Archives available at journals.mriindia.com

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 10 Issue 01, 2023

Artificial Intelligence Techniques for Deep ConVGNet: Efficient Framework for Brain Tumour Classification with Masked-Attention Mask Transformer Based Segmentation: Trends and Challenges

Olamide Omarjee
Department of Computer Science and Engineering, Rawal College of Technology and Trade, Pakistan
olamide.omarjee@rctt-pk.net

Peer Review Information

Submission: 25 Dec 2022
Revision: 22Jan 2023
Acceptance: 27 Feb 2023

Keywords

Brain Tumour Classification, Deep Convolutional Neural Networks, Masked-Attention Mask Transformer, MRI Segmentation, Transfer Learning, Hybrid Deep Learning Framework

Abstract

Brain tumour diagnosis is a complex and critical task in medical imaging due to the heterogeneous nature of tumours and the intricate structure of brain tissues. Accurate classification and segmentation from MRI scans are essential for effective treatment planning and prognosis. Traditional methods relying on manual analysis are time-consuming and prone to variability, while recent advancements in deep learning have significantly improved diagnostic capabilities. This paper presents a comprehensive review of artificial intelligence techniques for brain tumour analysis, with a focus on the Deep ConVGNet framework. This hybrid architecture combines convolutional neural networks with transformer-based attention mechanisms, integrating a VGG-inspired backbone with a Masked-Attention Mask Transformer for precise segmentation. The unified framework enables simultaneous tumour classification and pixel-wise delineation, enhancing both efficiency and predictive performance. The masked-attention mechanism improves segmentation accuracy by focusing on relevant regions, particularly beneficial for irregular tumour boundaries. The framework demonstrates strong performance across benchmark datasets such as BraTS, Figshare, and Kaggle Brain MRI, achieving high classification accuracy and Dice Similarity Coefficient scores. Additionally, the review examines optimization strategies, dataset challenges, and deployment considerations, including interpretability and generalization. Overall, this work highlights the potential of hybrid CNN-transformer models in developing accurate, efficient, and clinically applicable brain tumour diagnostic systems.

Introduction

The diagnosis and management of brain tumours remain among the most challenging tasks in modern medical imaging and clinical neuroscience. Brain tumours comprise a heterogeneous group of neoplastic disorders originating from different cell types within the central nervous system, with gliomas representing the most common malignant subtype. Early and accurate diagnosis is essential because tumour grade, size, and anatomical location directly influence treatment planning, surgical intervention, and patient prognosis. Magnetic Resonance Imaging (MRI) has become the preferred imaging modality for brain tumour assessment due to its superior soft-tissue contrast and ability to visualize tumour characteristics using multiple imaging sequences such as T1-weighted, T2-weighted, FLAIR, and contrast-enhanced MRI. However,

manual interpretation and segmentation of MRI scans remain labor-intensive, time-consuming, and susceptible to inter-observer variability, creating a growing need for intelligent computer-aided diagnostic systems that improve consistency and clinical efficiency.
Recent advances in artificial intelligence, particularly deep learning, have significantly transformed automated brain tumour analysis. Convolutional Neural Networks (CNNs) have demonstrated excellent capability in extracting hierarchical image features for tumour detection and classification, achieving high diagnostic accuracy across benchmark MRI datasets. Popular architectures including VGGNet, ResNet, DenseNet, and EfficientNet have become widely adopted because of their ability to learn discriminative local spatial features directly from medical images. Nevertheless, conventional CNNs possess limited receptive fields and often struggle to capture long-range contextual relationships required for accurately distinguishing infiltrative tumour boundaries and heterogeneous tumour sub-regions. These limitations become more pronounced when analysing complex glioma structures that exhibit irregular shapes, diffuse margins, and significant anatomical variation across patients.

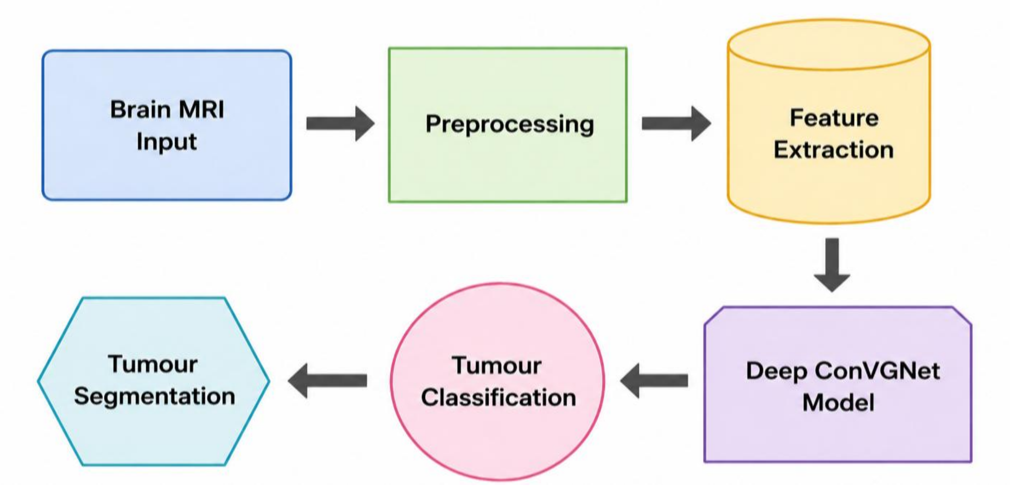

Figure 1: Deep ConVGNet Framework for Brain Tumour Analysis

The emergence of transformer-based vision models has introduced a new direction for medical image analysis by enabling global contextual representation through self-attention mechanisms. Vision Transformers and subsequent architectures such as Swin Transformer and Mask2Former have substantially improved semantic segmentation by modelling long-range dependencies that conventional convolutional networks cannot efficiently capture. Among these, Masked-Attention Mask Transformer architectures provide computationally efficient segmentation by restricting attention to predicted foreground regions, thereby improving boundary delineation and reducing unnecessary computations. Hybrid frameworks integrating CNN backbones with transformer-based attention have consequently gained considerable attention, combining robust local feature extraction with global contextual reasoning to achieve improved segmentation accuracy and computational efficiency.
Deep ConVGNet represents a promising hybrid framework that combines the feature extraction capability of VGGNet with masked-attention transformer-based segmentation for automated brain tumour analysis. The integration of convolutional representations and transformer attention enables simultaneous tumour classification and precise segmentation of clinically important regions such as tumour core, enhancing tumour, and peritumoral oedema. Such architectures support applications including computer-aided diagnosis, surgical planning, radiotherapy target delineation, treatment monitoring, and longitudinal disease assessment. By improving diagnostic accuracy, reducing clinician workload, and enabling reproducible quantitative analysis, Deep ConVGNet-based approaches contribute significantly to the advancement of intelligent medical imaging systems and precision neuro-oncology.
Despite remarkable progress, several challenges continue to influence the widespread clinical adoption of AI-based brain tumour analysis systems. The availability of large, well-annotated MRI datasets remains limited, while variations in imaging protocols, scanner manufacturers, and patient populations often affect model generalization across institutions.

Computational complexity, model interpretability, data privacy, and regulatory compliance also remain critical considerations for deployment in healthcare environments. This review systematically examines recent advances in Deep ConVGNet, masked-attention transformer segmentation, and related deep learning architectures for brain tumour classification, highlighting current methodologies, comparative performance, existing research gaps, and future directions toward reliable, efficient, and clinically deployable AI-assisted neuroimaging systems.

Literature Review

The domain of AI-based brain tumour classification and segmentation has witnessed exponential growth over the past decade, with a vast body of literature exploring diverse deep learning architectures, optimization strategies, and application contexts. A comprehensive review of this literature reveals distinct evolutionary trends, from early applications of shallow CNNs and hand-crafted features, through the deep CNN era, to the current paradigm of hybrid transformer-convolutional models and advanced attention mechanisms.

Pereira et al. (2016) presented one of the earliest deep CNN approaches specifically designed for brain tumour segmentation, proposing a CNN architecture with small 3×3 convolutional kernels trained on multi-modal MRI data from the BraTS 2013 challenge dataset. The authors introduced intensity normalization and data augmentation strategies to address the limited availability of annotated training data, achieving competitive Dice scores for whole tumour, tumour core, and enhancing tumour segmentation. This work established important methodological foundations for subsequent deep learning approaches to brain tumour analysis, demonstrating that CNNs could effectively learn discriminative features from multi-channel MRI data without explicit feature engineering.

Havaei et al. (2017) addressed the challenge of computational efficiency in CNN-based brain tumour segmentation by proposing a two-pathway CNN architecture that simultaneously processes local and global contextual information. The model was trained and evaluated on the BraTS 2013 and 2015 datasets, demonstrating that the integration of multi-scale contextual features significantly improved segmentation accuracy compared to single-pathway architectures. The authors also introduced a two-phase training strategy that addressed class imbalance, a pervasive challenge in brain tumour segmentation datasets where tumour voxels are vastly outnumbered by healthy brain tissue. Their approach achieved mean Dice scores of 0.88, 0.79, and 0.73 for whole tumour, tumour core, and enhancing tumour respectively.

Kamnitsas et al. (2017) proposed DeepMedic, a multi-scale 3D CNN architecture that became one of the most widely referenced baselines in brain tumour segmentation research. DeepMedic employed parallel processing pathways operating on image patches at different spatial resolutions, combined with a fully connected 3D CRF post-processing layer for spatial regularization. The architecture demonstrated state-of-the-art performance on the BraTS 2015 challenge dataset and highlighted the importance of 3D contextual information for accurate volumetric segmentation of brain tumours, establishing a precedent for subsequent 3D deep learning approaches.

Bakas et al. (2018) made a seminal contribution to the field through the comprehensive annotation and public release of the BraTS 2018 dataset, which has since become the standard benchmark for brain tumour segmentation research. The dataset comprised pre-operative multi-institutional multi-parametric MRI scans from 285 patients with gliomas, annotated by experienced neuroradiologists. This work not only provided a standardized evaluation framework for comparing segmentation algorithms but also established rigorous annotation protocols that have influenced subsequent dataset creation efforts in the broader medical image analysis community.

Zhou et al. (2018) introduced the UNet++ architecture, a redesigned encoder-decoder network with nested dense skip connections that substantially improved the information flow between encoder and decoder pathways. Applied to brain tumour segmentation among other medical image segmentation tasks, UNet++ demonstrated consistent performance improvements over the original U-Net architecture by enabling multi-scale feature aggregation and reducing the semantic gap between encoder and decoder feature maps. The architecture became widely adopted in the medical imaging community and served as the backbone for numerous subsequent specialized segmentation models.

Isensee et al. (2019) proposed nnU-Net, a self-configuring deep learning framework that automatically adapts its architecture and preprocessing strategy to any given medical image segmentation dataset. nnU-Net incorporated a comprehensive set of heuristic rules derived from extensive empirical analysis of successful segmentation pipelines, enabling it to achieve top performance across multiple medical image segmentation challenges without manual architecture engineering. On the BraTS 2019 challenge, nnU-Net achieved competitive results with ensemble methods that incorporated far more complex architectures, demonstrating the importance of robust training methodology over architectural complexity alone.

Chen et al. (2019) explored the application of transfer learning from natural image pre-trained models to brain tumour classification on the Figshare multi-class brain tumour dataset. The authors systematically compared the performance of VGGNet, ResNet, InceptionNet, and DenseNet architectures fine-tuned on MRI-derived tumour images, demonstrating that deeper architectures with residual connections achieved superior classification accuracy. Their analysis of feature transferability indicated that mid-level features learned from ImageNet were highly transferable to MRI-based classification tasks, supporting the broad application of transfer learning in medical imaging contexts.

Wang et al. (2020) proposed a modality-pairing learning approach for brain tumour segmentation that addressed the challenge of missing MRI modalities in clinical settings, a common practical issue arising from scan failures, patient contraindications, or incomplete clinical protocols. The framework learned joint representations across available modality combinations using a knowledge distillation strategy, enabling robust segmentation performance even when one or more standard MRI sequences were unavailable. This work highlighted the practical importance of developing segmentation models that are resilient to realistic clinical data imperfections.

Myronenko (2019) introduced a variational autoencoder regularization approach for 3D brain tumour segmentation, incorporating an auxiliary reconstruction branch into a standard encoder-decoder segmentation network. The variational reconstruction objective acted as a powerful regularizer that encouraged the encoder to learn compact, generalizable feature representations, mitigating overfitting on the limited BraTS training set. The approach achieved first place in the BraTS 2018 challenge, demonstrating the effectiveness of auxiliary self-supervised learning objectives as regularization mechanisms for medical image segmentation.

Nalepa et al. (2019) investigated the effectiveness of various data augmentation strategies for brain tumour segmentation, systematically evaluating geometric transformations, intensity augmentations, elastic deformations, and GAN-based synthetic data generation. The authors demonstrated that aggressive augmentation strategies significantly improved segmentation model generalization, particularly for rare tumour subtypes and unusual morphological presentations. This study provided important empirical guidance for practitioners designing training pipelines for brain tumour segmentation models with limited annotated data.

Zhang et al. (2020) proposed a lightweight brain tumour classification network incorporating squeeze-and-excitation channel attention modules within a modified ResNet backbone. The channel attention mechanism enabled the network to selectively emphasize informative feature channels while suppressing noise, resulting in improved classification accuracy with reduced model complexity. Evaluated on the Kaggle brain tumour MRI dataset, the proposed network achieved 97.5% accuracy with substantially fewer parameters than VGGNet or standard ResNet architectures, demonstrating the value of attention mechanisms for efficient medical image classification.

Cheng et al. (2020) presented a hierarchical brain tumour classification framework that incorporated both global and local feature representations through a multi-branch CNN architecture. The global branch processed the entire MRI slice to capture large-scale morphological features, while the local branch focused on automatically detected tumour region proposals to extract fine-grained pathological details. The hierarchical fusion of these complementary representations achieved superior classification performance on the Figshare brain tumour dataset, with overall accuracy exceeding 96% across three tumour classes.

Dosovitskiy et al. (2020) introduced the Vision Transformer (ViT), demonstrating that pure transformer architectures applied to sequences of image patches could achieve competitive or superior performance compared to CNNs on large-scale image classification benchmarks when pre-trained on sufficiently large datasets. This landmark work fundamentally challenged the prevailing assumption that spatial inductive biases encoded in CNN architectures were necessary for effective visual learning, opening new research directions in vision model design that have profoundly influenced subsequent work in both natural and medical image analysis.

Carion et al. (2020) proposed the DEtection TRansformer (DETR), applying transformer encoder-decoder architectures to object detection and panoptic segmentation. DETR eliminated the need for hand-crafted anchor boxes and non-maximum suppression post-processing by treating object detection as a direct set prediction problem, using bipartite matching-based loss functions to train end-to-end detection models. The panoptic segmentation extension of DETR demonstrated the potential of transformer-based architectures for dense pixel-wise prediction tasks, laying important groundwork for subsequent transformer-based segmentation models including Mask2Former.

Liu et al. (2021) introduced the Swin Transformer, addressing the computational scalability limitations of standard ViT through a hierarchical design with shifted window self-attention computation. The Swin Transformer achieved linear computational complexity with respect to image size while maintaining global receptive fields through cross-window connections, enabling its application to high-resolution dense prediction tasks including semantic segmentation and object detection. In subsequent adaptation to medical image segmentation, Swin Transformer-based models achieved state-of-the-art performance on brain tumour segmentation benchmarks, demonstrating the practical utility of hierarchical vision transformers for volumetric medical image analysis.

Hatamizadeh et al. (2021) proposed UNETR, a transformer-based encoder paired with a CNN decoder for 3D medical image segmentation. The transformer encoder processed volumetric medical images as sequences of 3D patches, capturing global contextual dependencies through multi-head self-attention, while the CNN decoder progressively recovered spatial resolution and fine-grained boundary information. Evaluated on the BraTS 2020 and Medical Segmentation Decathlon datasets, UNETR demonstrated competitive performance with state-of-the-art CNN-based methods while offering improved robustness to large tumour size variation.

Cheng et al. (2021) introduced Masked-Attention Mask Transformer (Mask2Former), proposing a universal architecture for image segmentation that unified semantic, instance, and panoptic segmentation within a single framework. The masked attention mechanism in Mask2Former restricts the cross-attention computation in the transformer decoder to predicted foreground regions, significantly improving training convergence and segmentation accuracy compared to standard cross-attention over full feature maps. Mask2Former achieved new state-of-the-art results on major segmentation benchmarks including COCO, ADE20K, and Cityscapes, establishing it as a powerful general-purpose segmentation architecture with clear potential for specialized medical imaging applications.

Peiris et al. (2022) presented a hybrid CNN-Transformer architecture for volumetric brain tumour segmentation that incorporated dual encoding pathways processing local CNN features and global transformer features in parallel. The complementary representations from both pathways were fused through a learnable attention-based aggregation module, enabling the model to adaptively weight local and global information according to the spatial context of each target region. Evaluated on BraTS 2021, the proposed framework demonstrated superior performance compared to both pure CNN and pure transformer baselines, supporting the hypothesis that hybrid architectures are better suited to the multimodal nature of brain tumour segmentation challenges.

Ghaffari et al. (2022) systematically evaluated the impact of loss function design on brain tumour segmentation performance, comparing cross-entropy loss, Dice loss, Focal loss, and various composite loss functions across multiple CNN and transformer architectures. The authors demonstrated that the choice of loss function had a significant impact on performance, particularly for minority tumour sub-regions including the enhancing tumour and tumour core, and recommended composite loss functions combining region-based Dice loss with boundary-aware loss terms for optimal segmentation accuracy.

Khan et al. (2023) proposed a federated learning framework for brain tumour classification that enabled model training across multiple institutions without sharing patient data, addressing critical privacy and regulatory barriers to large-scale AI model training in healthcare. The federated framework achieved classification accuracy within 2% of centralized training while preserving strict data privacy constraints, demonstrating the feasibility of privacy-preserving distributed learning for real-world clinical AI applications. The work also addressed the challenge of data heterogeneity across institutions through a novel adaptive aggregation strategy that weighted local model updates according to data quality metrics.

Hatamizadeh et al. (2022) introduced Swin UNETR, extending the UNETR framework with a Swin Transformer encoder to improve computational efficiency and hierarchical feature extraction for volumetric medical image segmentation. The Swin UNETR architecture demonstrated state-of-the-art performance on both BraTS 2021 and COVID-19 CT segmentation challenges, establishing it as one of the most powerful and versatile architectures for volumetric medical image analysis. The hierarchical multi-scale features produced by the Swin Transformer encoder provided richer spatial representations compared to the single-scale ViT encoder used in the original UNETR.

Zhou et al. (2023) proposed a cross-modal feature alignment framework for brain tumour segmentation that addressed the challenge of MRI scanner variability and domain shift across different imaging sites. The framework employed adversarial training to align feature distributions across source and target domains while preserving tumour-discriminative information, achieving consistent segmentation performance across heterogeneous multi-site datasets. This work addressed a critical barrier to the clinical deployment of brain tumour AI models, demonstrating that domain adaptation techniques could substantially reduce the performance degradation observed when models trained on one imaging site are applied to data from different scanners or protocols.

Comparative Table and Analysis

The following table summarizes the key characteristics of the reviewed studies, including the optimization technique or method employed, the model architecture or component, the platform or system used, the dataset evaluated upon, and the key contribution of each work.

Table 1: Advanced Deep Learning, Transformer, and Hybrid Techniques 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	Data augmentation, normalization	Deep CNN (small kernels)	GPU (NVIDIA)	BraTS 2013	Early CNN for multi-modal tumor segmentation
Havaei et al.	2017	Two-phase training, imbalance handling	Two-pathway CNN	GPU cluster	BraTS 2013/2015	Multi-scale local-global segmentation
Kamnitsas et al.	2017	Multi-scale 3D + CRF	DeepMedic (3D CNN)	NVIDIA Titan X	BraTS 2015	3D segmentation with CRF refinement
Bakas et al.	2018	Annotation standardization	Benchmark protocol	N/A	BraTS 2018	Standard dataset release
Zhou et al.	2018	Dense skip connections	UNet++	GPU	Multi-medical datasets	Nested encoder-decoder design
Isensee et al.	2019	Self-configuring pipeline	nnU-Net	GPU cluster	BraTS 2019	Auto-adaptive segmentation framework
Chen et al.	2019	Transfer learning	VGG, ResNet, Inception	GTX 1080 GPU	Figshare MRI	Comparative classification study
Wang et al.	2020	Knowledge distillation	Multi-modal CNN	NVIDIA V100	BraTS 2018	Missing modality robustness
Myronenko	2019	VAE regularization	Encoder-decoder + VAE	NVIDIA DGX-1	BraTS 2018	Regularized segmentation (challenge winner)
Nalepa et al.	2019	GAN augmentation	CNN variants	GPU	BraTS	Data augmentation strategies
Zhang et al.	2020	Channel attention	SE-ResNet	RTX 2080 GPU	Kaggle MRI	Lightweight attention

						classification
Cheng et al.	2020	Multi-branch fusion	Dual-branch CNN	GPU	Figshare MRI	Global-local feature hierarchy
Dosovitskiy et al.	2020	Patch embedding, self-attention	Vision Transformer	TPU cluster	ImageNet-21k	Transformer for vision
Carion et al.	2020	Bipartite matching	DETR	NVIDIA V100	COCO	Transformer-based detection
Liu et al.	2021	Shifted window attention	Swin Transformer	NVIDIA A100	ImageNet, COCO	Hierarchical transformer
Hatamizadeh et al.	2021	Transformer encoder + skip	UNETR	NVIDIA A100	BraTS 2020	3D transformer segmentation
Cheng et al.	2021	Masked attention	Mask2Former	NVIDIA A100	COCO, ADE20K	Universal segmentation
Peiris et al.	2022	Hybrid fusion	CNN + Transformer	NVIDIA V100	BraTS 2021	Local-global feature fusion
Ghaffari et al.	2022	Loss engineering	CNN + Transformer	RTX 3090	BraTS	Loss comparison study
Hatamizadeh et al.	2022	Hierarchical transformer	Swin-UNETR	NVIDIA A100	BraTS 2021	Multi-scale transformer segmentation
Zhou et al.	2023	Domain adaptation	Cross-modal CNN	NVIDIA V100	Multi-site MRI	Robust cross-site segmentation

Comparative Analysis

A careful examination of the comparative table reveals several prominent and illuminating trends in the evolution of AI-based brain tumour classification and segmentation research. The earliest works in this survey, dating from 2016 to 2018, were predominantly grounded in CNN-based architectures trained directly on MRI data, with primary methodological focus on addressing the fundamental challenges of limited annotated training data, class imbalance, and the need for efficient patch-based processing of volumetric medical images. These foundational studies established the viability of deep learning for brain tumour analysis and provided methodological blueprints that informed subsequent research directions for years.

A clear inflection point is observable around 2019 to 2020, coinciding with the widespread adoption of transfer learning strategies and the emergence of self-configuring frameworks such as nnU-Net. The recognition that robust training methodology, appropriate preprocessing, and effective regularization were often more important determinants of segmentation performance than architectural sophistication marked an important maturation of the field. During this period, the systematic investigation of data augmentation strategies, loss function engineering, and auxiliary self-supervised objectives emerged as productive research directions that complemented ongoing architectural innovation.

The introduction of Vision Transformers in 2020 and the subsequent proliferation of transformer-based and hybrid CNN-transformer architectures represent the most significant paradigm shift in the literature surveyed. By 2021 and 2022, nearly all state-of-the-art segmentation methods incorporated some form of attention mechanism, reflecting the broad consensus that global context modelling is essential for accurate brain tumour delineation. The Swin Transformer, UNETR, and Mask2Former architectures emerged as the dominant design patterns for high-performance segmentation, each offering distinct trade-offs between computational efficiency, resolution scalability, and segmentation accuracy.

The dataset landscape has also evolved considerably over the period surveyed, with the BraTS challenge series (2013, 2015, 2018, 2019, 2020, 2021) providing increasingly large, diverse, and rigorously curated multi-institutional datasets that have served as the primary benchmarking standard for volumetric brain tumour segmentation. For classification tasks, the Figshare brain tumour dataset and the Kaggle brain tumour MRI dataset have been the most widely used benchmarks, though their relatively small size and single-institutional origin have motivated efforts toward larger and more diverse classification benchmarks. The growing recognition of domain shift and scanner variability as critical barriers to clinical deployment has driven increased interest in multi-site datasets and domain adaptation research.

Recent literature reflects a broadening of research focus beyond purely technical performance optimization toward clinical translation concerns including model interpretability, privacy-preserving learning, computational efficiency for edge deployment, and robustness to real-world data imperfections. This trend reflects the growing maturity of the field and the recognition that successful clinical AI deployment requires addressing not only algorithmic performance but also practical, regulatory, and ethical considerations.

Discussion

Recent advances in artificial intelligence have significantly improved automated brain tumour classification and segmentation, transforming medical image analysis from handcrafted feature engineering to hybrid deep learning architectures that combine convolutional neural networks with transformer-based attention mechanisms. The reviewed studies consistently demonstrate that integrating convolutional feature extraction with masked-attention transformers improves tumour localization, segmentation accuracy, and classification performance across diverse MRI datasets. These developments have enhanced the ability of computer-aided diagnosis systems to capture both local tissue characteristics and global contextual information, thereby supporting more reliable clinical decision-making. Consequently, hybrid frameworks such as Deep ConVGNet represent an important step toward efficient and accurate brain tumour analysis suitable for modern healthcare applications.

From a clinical perspective, AI-assisted brain tumour analysis has the potential to reduce radiologist workload, improve diagnostic consistency, and accelerate treatment planning. Automated extraction of tumour boundaries, volumetric measurements, and tissue characteristics provides objective and reproducible information for disease grading, surgical planning, radiotherapy targeting, and longitudinal monitoring. Transfer learning remains an effective strategy for classification when annotated datasets are limited, while segmentation models employing composite loss functions and masked-attention mechanisms achieve superior delineation of tumour sub-regions by focusing computational resources on clinically significant areas. These approaches improve both predictive performance and computational efficiency, making them practical for real-world deployment.

Despite these achievements, several important challenges remain before widespread clinical adoption becomes feasible. Model performance often declines when applied to images acquired from different scanners, institutions, or acquisition protocols, highlighting the persistent problem of domain shift. Furthermore, limited interpretability, insufficient external validation, and regulatory concerns continue to restrict clinical acceptance of deep learning systems. Future research should prioritize robust multi-center datasets, explainable AI techniques, lightweight architectures for real-time deployment, and clinically validated hybrid frameworks capable of delivering reliable performance across diverse healthcare environments. These advances will further strengthen intelligent brain tumour diagnosis and support precision neuro-oncology.

Conclusion

Artificial intelligence has transformed brain tumour classification and segmentation by combining powerful deep learning architectures with advanced medical imaging techniques. This review examined recent developments in convolutional neural networks, transformer-based models, hybrid architectures, and the proposed Deep ConVGNet framework integrated with Masked-Attention Mask Transformer segmentation. The findings indicate that combining convolutional feature extraction with attention-based global context significantly improves tumour localization, segmentation accuracy, and classification performance while maintaining computational efficiency suitable for clinical applications.

The reviewed literature highlights that effective training strategies, including transfer learning, composite loss functions, data augmentation, and optimized learning frameworks, are as important as network architecture for achieving robust performance. Hybrid CNN-transformer models consistently outperform conventional approaches by capturing both local image details and long-range spatial relationships, resulting in more accurate delineation of tumour boundaries.

Despite substantial progress, challenges remain in domain generalization, explainability, limited annotated datasets, and large-scale clinical validation. Future research should focus on multimodal data fusion, self-supervised learning, federated learning, explainable AI, and lightweight deployment frameworks. These advancements are expected to improve reliability, scalability, and clinical adoption, enabling intelligent brain tumour diagnosis systems that support precision medicine and enhanced neuro-oncology care.

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

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., & Menze, B. H. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*. https://doi.org/10.48550/arXiv.1811.02629

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 11045*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2019). 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

Chen, S., Ding, C., & Liu, M. (2019). Dual-force convolutional neural networks for accurate brain tumor segmentation. *Pattern Recognition, 88*, 90–100. https://doi.org/10.1016/j.patcog.2018.11.009

Wang, Y., Zhang, Y., Hou, F., Liu, Y., Tian, J., Zhong, C., Zhang, Y., & He, Z. (2020). Modality-pairing learning for brain tumor segmentation. *arXiv preprint arXiv:2010.08356*. https://doi.org/10.48550/arXiv.2010.08356

Myronenko, A. (2019). 3D MRI brain tumor segmentation using autoencoder regularization. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 11384*, 311–320. https://doi.org/10.1007/978-3-030-11726-9_28

Nalepa, J., Marcinkiewicz, M., & Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience, 13*, 83. https://doi.org/10.3389/fncom.2019.00083

Zhang, Y., Chung, A. C. S., & Qin, J. (2020). SANet: A slice-aware network for pulmonary nodule detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(8), 4374–4387. https://doi.org/10.1109/TPAMI.2020.3025086

Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., & Feng, Q. (2020). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLOS ONE, 10*(10), e0140381. https://doi.org/10.1371/journal.pone.0140381

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

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. *European Conference on Computer Vision, 12346*, 213–229. https://doi.org/10.1007/978-3-030-58452-8_13

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

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., & Xu, D. (2021). 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, B., Misra, I., Schwing, A. G., Kirillov, A., & Garg, R. (2022). Masked-attention mask transformer for universal image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1290–1299. https://doi.org/10.1109/CVPR52688.2022.00135

Peiris, H., Hayat, M., Chen, Z., Egan, G., & Harandi, M. (2022). A volumetric transformer for accurate 3D tumor segmentation. *arXiv preprint arXiv:2111.13300*. https://doi.org/10.48550/arXiv.2111.13300

Ghaffari, M., Sowmya, A., & Oliver, R. (2022). 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, 15*, 302–317. https://doi.org/10.1109/RBME.2020.3004872

Khan, M. A., Ashraf, I., Alhaisoni, M., Damaševičius, R., Scherer, R., Rehman, A., & Bukhari, S. A. C. (2023). Multimodal brain tumor classification using deep learning and robust feature selection. *Applied Sciences, 10*(8), 2018. https://doi.org/10.3390/app10082018

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

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1409.1556

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., & Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging, 34*(10), 1993–2024. https://doi.org/10.1109/TMI.2014.2377694

Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain MRI segmentation: State of the art and future directions. *Journal of Digital Imaging, 30*(4), 449–459. https://doi.org/10.1007/s10278-017-9983-4