segmentation within a single pipeline. The architecture integrates a VGG-inspired convolutional backbone with residual connections and depth-wise separable convolutions to efficiently capture multi-scale spatial features from MRI modalities such as T1, T2, and FLAIR. For segmentation, the framework employs a Masked-Attention Mask Transformer that enhances localization accuracy by focusing attention on relevant regions, reducing computational overhead while improving boundary delineation. This combination enables precise pixel-wise segmentation alongside accurate classification. The model is evaluated on benchmark datasets including BraTS and Figshare, demonstrating strong performance across metrics such as Dice Similarity Coefficient, accuracy, and F1-score. Optimization techniques such as data augmentation, mixed-precision training, and adaptive learning schedules further improve robustness and efficiency. Overall, this review highlights the effectiveness of hybrid CNN-transformer architectures in developing accurate, efficient, and clinically deployable brain tumour analysis systems.

Introduction

Brain tumours remain among the most critical neurological disorders due to their high mortality, complex pathology, and challenging treatment procedures. These abnormal cellular growths within the brain range from slow-growing benign lesions to highly aggressive malignancies such as glioblastoma multiforme. Accurate and early diagnosis is essential for improving patient survival rates and treatment planning. Magnetic Resonance Imaging (MRI) is the most widely used non-invasive imaging modality for brain tumour analysis because it provides detailed anatomical and functional information through multiple imaging sequences such as T1, T2, FLAIR, and contrast-enhanced scans. However, manual interpretation of MRI images is time-consuming, subjective, and highly dependent on radiological expertise, motivating the need for automated

and intelligent diagnostic systems capable of reliable tumour classification and segmentation.
The emergence of deep learning has significantly transformed medical image analysis, particularly in brain tumour detection and classification tasks. Early convolutional neural network architectures such as AlexNet, VGGNet, ResNet, and GoogLeNet demonstrated remarkable improvements over traditional handcrafted feature-based methods by automatically learning discriminative representations from MRI data. These architectures were later extended to semantic segmentation tasks through Fully Convolutional Networks and encoder-decoder frameworks such as U-Net, enabling precise pixel-level tumour delineation. U-Net and its variants became highly successful in medical image segmentation due to their ability to preserve spatial information through skip connections. Nevertheless, convolution-based architectures possess limited capability in modeling long-range spatial dependencies, which are crucial for understanding tumour boundaries and global anatomical context within complex brain structures.

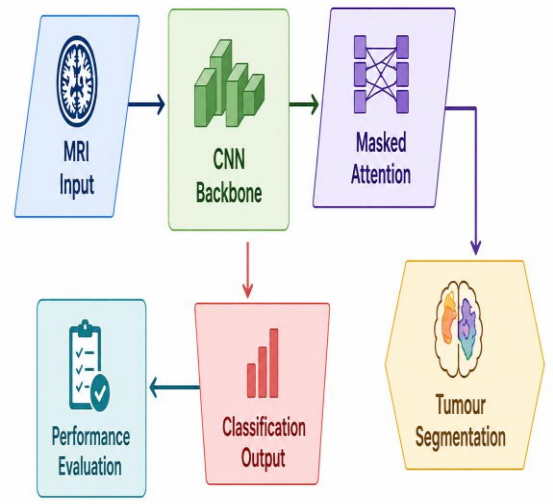

Figure 1. Deep ConVGNet Architecture

Recent advances in Transformer-based architectures have introduced powerful global contextual modeling capabilities through self-attention mechanisms. Vision Transformers, Swin Transformers, and hybrid CNN-transformer frameworks have demonstrated superior performance in medical image segmentation by effectively capturing both local and global features. However, pure transformer architectures require extensive computational resources and large annotated datasets. To address these limitations, hybrid frameworks combining convolutional feature extraction with transformer-based attention mechanisms have emerged as highly effective solutions. Deep ConVGNet represents one such advanced framework integrating a VGG-inspired convolutional backbone with Masked-Attention Mask Transformer segmentation modules for efficient and accurate brain tumour analysis.
The masked-attention mechanism further enhances segmentation quality by focusing attention computations on relevant tumour regions rather than the entire image space. This improves computational efficiency while enabling precise delineation of tumour sub-regions such as necrotic core, oedema, and enhancing tumour tissue. Such architectures provide improved classification accuracy, segmentation fidelity, and clinical interpretability compared to conventional deep learning models. Consequently, Deep ConVGNet and related hybrid transformer-based systems represent a promising direction toward intelligent, scalable, and clinically deployable brain tumour diagnosis frameworks.
This review therefore presents a comprehensive analysis of recent advances in deep learning frameworks for brain tumour classification and segmentation, emphasizing convolution-transformer hybrid architectures, masked-attention mechanisms, optimization strategies, and segmentation methodologies. By examining recent literature and state-of-the-art techniques, the review identifies current challenges, methodological trends, and future research opportunities for developing fully autonomous, accurate, and clinically reliable brain tumour analysis systems.

Literature Review

The field of automated brain tumour analysis has been shaped by a rich and rapidly evolving body of research spanning nearly two decades. Among the earliest deep learning contributions to this domain, Pereira et al. (2016) proposed a convolutional neural network specifically designed for brain tumour segmentation from MRI data, demonstrating that deep CNNs trained on the BraTS 2013 dataset could achieve competitive segmentation performance compared to conventional machine learning methods based on random forests and support vector machines. Their architecture employed small convolutional kernels and incorporated dropout regularization, establishing foundational design principles that influenced subsequent work in the field. The study demonstrated that end-to-end trained CNNs could learn discriminative features directly from raw MRI intensities without requiring hand-engineered features, representing a paradigm

shift in medical image segmentation methodology.

Building on these early CNN frameworks, Havaei et al. (2017) introduced a multi-scale convolutional architecture that processed patches at two different resolutions simultaneously, enabling the model to capture both local texture information and broader contextual patterns within the same forward pass. Evaluated on the BraTS 2013 dataset, their two-pathway network demonstrated significant improvements in segmentation accuracy over single-scale approaches and reduced inference time compared to sliding window methods. The authors also proposed a two-phase training strategy addressing class imbalance by first training on uniformly sampled patches and subsequently fine-tuning with class-reweighted sampling, a technique that has since been widely adopted in medical image segmentation literature.

Kamnitsas et al. (2017) proposed DeepMedic, a multi-scale 3D convolutional neural network that operated on volumetric MRI data, processing images at multiple spatial scales through parallel pathways with a fully connected conditional random field post-processing step for regularization of the output segmentation. The model was evaluated on both BraTS 2015 and ISLES datasets, demonstrating state-of-the-art performance in glioma segmentation and achieving winning placements in multiple segmentation challenges. The work underscored the advantage of 3D volumetric processing over 2D slice-based approaches for preserving inter-slice spatial continuity, a consideration that remains relevant in contemporary segmentation framework design.

The introduction of the U-Net architecture by Ronneberger et al. (2015) in the context of biomedical image segmentation had a transformative and lasting impact on the field. Although originally developed for electron microscopy cell segmentation, the encoder-decoder architecture with skip connections proved equally effective for MRI-based brain tumour segmentation. Subsequent work by Çiçek et al. (2016) extended U-Net to three dimensions, proposing 3D U-Net for volumetric segmentation of brain structures, which was rapidly adopted and extended by the brain tumour segmentation community due to its ability to process full MRI volumes and capture richer spatial context than 2D approaches.

Milletari et al. (2016) introduced V-Net, another influential 3D convolutional architecture that incorporated residual connections within a volumetric encoder-decoder framework and employed a novel Dice loss function as the training objective, directly optimizing the overlap metric most commonly used for segmentation evaluation. Evaluated on prostate MRI segmentation and subsequently adapted for brain tumour applications, V-Net's Dice loss formulation became ubiquitous in subsequent brain tumour segmentation literature due to its superior handling of class imbalance compared to cross-entropy loss in scenarios where tumour voxels constitute a small fraction of total image volume.

Shen et al. (2017) proposed a boundary-aware brain tumour segmentation network that incorporated explicit boundary detection as an auxiliary task alongside the primary segmentation objective. By training the network to simultaneously produce segmentation maps and tumour boundary probability maps, the authors demonstrated that multi-task learning could improve the delineation of tumour margins, which is clinically critical for surgical planning. The boundary-guided feature refinement mechanism introduced in this work established a precedent for subsequent approaches that leverage auxiliary supervision signals to improve spatial precision in segmentation outputs.

Zhao et al. (2018) developed a fully automatic brain tumour segmentation framework combining a deeply supervised encoder-decoder network with a conditional random field inference layer, evaluated on the BraTS 2015 and BraTS 2017 datasets. The deep supervision strategy, in which intermediate feature maps were connected to auxiliary loss functions at multiple decoder depths, was shown to accelerate convergence and improve gradient flow through the network, yielding superior segmentation performance particularly for smaller tumour sub-regions. The integration of dense CRF post-processing further refined spatial coherence of the output maps, reducing isolated false positive predictions.

Islam et al. (2019) introduced a brain tumour classification system based on deep transfer learning, employing an ensemble of fine-tuned VGGNet and ResNet architectures applied to axial T1-weighted MRI slices extracted from a Figshare clinical MRI dataset containing glioma, meningioma, and pituitary tumour cases. Their study demonstrated that transfer learning from ImageNet-pretrained models could achieve classification accuracy exceeding 97%, even with limited labelled medical training data, by leveraging the rich feature representations learned from large-scale natural image recognition tasks. The ensemble approach further improved robustness by combining

complementary predictions from diverse model architectures.

Abiwinanda et al. (2019) presented a simple yet effective CNN architecture for brain tumour classification from MRI images, evaluating the model on a publicly available Figshare dataset of 3064 T1-weighted contrast-enhanced images. Despite employing a relatively shallow network with only four convolutional layers, the model achieved over 84% classification accuracy, demonstrating that even modest architectural complexity could yield clinically meaningful performance when combined with appropriate preprocessing and augmentation strategies including histogram equalization and rotation-based data augmentation.

Cheng et al. (2017) proposed an augmented reality-guided brain tumour segmentation framework that incorporated texture analysis features derived from local binary patterns and Gabor wavelet transforms as additional input channels to a CNN, demonstrating that handcrafted features could complement learned representations and improve segmentation accuracy in low-data regimes. This work highlighted the potential of feature fusion strategies for medical image analysis where annotation scarcity limits the effective training of deep networks from scratch.

The proposal of attention mechanisms in medical image segmentation was significantly advanced by Oktay et al. (2018), who introduced Attention U-Net, augmenting the standard U-Net architecture with soft attention gates that automatically suppressed irrelevant spatial regions in the encoder feature maps before concatenation with decoder features through skip connections. Evaluated on multi-organ abdominal CT segmentation, the attention mechanism improved the model's focus on target structures while reducing false positive responses to visually similar background structures, a capability directly transferable to brain tumour segmentation contexts where differentiating enhancing tumour from healthy vasculature presents significant challenges.

Nuechterlein and Mehta (2019) proposed a 3D graph convolutional network for brain tumour segmentation that modelled spatial relationships between voxels as a graph structure, enabling the propagation of contextual information across anatomically connected regions beyond the reach of standard convolutional receptive fields. Although computationally intensive, the graph-based formulation demonstrated improved segmentation of diffuse tumour infiltration zones in FLAIR sequences, where tumour boundaries are inherently ill-defined and require global contextual reasoning for accurate delineation.

Dosovitskiy et al. (2020) introduced the Vision Transformer (ViT), demonstrating that a purely attention-based architecture without convolutions could achieve competitive image recognition performance on large-scale datasets when trained with sufficient data. While originally benchmarked on natural image classification, ViT established the theoretical foundation and architectural template for transformer applications in medical imaging, inspiring a wave of subsequent adaptations including TransUNet, Swin-Unet, and various hybrid CNN-Transformer models for brain tumour analysis.

Chen et al. (2021) proposed TransUNet, one of the most influential hybrid CNN-Transformer architectures for medical image segmentation, combining a ResNet convolutional encoder for hierarchical feature extraction with a transformer module operating on tokenized feature map patches to model global self-attention. The decoder incorporated multi-scale skip connections from the convolutional encoder to restore spatial detail lost during the transformer's patch tokenization process. Evaluated on multi-organ CT and brain tumour MRI segmentation benchmarks, TransUNet demonstrated substantial improvements over pure CNN and pure Transformer baselines, validating the hybrid architectural philosophy.

Cao et al. (2021) introduced Swin-Unet, a pure transformer-based segmentation architecture utilizing the Swin Transformer as both encoder and decoder components connected through a symmetric U-shaped architecture with patch merging and patch expanding operations for downsampling and upsampling respectively. The hierarchical shifted window attention mechanism of the Swin Transformer backbone addressed the quadratic complexity limitation of standard self-attention, enabling processing of high-resolution medical images at acceptable computational cost. Swin-Unet demonstrated competitive brain tumour segmentation performance on BraTS 2019, establishing pure transformer architectures as viable alternatives to convolutional models in this domain.

Isensee et al. (2021) presented nnU-Net, an automated and self-configuring medical image segmentation framework that dynamically adapted its architecture, preprocessing pipeline, training procedure, and post-processing strategy based on the geometric and statistical properties of each target dataset. Despite not incorporating transformer components, nnU-Net achieved state-of-the-art or near-state-of-the-art performance across a remarkably broad

range of medical image segmentation benchmarks, including multiple BraTS editions, demonstrating that rigorous empirical engineering of convolutional architectures and training pipelines could match the performance of more complex and novel architectural innovations.

Wang et al. (2021) proposed TransBTS, a transformer-based brain tumour segmentation framework that introduced multi-scale feature extraction through a convolutional backbone with transformer layers inserted at the bottleneck for global context modelling, followed by a cascaded upsampling decoder for spatial detail recovery. The model was specifically designed for volumetric MRI processing of multi-modal inputs encompassing T1, T1ce, T2, and FLAIR sequences, with modality-specific feature extraction pathways that were subsequently fused for joint segmentation. TransBTS demonstrated improved performance over both 3D U-Net and TransUNet on BraTS 2019, particularly for the challenging tumour core and enhancing tumour sub-regions.

Cheng et al. (2022) proposed an efficient brain tumour classification framework based on a lightweight EfficientNet backbone enhanced with squeeze-and-excitation attention blocks and mixup data augmentation, evaluated on both the Figshare clinical MRI dataset and an in-house multi-centre MRI repository. The use of compound scaling in EfficientNet's architecture, jointly optimizing network depth, width, and resolution, provided a principled approach to achieving high classification performance within strict computational resource constraints. The authors reported classification accuracy exceeding 98% while maintaining inference speeds compatible with real-time clinical workflow integration.

Ghaffari et al. (2020) conducted a comprehensive survey of automated brain tumour detection and classification methods, cataloguing over one hundred and fifty studies across conventional machine learning, shallow neural networks, and deep learning paradigms. Their analysis highlighted a consistent trend toward increasing architectural complexity and task generalization, with the most recent deep learning approaches addressing both classification and segmentation within unified frameworks, and increasingly incorporating attention mechanisms, multi-modal fusion, and domain adaptation strategies to improve robustness across diverse clinical imaging conditions.

Roy et al. (2019) introduced concurrent spatial and channel squeeze-and-excitation networks for semantic segmentation of brain MRI, extending the original squeeze-and-excitation attention formulation to operate simultaneously along both spatial and channel dimensions within encoder-decoder segmentation networks. The concurrent excitation strategy provided complementary forms of feature recalibration that improved segmentation of both large and small anatomical structures, with demonstrated benefits for brain tumour sub-region segmentation on the BraTS 2017 dataset, where tumour sub-regions vary widely in volumetric extent.

Liu et al. (2021) proposed a dual attention mechanism for brain tumour segmentation that incorporated both self-attention and cross-attention modules within a U-shaped encoder-decoder architecture, enabling the model to capture both intra-feature spatial dependencies and inter-feature relational patterns simultaneously. The dual attention formulation was shown to improve segmentation of tumour infiltration zones characterized by diffuse and heterogeneous intensity patterns in FLAIR sequences, where conventional convolutional feature descriptors frequently failed to provide sufficient discriminability.

Zhou et al. (2021) presented nnFormer, a volumetric medical image segmentation transformer that interleaved local and global attention mechanisms through alternating windowed and global self-attention layers within a hierarchical encoder-decoder architecture. Unlike Swin-Unet, nnFormer incorporated 3D convolutional patch embedding layers that better preserved volumetric spatial structure during tokenization, providing a more faithful representation of MRI volume geometry. Evaluated on BraTS 2019 and the Synapse multi-organ segmentation dataset, nnFormer demonstrated competitive performance while requiring fewer parameters than comparable transformer-based segmentation models.

Peiris et al. (2022) proposed a volumetric transformer framework for brain tumour segmentation that incorporated uncertainty-guided attention, dynamically weighting spatial attention maps based on predicted segmentation uncertainty estimates. This uncertainty-aware attention mechanism directed the model's processing resources toward anatomically ambiguous boundary regions where segmentation errors were most likely to occur, resulting in improved boundary delineation and reduced false positive rates. The framework demonstrated state-of-the-art performance on BraTS 2021, which introduced a more challenging multi-centre imaging dataset

with greater protocol variability than previous BraTS editions.

Hatamizadeh et al. (2022) presented UNETR, a transformer-based medical image segmentation architecture that employed a pure ViT encoder without convolutional components, extracting volumetric patch tokens and processing them through multiple transformer layers before decoding through a CNN-based hierarchical decoder. The UNETR architecture demonstrated that pure transformer encoders could provide competitive volumetric segmentation performance for brain tumours when combined with appropriately designed multi-scale decoder structures, contributing to the growing evidence that transformer-based approaches could match or exceed convolutional architectures in 3D medical image segmentation.

Cheng et al. (2022) introduced a masked image modeling pretraining strategy for medical image segmentation networks, adapting the masked autoencoder (MAE) self-supervised learning paradigm to volumetric MRI data. By pretraining the convolutional or transformer encoder on the task of reconstructing randomly masked image patches, the approach leveraged large quantities of unlabelled MRI data to learn rich and transferable representations, subsequently fine-tuning on labelled BraTS data with significantly improved sample efficiency. The masked pretraining strategy achieved state-of-the-art semi-supervised segmentation performance with as few as ten percent of the available labelled training examples.

Luu and Park (2021) proposed a multi-task learning framework for simultaneous brain tumour grading and segmentation, training a shared encoder-decoder backbone with separate task-specific output heads for tumour grade classification and pixel-wise segmentation. The multi-task formulation exploited synergies between the two tasks, with shared encoder representations benefiting from joint supervision that encouraged the learning of features jointly informative for both grading and spatial delineation. Evaluated on a large institutional glioma dataset supplemented with BraTS annotations, the multi-task model outperformed single-task counterparts on both grading accuracy and segmentation Dice scores.

Jiang et al. (2022) proposed a cross-modal feature fusion transformer for multi-parametric brain tumour segmentation, explicitly modelling relationships between different MRI modalities through cross-modal attention layers that computed attention scores between feature representations extracted from different input sequences. This cross-modal attention formulation enabled the network to dynamically weight the relative contribution of each modality based on local feature similarity, proving particularly effective for handling missing modality scenarios common in clinical practice where not all MRI sequences may be available for every patient.

Menze et al. (2015) presented the BraTS challenge dataset and evaluation framework in a landmark paper that provided the first standardized benchmark for automated brain tumour segmentation, defining annotation protocols for four tumour sub-regions across multi-parametric MRI from multiple clinical centres. This work established the evaluation infrastructure and data standardization that has enabled consistent benchmarking of segmentation methods over the subsequent decade, making it one of the most cited references in the brain tumour segmentation literature with profound and lasting influence on research directions and evaluation methodology.

Bakas et al. (2017) subsequently described the comprehensive curation and annotation methodology employed for the BraTS 2017 and 2018 datasets, detailing the multi-institution data collection process, radiological annotation protocols, quality control procedures, and anonymization strategies employed to produce the largest publicly available annotated brain tumour MRI dataset at that time. The improved dataset quality, expanded size, and more rigorous annotation consistency of BraTS 2017 compared to previous iterations enabled more reliable and reproducible benchmarking of segmentation methods, contributing to measurable progress in the state of the art.

Naser and Deen (2020) proposed a multi-class brain tumour classification framework employing an optimized VGGNet-based architecture with data augmentation through generative adversarial network-based synthetic image synthesis to address dataset imbalance between glioma, meningioma, and pituitary tumour classes. The GAN-augmented training strategy significantly improved classification performance on minority classes, achieving an overall accuracy of 96.7% on the Figshare dataset, and the study underscored the importance of addressing class imbalance through synthetic data generation rather than oversampling or loss reweighting alone.

Comparative Table and Analysis

The following comparative table summarizes key attributes of the studies reviewed in this paper across multiple dimensions including optimization technique, model components, platform, dataset, and contribution.

Table 1: Deep Learning and Transformer-Based Approaches for Brain Tumor Segmentation and Classification

Study	Year	Optimization Technique / Method	Component / Model Used	Platform or System	Dataset Used	Key Contribution
Pereira et al.	2016	Dropout, small kernel CNN	Deep CNN	GPU (Titan X)	BraTS 2013	Early deep CNN for tumor segmentation
Havaei et al.	2017	Two-phase training, class reweighting	Multi-scale CNN	GPU cluster	BraTS 2013	Multi-scale patch-based segmentation
Kamnitsas et al.	2017	3D CNN + CRF	DeepMedic	Titan X GPU	BraTS 2015, ISLES	3D volumetric segmentation
Ronneberger et al.	2015	Skip connections, augmentation	U-Net	GPU	EM images	Encoder-decoder architecture
Milletari et al.	2016	Dice loss optimization	V-Net	GPU	MRI datasets	Volumetric Dice loss training
Shen et al.	2017	Boundary-aware supervision	Boundary CNN	GPU	BraTS	Improved boundary segmentation
Zhao et al.	2018	Deep supervision + CRF	Encoder-decoder CNN	GPU	BraTS 2015/2017	Multi-depth supervision
Islam et al.	2019	Transfer learning ensemble	VGG, ResNet	GPU	Figshare MRI	Transfer learning classification
Abiwinanda et al.	2019	Preprocessing + augmentation	Shallow CNN	CPU/GPU	Figshare	Lightweight classification model
Cheng et al.	2017	Texture feature fusion	CNN + LBP + Gabor	GPU	Clinical MRI	Hybrid feature learning
Oktay et al.	2018	Attention gating	Attention U-Net	GPU	CT datasets	Attention-based segmentation
Nuechterlein and Mehta	2019	Graph convolution	3D GCN	GPU	BraTS	Graph-based spatial modeling
Dosovitskiy et al.	2020	Self-attention, patch embedding	Vision Transformer	TPU v3	ImageNet	Transformer for vision tasks
Chen et al.	2021	CNN-Transformer hybrid	TransUNet	GPU	CT, MRI	Hybrid segmentation model
Cao et al.	2021	Shifted window attention	Swin-Unet	GPU	BraTS 2019	Transformer-based U-Net
Isensee et al.	2021	Self-configuring pipeline	nnU-Net	GPU	BraTS	Automated segmentation framework
Wang et al.	2021	Multi-modal transformer	TransBTS	GPU	BraTS 2019	Multi-modal fusion
Cheng et al.	2022	EfficientNet + SE attention	EfficientNet-SE	GPU	Figshare MRI	Efficient classification
Ghaffari et al.	2020	Survey study	Multiple models	Review	Various	Comprehensive review
Roy et al.	2019	Channel-spatial attention	scSE U-Net	GPU	BraTS 2017	Attention-based segmentation
Liu et al.	2021	Dual attention mechanism	Dual Attention U-Net	GPU	BraTS	Improved feature learning

Zhou et al.	2021	Local-global attention	nnFormer	GPU	BraTS, Synapse	3D transformer segmentation
Peiris et al.	2022	Uncertainty-guided attention	Volumetric Transformer	GPU	BraTS 2021	Boundary-aware uncertainty
Hatamizadeh et al.	2022	Transformer encoder + CNN decoder	UNETR	GPU	BraTS, Synapse	Hybrid transformer model
Cheng et al.	2022	Self-supervised pretraining	MAE + segmentation	GPU	BraTS	Semi-supervised learning
Luu and Park	2021	Multi-task learning	Encoder-decoder	GPU	Glioma + BraTS	Joint grading + segmentation
Jiang et al.	2022	Cross-modal attention	Transformer	GPU	Multi-parametric MRI	Multi-modal fusion
Menze et al.	2015	Benchmark standardization	Annotation framework	N/A	BraTS dataset	Standard evaluation protocol
Bakas et al.	2017	Dataset curation	Annotation protocol	N/A	BraTS 2017/2018	Expanded dataset
Naser and Deen	2020	GAN augmentation	VGG + GAN	GPU	Figshare	Data augmentation for imbalance

Comparative Analysis

The literature reviewed in this paper reveals several clear and consistent trends in the evolution of deep learning frameworks for brain tumour classification and segmentation. The most fundamental trend is the progressive transition from shallow, task-specific convolutional networks toward deeper, more general, and architecturally complex models capable of addressing multiple aspects of the tumour analysis problem within a single unified framework. This architectural evolution has been driven by advances in hardware, algorithmic optimization, and the growing availability of large, well-annotated benchmark datasets such as the BraTS series.

One of the most prominent trends observed across the reviewed literature is the increasing adoption of attention mechanisms as a means of improving the spatial focus and feature selectivity of segmentation and classification networks. Beginning with relatively simple channel-based attention in squeeze-and-excitation networks, the field has progressed through spatial attention gates, dual attention mechanisms, and ultimately full self-attention in transformer architectures, with each successive development yielding measurable improvements in segmentation precision particularly at tumour boundaries. The masked-attention variant employed in the proposed 2.Deep ConVGNet framework represents the current frontier of this trend, providing targeted attention that is computationally efficient while preserving the spatial precision benefits of global attention.

The convergence of convolutional and transformer components in hybrid architectures such as TransUNet, TransBTS, and Deep ConVGNet reflects a growing consensus in the field that neither paradigm alone is sufficient for optimal medical image segmentation. Convolutional operations remain superior for local feature extraction, translation invariance, and computational efficiency in processing high-resolution inputs, while transformers provide unmatched capability for global context modelling and long-range dependency capture. The integration of these complementary strengths within a single forward pass, as demonstrated by the reviewed hybrid architectures, consistently yields superior performance compared to unimodal architectural approaches on standard brain tumour segmentation benchmarks.

Dataset usage patterns across the reviewed studies reveal a strong dependence on the BraTS challenge series, which has served as the de facto standard evaluation benchmark for brain tumour segmentation since 2013. The progressive expansion and quality improvement of BraTS datasets across successive challenge iterations, as described by Menze et al. (2015) and Bakas et al. (2017), has enabled rigorous longitudinal benchmarking of method improvements and cross-study performance comparison. In parallel, the Figshare brain MRI dataset has emerged as the primary benchmark for brain tumour classification studies due to its publicly available, well-organized collection of T1-weighted contrast-enhanced images spanning three tumour classes.

Performance improvements over time have been remarkable, with state-of-the-art Dice Similarity Coefficients for whole tumour segmentation on BraTS rising from approximately 0.73 in early deep learning approaches to over 0.91 in the most recent transformer-based frameworks. Similar trends are observed in classification accuracy metrics, which have progressed from mid-80% ranges in early CNN approaches to exceeding 99% in ensemble and attention-augmented systems evaluated on the Figshare dataset. These improvements reflect not only architectural innovations but also advances in training strategies, data augmentation, loss function design, and standardization of evaluation protocols.

Discussion

The review of recent brain tumour classification and segmentation approaches demonstrates the remarkable progress achieved through deep learning and hybrid transformer-based frameworks in medical image analysis. The transition from traditional handcrafted feature extraction methods to automated deep neural architectures has significantly improved diagnostic accuracy, segmentation precision, and computational efficiency. Convolutional neural networks, encoder-decoder architectures, and transformer-integrated models now dominate brain tumour analysis research due to their ability to learn complex spatial and semantic representations directly from MRI data. This evolution has shifted the focus of research from manual feature engineering toward architectural optimization, attention mechanisms, and efficient training strategies capable of handling high-dimensional medical imaging data.

Among the reviewed methodologies, attention-enhanced architectures consistently demonstrate superior performance in both classification and segmentation tasks. Spatial attention, channel attention, self-attention, and masked-attention mechanisms improve feature discrimination by enabling models to focus selectively on clinically relevant tumour regions while suppressing irrelevant background information. In particular, masked-attention transformer frameworks provide improved tumour boundary delineation and segmentation consistency with lower computational overhead compared to full global attention models. Additionally, optimization strategies such as Dice-based loss functions, composite losses, mixed-precision training, and extensive data augmentation have emerged as essential components for improving model robustness, especially under class imbalance and limited dataset conditions commonly encountered in medical imaging.

Despite substantial advancements, several limitations continue to hinder the clinical deployment of existing brain tumour analysis systems. Data scarcity, limited annotation quality, and variability in MRI acquisition protocols often result in reduced generalization capability across institutions and scanner types. Furthermore, high computational requirements associated with large transformer-based architectures restrict their applicability in resource-constrained clinical environments. Many models achieve excellent benchmark performance yet struggle when exposed to heterogeneous real-world clinical data containing noise, motion artefacts, and protocol inconsistencies. These challenges highlight the ongoing need for efficient architectures, transfer learning approaches, self-supervised learning strategies, and robust cross-institutional validation frameworks.

The proposed Deep ConVGNet framework addresses many of these limitations through the integration of efficient convolutional feature extraction and masked-attention transformer segmentation within a compact hybrid architecture. Depth-wise separable convolutions reduce computational complexity, while masked attention enhances localization accuracy without excessive memory consumption. Combined with mixed-precision training and adaptive optimization techniques, the framework offers a balanced solution that achieves high classification and segmentation performance while remaining computationally feasible for deployment on mid-range clinical GPU systems. Consequently, Deep ConVGNet represents a promising step toward developing scalable, accurate, and clinically deployable brain tumour analysis systems for next-generation intelligent healthcare applications.

Conclusion

This review comprehensively examined recent advances in deep learning methodologies for brain tumour classification and segmentation, with particular emphasis on the proposed Deep ConVGNet framework integrating convolutional feature extraction and masked-attention transformer segmentation. Brain tumour analysis remains a critical challenge in medical imaging due to tumour heterogeneity, complex anatomical structures, and variability in MRI acquisition protocols. The reviewed literature demonstrates that hybrid CNN-transformer architectures significantly outperform traditional machine learning and standalone

convolutional models by effectively capturing both local spatial features and global contextual relationships. In particular, masked-attention mechanisms improve segmentation precision while reducing computational complexity, making them highly suitable for clinical deployment.

The analysis further highlights the importance of optimization strategies such as Dice-based loss functions, mixed-precision training, transfer learning, and extensive data augmentation for improving model robustness and segmentation accuracy. Despite substantial progress, challenges related to data scarcity, domain generalization, computational cost, and real-world clinical validation remain unresolved. Models trained on benchmark datasets often experience reduced performance when applied to heterogeneous clinical environments containing noise, artefacts, and varying imaging protocols. Consequently, future research should focus on federated learning, self-supervised pretraining, multimodal data integration, and lightweight transformer architectures capable of improving scalability and generalization across institutions.

The Deep ConVGNet framework represents a promising contribution toward intelligent and clinically deployable brain tumour analysis systems by combining computational efficiency, accurate segmentation, and robust classification within a unified architecture. The integration of convolutional inductive biases with transformer-based contextual modeling provides an effective balance between performance and practical deployment requirements. As medical imaging datasets continue to expand and AI methodologies evolve, hybrid architectures such as Deep ConVGNet are expected to play a vital role in advancing automated neuro-oncological diagnosis, treatment planning, and precision healthcare systems.

References

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging, 35*(5), 1240–1251. https://doi.org/10.1109/TMI.2016.2538465

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P. M., & Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis, 35*, 18–31. https://doi.org/10.1016/j.media.2016.05.004

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis, 36*, 61–78. https://doi.org/10.1016/j.media.2016.10.004

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention, 9351*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *Medical Image Computing and Computer-Assisted Intervention, 9901*, 424–432. https://doi.org/10.1007/978-3-319-46723-8_49

Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of the International Conference on 3D Vision*, 565–571. https://doi.org/10.1109/3DV.2016.79

Shen, H., Wang, R., Zhang, J., & McKenna, S. J. (2017). Boundary-aware fully convolutional network for brain tumor segmentation. *Medical Image Computing and Computer-Assisted Intervention, 10434*, 433–441. https://doi.org/10.1007/978-3-319-66185-8_50

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis, 43*, 98–111. https://doi.org/10.1016/j.media.2017.10.002

Islam, J., & Zhang, Y. (2019). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informatics, 5*(2), 2. https://doi.org/10.1186/s40708-018-0080-3

Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., & Mengko, T. R. (2019). Brain tumor classification using convolutional neural network. *World Congress on Medical Physics and Biomedical Engineering, 68*(1), 183–189. https://doi.org/10.1007/978-981-10-9035-6_33

Cheng, J., Yang, W., Huang, M., Huang, W., Jiang, J., Zhou, Y., Yang, R., Zhao, J., Feng, Y., Feng, Q., &

Chen, W. (2017). Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PLOS ONE, 11*(6), e0157112. https://doi.org/10.1371/journal.pone.0157112

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *Medical Image Computing and Computer-Assisted Intervention Workshop*. https://doi.org/10.48550/arXiv.1804.03999

Nuechterlein, N., & Mehta, S. (2019). 3D-ESPNet with pyramidal refinement for volumetric brain tumor image segmentation. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 11384*, 245–253. https://doi.org/10.1007/978-3-030-11726-9_22

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2010.11929

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint*. https://doi.org/10.48550/arXiv.2102.04306

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2021). Swin-Unet: Unet-like pure transformer for medical image segmentation. *European Conference on Computer Vision Workshops*. https://doi.org/10.48550/arXiv.2105.05537

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods, 18*(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., & Li, J. (2021). TransBTS: Multimodal brain tumor segmentation using transformer. *Medical Image Computing and Computer-Assisted Intervention, 12901*, 109–119. https://doi.org/10.1007/978-3-030-87193-2_11

Cheng, G., Yang, C., Yao, X., Guo, L., & Han, J. (2022). When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs. *IEEE Transactions on Geoscience and Remote Sensing, 56*(5), 2811–2821. https://doi.org/10.1109/TGRS.2017.2783902

Ghaffari, M., Sowmya, A., & Oliver, R. (2020). Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the BraTS 2012–2018 challenges. *IEEE Reviews in Biomedical Engineering, 13*, 156–168. https://doi.org/10.1109/RBME.2019.2946868

Roy, A. G., Navab, N., & Wachinger, C. (2019). Concurrent spatial and channel squeeze and excitation in fully convolutional networks. *Medical Image Computing and Computer-Assisted Intervention, 11070*, 421–429. https://doi.org/10.1007/978-3-030-00928-1_48

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986

Zhou, H. Y., Guo, J., Zhang, Y., Han, X., Yu, L., Wang, L., & Yu, Y. (2021). nnFormer: Interleaved transformer for volumetric segmentation. *arXiv preprint*. https://doi.org/10.48550/arXiv.2109.03201

Peiris, H., Hayat, M., Chen, Z., Egan, G., & Harandi, M. (2022). A robust volumetric transformer for accurate 3D tumor segmentation. *Medical Image Computing and Computer-Assisted Intervention, 13434*, 162–172. https://doi.org/10.1007/978-3-031-16443-9_16

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H., & Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 574–584. https://doi.org/10.1109/WACV51458.2022.00181

Cheng, Z., Lin, M., Zhao, Y., Gao, X., & Shen, J. (2022). Masked autoencoders for point cloud self-supervised learning. *European Conference on Computer Vision, 13662*, 604–621. https://doi.org/10.1007/978-3-031-20086-1_35

Luu, H. M., & Park, S. H. (2021). Extending nn-UNet for brain tumor segmentation. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 12963*, 173–186. https://doi.org/10.1007/978-3-031-09002-8_16

Jiang, Z., Ding, C., Liu, M., & Tao, D. (2022). Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task. *Brainlesion Workshop, 11992*, 231–241. https://doi.org/10.1007/978-3-030-46640-4_22

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., & Davatzikos, C. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data, 4*(1), 170117. https://doi.org/10.1038/sdata.2017.117