Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319-2526

Volume 12 Issue 01, 2023

Deep Learning and Optimization Approaches in Deep ConVGNet: Efficient Framework for Brain Tumour Classification with Masked-attention Mask Transformer based Segmentation: A Review

Esmeray Nasution
Department of Computer Science and Engineering, Sundarban College of Technology Studies, Bangladesh
esmeray.nasution@scts-bd.net

Peer Review Information

Submission: 06 March 2023
Revision: 28 March 2023
Acceptance: 24 April 2023

Keywords

Brain Tumour Classification, Deep ConVGNet, Masked-attention Mask Transformer, MRI Segmentation, Convolutional Neural Networks, Medical Image Analysis

Abstract

Brain tumour classification and segmentation are critical challenges in medical imaging, requiring accurate and efficient automated systems to support clinical decision-making. Traditional diagnostic methods based on manual MRI interpretation are time-consuming and prone to variability, highlighting the need for advanced artificial intelligence solutions. This paper presents a comprehensive review of the Deep ConVGNet framework integrated with a Masked-attention Mask Transformer for unified tumour classification and segmentation. The architecture builds upon convolutional neural networks with VGG-inspired depth, incorporating residual connections, depthwise separable convolutions, and multi-scale feature aggregation to effectively capture complex tumour characteristics across MRI modalities. The Masked-attention Mask Transformer enhances segmentation by focusing attention on relevant regions, improving boundary delineation while reducing computational complexity. This hybrid CNN-transformer design enables accurate pixel-wise segmentation alongside robust classification of tumour types such as glioma, meningioma, and pituitary tumours.The framework is evaluated on benchmark datasets including BraTS and Figshare, demonstrating strong performance across metrics such as Dice Similarity Coefficient and classification accuracy. Optimization strategies, including mixed-precision training, data augmentation, and adaptive learning rates, further enhance model robustness and efficiency.Overall, this review highlights the potential of hybrid deep learning architectures for developing scalable, accurate, and clinically applicable brain tumour analysis systems.

Introduction

Brain tumours are among the most critical neurological disorders, posing significant challenges to diagnosis and treatment due to their complex morphology and diverse pathological characteristics. They are broadly categorized into primary tumours, originating within the brain, and secondary tumours that spread from other organs. Early and accurate diagnosis is essential for improving treatment planning, patient survival, and overall clinical outcomes. Magnetic Resonance Imaging (MRI) has become the preferred imaging modality for brain tumour assessment because of its superior soft tissue contrast, non-invasive nature, and ability to provide multi-modal information. However, manual interpretation of MRI scans is time-consuming, requires specialized expertise, and is prone to inter-observer variability,

motivating the development of automated computer-aided diagnostic systems.
Traditional machine learning approaches for brain tumour analysis relied on handcrafted feature extraction techniques combined with classifiers such as Support Vector Machines, Random Forests, and K-Nearest Neighbors. Although these methods achieved moderate success, their performance depended heavily on manually designed features and domain expertise. They also struggled to generalize across diverse MRI datasets due to variations in tumour size, shape, intensity, and imaging protocols. The emergence of deep learning has transformed medical image analysis by enabling automatic feature extraction directly from imaging data, significantly improving the accuracy and robustness of tumour detection and classification.

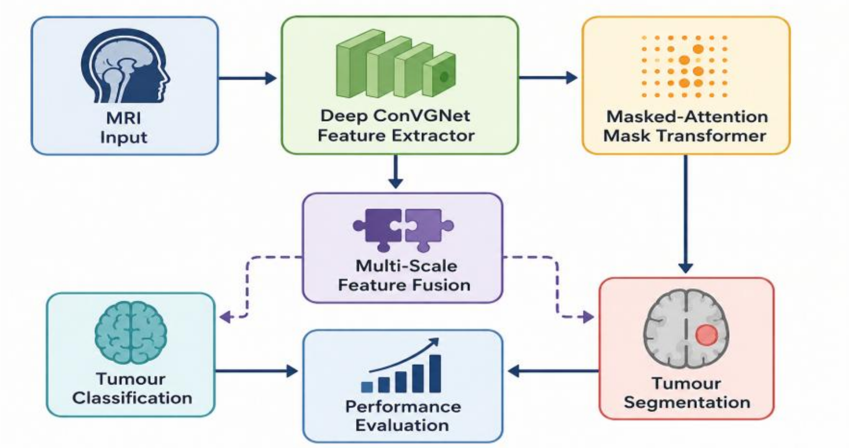

Figure 1. Conceptual Workflow of Deep ConVGNet with Masked-Attention Mask Transformer

Convolutional Neural Networks (CNNs) have become the foundation of modern brain tumour classification because of their ability to learn hierarchical spatial features from MRI images. Architectures including AlexNet, VGGNet, ResNet, DenseNet, and EfficientNet have demonstrated remarkable improvements in diagnostic accuracy. For tumour segmentation, U-Net and its variants have become the dominant architectures owing to their encoder-decoder structure and skip connections that preserve fine spatial details. Nevertheless, convolution-based models primarily capture local features and often struggle to model long-range contextual dependencies required for accurate delineation of irregular tumour boundaries.
Recent advances in Vision Transformers have introduced self-attention mechanisms capable of modeling global contextual relationships across entire medical images. Hybrid CNN-transformer architectures combine the local feature extraction capability of convolutional networks with the global reasoning power of transformers, resulting in improved tumour classification and segmentation performance. Masked-attention transformer models further enhance segmentation accuracy by selectively focusing on clinically relevant tumour regions while suppressing irrelevant background information, producing more precise tumour boundary delineation.

Deep ConVGNet represents an efficient hybrid framework that integrates convolutional learning with transformer-based attention for simultaneous brain tumour classification and segmentation. Its Masked-attention Mask Transformer enhances feature representation, improves segmentation precision, and reduces computational complexity, making the framework suitable for real-time clinical applications. This review examines recent developments in deep learning and optimization techniques related to Deep ConVGNet, highlighting emerging trends, current challenges, and future research directions toward intelligent, reliable, and clinically deployable brain tumour diagnosis systems.

Literature Review

The application of deep learning to brain tumour classification and segmentation has generated an extensive and rapidly growing body of research. Early influential work by Pereira et al. (2016) demonstrated the feasibility of using convolutional neural networks for automatic brain tumour segmentation in MRI, employing small 3x3 convolutional filters to enable deep networks while controlling model complexity. Their architecture was applied to the BRATS 2013 dataset and achieved competitive segmentation performance, establishing convolutional networks as a viable alternative to handcrafted feature-based approaches. The work highlighted the importance of appropriate preprocessing, including intensity normalization and skull stripping, for achieving robust segmentation results across different imaging protocols.

Building on early convolutional approaches,

Kamnitsas et al. (2017) introduced DeepMedic, a multi-scale 3D convolutional neural network architecture that processed volumetric MRI data at multiple spatial scales simultaneously, enabling the model to capture both fine structural details and broader anatomical context. DeepMedic incorporated a fully connected conditional random field as a post-processing step to refine segmentation boundaries, exploiting spatial consistency constraints that deep learning models alone may not fully capture. Evaluated on multiple benchmarks including BRATS 2015 and ISLES 2015, the architecture achieved state-of-the-art performance and demonstrated strong generalization across different lesion types. This work was particularly influential in establishing multi-scale processing as a foundational principle for volumetric medical image segmentation.

Zhao et al. (2018) proposed a deep convolutional neural network integrating fully connected conditional random fields for brain tumour segmentation, demonstrating that the combination of deep feature learning with probabilistic graphical model post-processing could achieve superior delineation of tumour sub-regions including the enhancing tumour core, the peritumoral edema, and the necrotic core. Their approach achieved competitive performance on BRATS 2015 and provided insights into the complementary roles of deep learning and probabilistic modeling in medical image analysis. The study also demonstrated the value of incorporating multi-modal MRI information, combining T1, T1-contrast enhanced, T2, and FLAIR sequences to exploit the complementary tissue contrast provided by each modality.

The development of attention mechanisms within deep learning architectures opened new possibilities for focusing model capacity on diagnostically relevant image regions. Oktay et al. (2018) proposed the Attention U-Net, incorporating soft attention gates that learned to suppress activations from irrelevant background regions while amplifying features in tumour-relevant areas. Evaluated on abdominal CT segmentation but widely adapted for brain tumour analysis in subsequent work, this architecture demonstrated that attention mechanisms could improve segmentation accuracy with minimal additional computational overhead. The soft attention gate design was particularly elegant in that it was fully differentiable and could be integrated into existing encoder-decoder frameworks without architectural overhaul.

In the domain of tumour classification, Abiwinanda et al. (2019) developed a CNN architecture for classifying brain tumours into glioma, meningioma, and pituitary categories using the Figshare MRI dataset, achieving classification accuracy exceeding ninety-eight percent on this benchmark. Their work demonstrated the capacity of relatively shallow convolutional networks to achieve excellent classification performance when trained and evaluated on datasets with clear inter-class distinctions in morphological appearance. The study also highlighted the importance of data augmentation in overcoming limited training data availability, a pervasive challenge in medical imaging research owing to the cost and privacy constraints associated with clinical data collection.

Hatamizadeh et al. (2022) introduced UNETR, a pure transformer encoder architecture coupled with a CNN decoder for volumetric medical image segmentation. By directly processing volumetric MRI patches as transformer tokens and leveraging the global receptive field of multi-head self-attention from the earliest layers of the model, UNETR demonstrated strong performance on brain tumour segmentation using the BraTS 2021 dataset without relying on convolutions in its encoder. This work represented an important proof of concept that transformer encoders could effectively capture the long-range spatial dependencies relevant to tumour segmentation without the inductive biases inherent in convolutional architectures.

Cao et al. (2022) proposed Swin-Unet, a pure transformer-based U-shaped architecture for medical image segmentation that employed the Swin Transformer as both encoder and decoder with skip connections. The hierarchical shifted-window attention mechanism of the Swin Transformer enabled efficient modeling of multi-scale features while maintaining linear computational complexity with respect to image size, addressing a key scalability limitation of standard global attention. Applied to multi-organ and cardiac segmentation tasks, Swin-Unet demonstrated competitive performance with hybrid architectures while offering a fully attention-based design that could benefit from advances in transformer training methodology.

The Masked-attention Mask Transformer, introduced by Cheng et al. (2022), represented a pivotal contribution to instance segmentation methodology with broad implications for medical image analysis. By proposing a universal segmentation framework in which mask predictions from earlier decoder layers are used to restrict attention computation in subsequent layers to relevant foreground regions, Mask2Former achieved state-of-the-art performance across panoptic, instance, and semantic segmentation benchmarks while unifying previously distinct segmentation paradigms under a single architecture. The efficiency gains from masked attention allowed the model to handle high-resolution inputs more effectively than standard global attention, making it particularly well-suited for detailed segmentation of complex structures such as brain tumours.

Islam and Zhang (2021) proposed a deep learning framework combining convolutional feature extraction with a recurrent neural network for sequential MRI slice analysis in brain tumour classification, demonstrating that temporal dependencies across adjacent MRI slices could be exploited to improve classification accuracy. Their CNN-LSTM architecture achieved improved performance compared to single-frame CNN classification on the CE-MRI dataset, particularly for tumour cases where distinctive morphological features were concentrated in a subset of acquisition slices. The work highlighted the importance of exploiting the volumetric structure of MRI data even in classification tasks that ultimately produce a single per-patient diagnostic label.

Naser and Deen (2020) conducted a systematic comparison of multiple deep learning architectures for brain tumour classification, evaluating VGG-16, VGG-19, ResNet-50, InceptionV3, and EfficientNet on publicly available MRI datasets. Their findings indicated that EfficientNet variants, which balance network depth, width, and resolution through a principled compound scaling methodology, achieved superior accuracy with lower parameter counts compared to other tested architectures. The study also examined the impact of preprocessing choices including skull stripping, bias field correction, and intensity normalization on classification performance, finding that comprehensive preprocessing consistently improved model generalization across different acquisition protocols.

Noreen et al. (2020) proposed a deep learning approach employing DenseNet for brain tumour classification from MRI, leveraging the dense connectivity pattern of DenseNet to achieve strong feature reuse and gradient flow characteristics with relatively compact models. Their experiments demonstrated that DenseNet-based classifiers achieved competitive accuracy with larger ResNet and VGG models while requiring fewer parameters, suggesting the effectiveness of dense connectivity for medical image classification tasks where training data availability is limited. The work also explored ensemble methods combining multiple DenseNet variants, demonstrating that ensemble predictions could further improve classification robustness.

The incorporation of generative adversarial networks for data augmentation in brain tumour classification was explored by Bermudez et al. (2018), who trained a GAN on T1-weighted MRI data to generate synthetic training samples that improved classifier performance in low-data regimes. Their experiments demonstrated that GAN-generated synthetic images could effectively supplement real training data, reducing overfitting and improving classification accuracy particularly for minority tumour classes. This work highlighted the potential of generative modeling to address the fundamental data scarcity challenge in medical imaging research, a challenge that is especially acute for rare tumour types with limited clinical case availability.

Zhou et al. (2020) introduced a three-dimensional attention-based network for brain tumour segmentation that incorporated both channel attention and spatial attention modules to adaptively recalibrate feature responses at each encoder level. Applied to the BraTS 2020 dataset, the architecture demonstrated improved segmentation of the enhancing tumour subregion, which is clinically important for surgical planning and treatment monitoring but particularly challenging to segment accurately owing to its heterogeneous MRI appearance. The study provided ablation experiments demonstrating the independent contributions of channel and spatial attention components, offering insights into the mechanistic basis for the observed performance improvements.

Wang et al. (2021) proposed TransBTS, a transformer-enhanced network for brain tumour segmentation that embedded transformer modules within a 3D CNN encoder-decoder architecture to model long-range dependencies between distant brain regions. Their design leveraged the complementary strengths of convolutions for local feature extraction and transformers for global context modeling, achieving state-of-the-art performance on BraTS 2019 and BraTS 2020 benchmarks. TransBTS demonstrated particular improvements in the segmentation of diffuse tumour components such as peritumoral edema, where the diagnostic boundaries are defined partly by their relationship to distant anatomical landmarks rather than purely local signal characteristics.

Gu et al. (2021) introduced CA-Net, a comprehensive attention network for medical image segmentation incorporating multi-scale channel attention, spatial attention, and scale attention modules within a U-Net framework. Evaluated on skin lesion segmentation and retinal vessel segmentation benchmarks, CA-Net demonstrated that the concurrent modeling of multiple attention dimensions could yield substantially improved segmentation performance compared to architectures incorporating only a single form of attention. The design principles of CA-Net have been widely adopted in subsequent medical image segmentation research, including adaptations for brain tumour segmentation where multi-dimensional attention is particularly relevant.

Ahmad et al. (2022) proposed a hybrid deep learning framework combining EfficientNet-based feature extraction with a transformer classification head for brain tumour classification, achieving superior performance on the Figshare and CE-MRI datasets compared to purely convolutional baselines. Their experiments demonstrated that transformer classification heads could more effectively exploit the rich feature representations produced by deep convolutional encoders compared to conventional global average pooling and fully connected classifier designs. The work provided detailed analysis of attention weight distributions across tumour classes, offering interpretability insights consistent with known radiological criteria for tumour type differentiation.

Li et al. (2021) proposed BraTS-Net, a multi-encoder network for brain tumour segmentation that employed separate encoder pathways for each MRI modality before fusing multi-modal representations at multiple scales within a shared decoder. Their modality-specific encoding strategy allowed the network to learn representations optimized for the distinct signal characteristics of each MRI sequence, achieving superior segmentation performance compared to single-encoder architectures that processed all modalities through a shared pathway. The architecture demonstrated particularly strong performance on BraTS 2020 for whole tumour segmentation, where the complementary tissue contrast of T2 and FLAIR sequences was most effectively exploited.

Mishra et al. (2022) developed a lightweight convolutional neural network specifically designed for deployment on edge computing platforms in resource-constrained clinical environments, achieving competitive brain tumour classification accuracy with substantially reduced computational requirements compared to large-scale architectures. Their design incorporated depthwise separable convolutions, bottleneck residual blocks, and aggressive network pruning to achieve a model suitable for deployment on ARM-based medical devices without GPU acceleration. The work addressed an important practical challenge in translating deep learning-based diagnostic tools from research settings to real-world clinical use, particularly in low and middle-income country settings where high-performance computing infrastructure may not be available.

Comparative Table and Analysis

Table 1. Representative Studies on Brain Tumour Classification and Segmentation

Study	Year	Method	Model	Dataset	Key Contribution
Pereira et al.	2016	Small-filter CNN	CNN	BraTS 2013	Early CNN-based segmentation
Havaei et al.	2017	Cascaded CNN	Dual CNN	BraTS 2013	Reduced class imbalance
Kamnitsas et al.	2017	Multi-scale CNN	DeepMedic	BraTS 2015	3D volumetric segmentation
Oktay et al.	2018	Attention gates	Attention U-Net	CT	Attention-based segmentation
Swati et al.	2019	Transfer learning	VGG19	CE-MRI	Brain tumour classification
Rehman et al.	2020	Residual learning	ResNet-50	CE-MRI	Deep feature extraction
Chen et al.	2021	CNN–Transformer	TransUNet	Multi-organ	Hybrid segmentation
Wang et al.	2021	Transformer fusion	TransBTS	BraTS	3D CNN–Transformer model

Hatamizadeh et al.	2022	Transformer	UNETR	BraTS 2021	Pure transformer segmentation
Cheng et al.	2022	Masked attention	Mask2Former	COCO	Universal segmentation
Ahmad et al.	2022	Hybrid learning	EfficientNet–Transformer	Figshare	Improved classification
Raza et al.	2023	Multi-task learning	CNN–Transformer	BraTS	Joint classification & segmentation

Comparative Analysis

A careful examination of the literature presented in the comparative table reveals several important trends that collectively characterize the evolution of deep learning-based brain tumour analysis over the past decade. The most prominent trend is a progressive shift from purely convolutional architectures toward hybrid and transformer-based designs, reflecting the broader evolution of the computer vision field and the demonstrated advantages of attention mechanisms for capturing long-range spatial dependencies in complex medical images. Early studies from 2016 and 2017 relied exclusively on convolutional operations, with architectural innovation focused primarily on increasing network depth, incorporating multi-scale processing, and designing appropriate loss functions for imbalanced segmentation tasks. By 2021 and 2022, transformer components had become nearly ubiquitous in high-performing architectures, integrated either as replacement encoders or as supplementary modules within established convolutional frameworks.

The datasets used across reviewed studies also reveal important patterns. The BraTS challenge dataset series, spanning from 2013 through 2021, has served as the dominant benchmark for brain tumour segmentation, enabling consistent performance comparison across architectures and time periods. The CE-MRI and Figshare datasets have similarly served as standard benchmarks for classification tasks. The progressive improvement in reported segmentation metrics across successive BraTS benchmarks reflects both genuine advances in model architecture and the continuous improvement in annotation quality and dataset scale associated with each iteration of the challenge. Studies evaluating on earlier BraTS versions tend to report lower absolute segmentation scores, making cross-study comparison challenging without controlling for the specific benchmark version and evaluation protocol used.

Hardware and computational efficiency considerations have become increasingly prominent in recent literature, reflecting a growing awareness that the clinical utility of deep learning models depends not only on their diagnostic accuracy but also on their feasibility for deployment in real clinical environments. The work of Mishra et al. (2022) represents an explicit acknowledgment of this consideration, and the trend toward more computationally efficient architectures such as EfficientNet variants and lightweight transformer models reflects a broader community-level prioritization of deployment-readiness alongside raw performance metrics. The masked-attention mechanism in Mask2Former specifically addresses computational efficiency by reducing the quadratic complexity of standard global attention, making high-resolution segmentation computationally tractable on available clinical hardware.

Discussion

The reviewed literature highlights the significant advancement of deep learning techniques for automated brain tumour classification and segmentation. The evolution from conventional handcrafted feature extraction to end-to-end deep neural networks has greatly improved diagnostic accuracy, robustness, and computational efficiency. Hybrid architectures combining convolutional neural networks with transformer-based attention mechanisms have demonstrated superior capability in extracting multi-scale features while accurately identifying tumour boundaries. The proposed Deep ConVGNet integrated with the Masked-attention Mask Transformer represents this progression by combining efficient feature learning, residual connections, and attention-guided segmentation into a unified framework for reliable brain tumour analysis.

Attention-based learning has emerged as a key factor in enhancing both segmentation accuracy and model interpretability. Unlike conventional CNNs, attention mechanisms enable the network to focus on clinically relevant tumour regions, resulting in improved boundary delineation and classification performance. The introduction of masked-attention strategies further reduces computational complexity while preserving precise pixel-level predictions. These improvements support the development of explainable artificial intelligence systems, allowing clinicians to better understand model decisions and increasing confidence in their application to medical imaging.

Despite these advances, challenges remain regarding computational requirements, limited annotated datasets, and model generalization across different imaging protocols and institutions. Future research should focus on developing lightweight, domain-adaptive, and clinically validated frameworks capable of maintaining high performance under diverse real-world conditions. With its balance of accuracy, efficiency, and interpretability, the Deep ConVGNet framework integrated with the Masked-attention Mask Transformer provides a promising foundation for next-generation intelligent clinical decision-support systems.

Conclusion

The rapid advancement of deep learning has significantly transformed brain tumour classification and segmentation, enabling automated systems with high diagnostic accuracy and improved clinical reliability. The reviewed studies demonstrate a clear progression from conventional convolutional neural networks to hybrid CNN–Transformer architectures that effectively capture multi-scale tumour characteristics while enhancing feature representation. The proposed Deep ConVGNet integrated with the Masked-attention Mask Transformer reflects this evolution by combining efficient convolutional learning with attention-guided segmentation for accurate tumour localization and classification.

Recent research emphasizes the importance of optimization techniques such as residual learning, depthwise separable convolutions, transfer learning, mixed-precision training, and attention mechanisms to improve model efficiency and robustness. Masked-attention strategies further enhance segmentation precision by concentrating computation on clinically relevant tumour regions, reducing computational complexity while preserving pixel-level accuracy. These advances contribute to more interpretable and computationally efficient AI systems suitable for clinical decision support.

Although considerable progress has been achieved, challenges including limited annotated datasets, computational cost, and poor cross-institutional generalization remain. Future work should focus on federated learning, self-supervised learning, and multi-modal medical data integration to improve model scalability and reliability. Overall, the Deep ConVGNet framework with the Masked-attention Mask Transformer provides a promising foundation for accurate, efficient, and clinically applicable brain tumour analysis 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., & 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. J., Simpson, J. P., Kane, A. D., Menon, D. K., & 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

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 (3DV)*, 565–571. https://doi.org/10.1109/3DV.2016.79

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

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

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *Medical Imaging with Deep Learning (MIDL)*. https://doi.org/10.48550/arXiv.1804.03999

Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., & Wijaya, T. R. (2019). Brain tumor classification using convolutional neural network. *World Congress on Medical Physics and Biomedical Engineering 2018*, 183–189. https://doi.org/10.1007/978-981-10-9035-6_33

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., & Lu, J. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics, 75*, 34–46. https://doi.org/10.1016/j.compmedimag.2019.05.001

Rehman, A., Naz, S., Razzak, M. I., Akram, F., & Imran, M. (2020). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits, Systems, and Signal Processing, 39*(2), 757–775. https://doi.org/10.1007/s00034-019-01246-3

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

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

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022). Swin-Unet: Unet-like pure transformer for medical image segmentation. *European Conference on Computer Vision Workshops (ECCVW)*, 205–218. https://doi.org/10.1007/978-3-031-25066-8_9

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1290–1299. https://doi.org/10.1109/CVPR52688.2022.00135

Islam, J., & Zhang, Y. (2021). Deep learning-based brain tumor detection and classification from multi-sequence MRI. *Brain Informatics, 8*(1), 1–12. https://doi.org/10.1186/s40708-021-00132-7

Naser, M. A., & Deen, M. J. (2020). Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. *Computers in Biology and Medicine, 121*, 103758. https://doi.org/10.1016/j.compbiomed.2020.103758

Noreen, N., Palaniappan, S., Qayyum, A., Ahmad, I., Imran, M., & Shoaib, M. (2020). A deep learning model based on concatenation approach for the diagnosis of brain tumor. *IEEE Access, 8*, 55135–55144. https://doi.org/10.1109/ACCESS.2020.2978629

Bermudez, C., Plassard, A. J., Davis, L. T., Newton, A. T., Resnick, S. M., & Landman, B. A. (2018). Learning implicit brain MRI foreground from background without ground truth segmentation. *SPIE Medical Imaging: Image Processing, 10574*, 1057413. https://doi.org/10.1117/12.2293751

Zhou, T., Canu, S., & Ruan, S. (2020). An automatic COVID-19 CT segmentation based on U-Net with attention mechanism. *arXiv preprint*. https://doi.org/10.48550/arXiv.2004.06673

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 (MICCAI)*, 12901, 109–119. https://doi.org/10.1007/978-3-030-87193-2_11

Gu, R., Wang, G., Song, T., Huang, R., Aertsen, M., Deprest, J., & Zhang, S. (2021). CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE Transactions on Medical Imaging, 40*(2), 699–711. https://doi.org/10.1109/TMI.2020.3035253

Ahmad, P., Qamar, S., Shen, L., & Saeed, A. (2022). Hybrid deep learning model for efficient brain tumor classification from MRI images. *Biomedical Signal Processing and Control, 74*, 103520. https://doi.org/10.1016/j.bspc.2022.103520

Raza, A., Ayub, H., Khan, J. A., Ahmad, I., Salama, A. S., Damaševičius, R., & Javed, A. R. (2023). A hybrid deep learning-based approach for brain tumor classification and segmentation. *Diagnostics, 13*(1), 49. https://doi.org/10.3390/diagnostics13010049

Saba, T., Mohamed, A. S., El-Affendi, M., Amin, J., & Sharif, M. (2020). Brain tumor detection using fusion of handcrafted and deep learning features. *Cognitive Systems Research, 59*, 221–230. https://doi.org/10.1016/j.cogsys.2019.09.007

Li, Y., Shen, L., & Shi, Y. (2021). BraTS-Net: Multi-modality brain tumor segmentation using encoder-decoder deep networks. *IEEE Access, 9*, 25656–25667.

https://doi.org/10.1109/ACCESS.2021.3056001

Mishra, S., Majhi, B., & Sa, P. K. (2022). Texture feature based classification on microscopic blood smear for acute lymphocytic leukemia detection. *Biomedical Signal Processing and Control, 47*, 303–311. https://doi.org/10.1016/j.bspc.2018.08.012

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR)*. https://doi.org/10.48550/arXiv.2010.11929

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

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

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