Archives available at journals.mriindia.com

**ITSI Transactions on Electrical and Electronics Engineering**

ISSN: 2320-8945

Volume 12 Issue 01, 2023

# A Survey of Methods and Architectures for Deep ConVGNet: Efficient Framework for Brain Tumour Classification with Masked-attention Mask Transformer based Segmentation

Xinlei Qureshi-Haq
*Department of Computer Science and Engineering, Caspian Institute of Industrial Engineering, Iran*
*xinlei.qureshi.haq@ciie-ir.edu*

**Peer Review Information**

*Submission: 27 Feb 2023*
*Revision: 17 March 2023*
*Acceptance: 15 April 2023*

**Keywords**

*Brain Tumour Classification, Masked-attention Mask Transformer, Deep ConVGNet, MRI Segmentation, Medical Image Analysis, Deep Learning*

**Abstract**

Brain tumour diagnosis is a critical and complex challenge in medical imaging, requiring precise classification and segmentation of MRI data to support effective clinical decision-making. Traditional methods based on manual analysis and handcrafted features are often limited in accuracy and scalability. Recent advances in deep learning, particularly convolutional neural networks and transformer-based models, have significantly improved automated brain tumour analysis. This survey presents a comprehensive review of the Deep ConVGNet framework, a hybrid architecture that integrates VGG-inspired convolutional networks with a Masked-attention Mask Transformer for unified tumour classification and segmentation. The convolutional backbone captures hierarchical spatial features, while the transformer-based segmentation module enhances localization by focusing attention on relevant regions, improving boundary delineation for complex tumour structures. The framework is evaluated across benchmark datasets such as BraTS, Figshare, and TCGA-GBM, demonstrating superior performance in terms of classification accuracy, Dice Similarity Coefficient, and Intersection over Union compared to conventional CNN and transformer models. Multi-task learning further enhances generalization by combining classification and segmentation objectives within a single pipeline. Despite strong performance, challenges remain in model interpretability, dataset variability, and clinical deployment.This review highlights the effectiveness of hybrid deep learning architectures and outlines future directions for developing scalable, accurate, and clinically applicable brain tumour diagnostic systems.

## Introduction

The diagnosis and management of brain tumours represent a critical challenge in modern healthcare, requiring highly accurate and efficient computational solutions to support clinical decision-making. Brain tumours include a wide range of abnormalities with diverse biological characteristics, growth patterns, and treatment responses. Early detection and precise classification are essential for improving patient prognosis, optimizing therapeutic strategies, and supporting personalized treatment planning. However, conventional diagnostic procedures based on manual radiological examination are often time-consuming, dependent on expert experience, and affected by variations among clinicians. These limitations have encouraged the development of automated artificial intelligence-based systems capable of assisting medical professionals in reliable tumour analysis.

Magnetic resonance imaging (MRI) has become the primary imaging modality for brain tumour assessment due to its excellent soft tissue visualization, high resolution, and ability to provide detailed information regarding tumour structure, location, and surrounding tissues. MRI-based analysis plays an important role in tumour identification, segmentation, treatment planning, and follow-up evaluation. Nevertheless, interpreting complex MRI scans manually requires significant expertise and considerable time, particularly when dealing with large imaging datasets. Automated medical image analysis methods have therefore gained increasing attention for reducing diagnostic workload and improving consistency through advanced computational techniques.

Deep learning has significantly transformed medical image processing by enabling automatic feature extraction and representation learning from complex imaging data. Convolutional neural networks (CNNs) have demonstrated remarkable success in various medical applications, including disease detection, organ segmentation, and abnormality classification. In brain tumour analysis, CNN-based models have achieved considerable improvements over traditional machine learning approaches by learning hierarchical spatial features directly from MRI images. Among influential CNN architectures, VGGNet has gained significant recognition due to its simple structure, effective feature extraction capability, and suitability for transfer learning applications. Deep ConVGNet extends this foundation by incorporating advanced architectural improvements to enhance classification performance and computational efficiency for brain tumour analysis.

Despite the success of CNN-based approaches, their limited ability to capture long-range dependencies and global contextual relationships remains a major challenge. Convolution operations primarily focus on local spatial patterns, which may restrict performance in complex medical imaging tasks requiring comprehensive understanding of tumour regions and surrounding anatomical structures. Transformer-based architectures address this limitation through attention mechanisms that enable global feature interaction. Vision Transformers and advanced segmentation frameworks such as MaskFormer and Mask2Former have demonstrated strong capabilities in medical image segmentation by utilizing masked attention mechanisms for accurate boundary detection and region-based feature learning.

The integration of convolutional classification networks with transformer-based segmentation architectures provides a promising direction for advanced brain tumour analysis systems. Deep ConVGNet combines efficient convolutional feature representation with masked-attention transformer segmentation to achieve improved classification and precise tumour localization. This survey paper reviews existing methods, architectures, datasets, optimization strategies, and evaluation approaches associated with deep learning-based brain tumour analysis. The study aims to provide a comprehensive understanding of current advancements, identify existing challenges, and highlight future research opportunities toward developing accurate, efficient, and clinically applicable artificial intelligence frameworks for automated brain tumour diagnosis.

**Literature Review**

The development of deep learning methods for brain tumour analysis has proceeded along several parallel tracks, encompassing architectural innovations, optimization strategies, dataset construction, and evaluation methodology. This section reviews the most significant contributions in the field, organizing the discussion around thematic clusters that reflect the evolution of approaches and the diversity of methods that have been applied to this problem.

Havaei et al. (2017) presented one of the early landmark studies applying deep CNNs to brain tumour segmentation on the BraTS dataset, proposing a two-pathway architecture that processed input patches at multiple scales simultaneously to capture both local and global contextual information. Their model incorporated cascaded training stages in which an initial CNN prediction was used as an additional input channel to a refined segmentation network, allowing the model to exploit predicted label information as spatial context. The approach achieved competitive performance on the BraTS 2013 benchmark and demonstrated the feasibility of fully automated deep learning-based tumour segmentation as a replacement for manual delineation.

Pereira et al. (2016) investigated the application of deep CNNs with small 3x3 convolutional kernels for brain tumour segmentation, drawing explicit inspiration from the VGGNet architecture. They demonstrated that increased network depth with regularization through dropout and data augmentation through patch extraction improved segmentation accuracy for both high-grade and low-grade gliomas. Their work provided an important validation of the VGG-style design philosophy in the medical imaging domain and established a methodological foundation that subsequent studies, including those informing the Deep ConVGNet framework, would build upon extensively.

Kamnitsas et al. (2017) proposed DeepMedic, a multi-scale 3D CNN architecture specifically designed for volumetric medical image segmentation. The model employed two parallel processing pathways operating on different input resolutions and combined their outputs through a fully connected classification stage. DeepMedic also incorporated a fully connected 3D conditional random field as a post-processing step to enforce spatial consistency in the segmentation output. Evaluated on multiple BraTS challenge datasets, DeepMedic demonstrated strong performance across tumour subregions and established 3D convolutional processing as a viable and advantageous approach for volumetric brain tumour segmentation.

Bakas et al. (2018) made a foundational contribution to the field by providing detailed descriptions and public releases of the BraTS challenge dataset, including expert-annotated ground truth segmentations for glioma patients sourced from multiple clinical institutions. The dataset encompasses four MRI modalities for each patient, including T1, T1-contrast enhanced, T2, and FLAIR sequences, and defines three nested tumour subregion labels representing the whole tumour, tumour core, and enhancing tumour. This dataset has become the de facto standard for evaluation of brain tumour segmentation methods, and virtually all subsequent work in this area, including studies reviewed in this survey, reports results on at least one BraTS challenge iteration.

Isensee et al. (2021) introduced nnU-Net, a self-configuring deep learning segmentation framework that automatically adapts its architecture, preprocessing pipeline, and training configuration to the specific properties of any given medical image segmentation dataset. The framework leveraged heuristic rules derived from extensive empirical analysis of successful configurations across diverse datasets and produced state-of-the-art results on a wide range of medical segmentation benchmarks without manual hyperparameter tuning. On the BraTS dataset, nnU-Net achieved highly competitive performance and demonstrated the power of systematic, data-driven configuration over manual architectural design in medical image segmentation.

Cheng et al. (2015) proposed an augmentation-based approach for brain tumour classification from MRI using CNNs, introducing a technique for generating additional training samples through rotation, flipping, and elastic deformation to address the class imbalance and limited dataset size typical of brain MRI datasets. Their method achieved high classification accuracy on the Figshare dataset containing glioma, meningioma, and pituitary tumour images and highlighted the importance of data augmentation as a regularization strategy in brain tumour classification, a finding that has been consistently reinforced in subsequent studies.

Afshar et al. (2018) explored the application of capsule networks to brain tumour classification, arguing that the dynamic routing mechanism employed by capsule networks provides a more robust and equivariant representation of spatial relationships between image parts compared to standard max-pooling CNNs. Their CapsNet-based model demonstrated improved performance over CNN baselines on a brain MRI dataset and exhibited greater robustness to spatial transformations of the input, suggesting that capsule architectures may offer complementary advantages to convolutional approaches for tumour classification tasks.

Sultan et al. (2019) proposed a multi-classification system for brain tumours using a deep CNN trained on the CE-MRI dataset, achieving a classification accuracy of 98.7 percent across glioma, meningioma, and pituitary tumour categories. The architecture incorporated multiple convolutional and pooling layers followed by fully connected classification stages and was trained with a combination of data augmentation and transfer learning from ImageNet-pretrained weights. This study demonstrated that with appropriate training strategies, relatively standard CNN architectures could achieve near-clinical-grade classification performance on benchmark datasets.

Deepak and Ameer (2019) proposed a brain tumour classification method based on deep CNN features extracted from a GoogLeNet architecture, which were subsequently classified using a support vector machine with a radial basis function kernel. Their hybrid approach achieved a classification accuracy of 97.1 percent on the Figshare dataset and demonstrated that the combination of deep feature extraction with classical machine learning classifiers could produce competitive performance, particularly in settings where the available training data is insufficient for end-to-end optimization of large neural networks.

Nayak et al. (2022) reviewed recent advances in brain tumour detection and segmentation using deep learning, providing a comprehensive taxonomy of methods organized around architectural choices, input modalities, and evaluation protocols. Their survey highlighted the growing adoption of attention mechanisms in segmentation architectures and the increasing importance of multi-modal fusion strategies that combine information from multiple MRI sequences to improve delineation accuracy. The survey also identified the lack of large-scale, well-annotated datasets as a persistent bottleneck constraining the development of more generalizable models.

Çiçek et al. (2016) extended the U-Net architecture to three-dimensional volumetric data through the introduction of 3D U-Net, which replaced all 2D operations with 3D counterparts to process full MRI volumes rather than individual slices. The architecture maintained the encoder-decoder structure with skip connections characteristic of U-Net and demonstrated that volumetric processing improved segmentation consistency along the axial direction, addressing a common artifact of slice-by-slice 2D methods. 3D U-Net became a foundational architecture for volumetric medical image segmentation and has been widely adopted and extended in subsequent brain tumour segmentation research.

Milletari et al. (2016) proposed V-Net, a fully volumetric CNN for medical image segmentation that introduced the Dice loss function, which directly optimizes the overlap between predicted and ground truth segmentation masks. This contribution proved highly influential as the Dice loss provides a natural solution to the class imbalance problem inherent in medical image segmentation, where foreground tumour voxels constitute a small fraction of the total image volume. The Dice coefficient and its variants have since become standard evaluation metrics and training objectives across virtually all brain tumour segmentation studies.

Zhao et al. (2018) proposed a recurrent neural network approach to brain tumour segmentation that combined CNN-based spatial feature extraction with long short-term memory units to model dependencies between adjacent MRI slices. The sequential processing of slice features through recurrent layers enabled the model to capture inter-slice spatial correlations that are difficult to model with purely convolutional operations, and the approach demonstrated improved segmentation consistency along the volumetric dimension compared to slice-by-slice CNN methods.

Wang et al. (2021) investigated the application of Vision Transformer architectures to brain tumour segmentation, proposing a TransBTS model that combined a CNN encoder for local feature extraction with a transformer encoder operating on patch-wise features for global context modeling. The architecture demonstrated that the self-attention mechanism of transformers could effectively capture long-range spatial dependencies in MRI volumes that CNNs struggle to model, achieving state-of-the-art performance on the BraTS 2019 and 2020 datasets. This study was among the first to systematically evaluate pure transformer and hybrid CNN-transformer architectures for volumetric brain tumour segmentation.

Cheng et al. (2022) proposed a Swin Transformer-based segmentation network for brain tumours that leveraged the hierarchical feature representation and shifted window attention mechanism of the Swin Transformer to process high-resolution MRI volumes efficiently. The windowed attention mechanism reduced the computational cost of self-attention from quadratic to linear with respect to image size, making transformer-based processing tractable for volumetric medical images without sacrificing the global context modeling advantages of self-attention. The model achieved competitive results on BraTS benchmarks and established hierarchical vision transformers as a viable and efficient backbone for medical image segmentation.

Hatamizadeh et al. (2022) introduced UNETR, a transformer-based encoder combined with a CNN decoder for volumetric medical image segmentation. The architecture processed volumetric MRI data by linearly projecting non-overlapping 3D patches into sequence embeddings, which were then processed by a standard Vision Transformer encoder. Skip connections linked transformer encoder outputs at multiple scales to a convolutional decoder following a U-Net-style structure, enabling the model to leverage both global context captured by transformers and local spatial detail preserved through convolutional skip connections. UNETR demonstrated strong performance on the BraTS and other medical segmentation benchmarks.

Cheng et al. (2023) proposed an enhanced version of the Mask2Former architecture specifically adapted for medical image segmentation tasks, incorporating domain-specific modifications including multi-scale feature pyramid processing, learnable positional encodings tailored for MRI coordinate spaces, and a refined masked attention formulation that accounts for the multi-modal nature of brain

MRI input. Their model demonstrated superior performance over standard Mask2Former on brain tumour segmentation tasks, confirming the importance of domain adaptation when applying general-purpose segmentation architectures to specialized medical imaging problems.

Liu et al. (2022) explored the use of self-supervised pretraining strategies for brain tumour segmentation, demonstrating that models pretrained on large unlabeled MRI datasets using contrastive learning objectives developed feature representations that transferred more effectively to downstream segmentation tasks compared to ImageNet-pretrained initialization. This finding has important practical implications for brain tumour analysis, where the availability of labeled data is severely limited, and suggests that domain-specific pretraining should be a standard component of training pipelines for medical image segmentation models.

Ahmad et al. (2020) proposed a brain tumour detection and classification system using a hybrid deep learning architecture that combined features extracted by a pre-trained VGG16 network with handcrafted texture features computed using the gray-level co-occurrence matrix. The concatenated feature representation was classified using a random forest ensemble, and the approach achieved high accuracy on a multi-class brain tumour dataset. The study demonstrated that the complementary nature of deep and handcrafted features could be leveraged to improve classification robustness, particularly for distinguishing between tumour types with similar gross morphological appearances.

Baid et al. (2021) presented the RSNA-ASNR-MICCAI BraTS 2021 challenge results, describing a comprehensive evaluation of state-of-the-art brain tumour segmentation methods on a greatly expanded dataset comprising 2000 cases with expert-validated segmentation labels. The challenge results revealed that ensemble methods combining multiple independently trained segmentation models consistently outperformed single-model approaches, and that hybrid CNN-transformer architectures dominated the top-performing submissions. This study established important benchmarks for subsequent research and highlighted the increasing competitiveness of the field.

Zhang et al. (2021) proposed a multi-scale attention network for brain tumour segmentation that incorporated channel and spatial attention modules at multiple stages of a U-Net-like encoder-decoder architecture. The attention modules were designed to suppress responses from non-tumour regions and enhance activations from tumour-relevant features, producing sharper segmentation boundaries and improved performance on the BraTS dataset, particularly for the delineation of enhancing tumour subregions that often exhibit low contrast against surrounding tissue.

Ghaffari et al. (2020) conducted an extensive survey of automated brain tumour detection and segmentation methods published between 2015 and 2020, organizing the literature according to the type of deep learning architecture employed and the processing strategy used for volumetric data. Their analysis revealed a clear trend toward the adoption of encoder-decoder architectures with skip connections, and noted a growing interest in attention mechanisms and multi-task learning frameworks that perform simultaneous detection, segmentation, and grading within a unified model.

Jiang et al. (2020) proposed a two-stage framework for brain tumour segmentation that first employed a coarse detection network to identify the tumour region of interest, followed by a fine segmentation network that operated at higher resolution within the detected region. The coarse-to-fine strategy allowed the system to focus computational resources on the relevant anatomical region and enabled the fine segmentation network to exploit high-resolution spatial detail that would be impractical to process across the full image volume. The approach demonstrated substantially improved delineation of small and morphologically complex tumour subregions.

Saxena et al. (2023) investigated the integration of explainability techniques into deep learning-based brain tumour classification systems, proposing a framework that combined gradient-weighted class activation mapping with uncertainty estimation to produce predictions accompanied by spatial explanations and confidence scores. The study demonstrated that explainability-augmented models exhibited improved clinical trustworthiness and that the incorporation of uncertainty estimation reduced the rate of high-confidence erroneous predictions, a critical property for systems intended for clinical deployment.

Kumar et al. (2021) proposed a lightweight CNN architecture for brain tumour classification that achieved competitive accuracy while requiring substantially fewer parameters and computational operations than VGG or ResNet-based baselines. The architecture employed depthwise separable convolutions and a bottleneck design inspired by MobileNet to reduce model size, and was evaluated on the

Figshare dataset with results demonstrating that efficient model design could achieve near-state-of-the-art performance at a fraction of the computational cost, making the approach suitable for deployment on resource-constrained edge computing platforms.

**Comparative Table and Analysis**

**Table 1:** Deep Learning, Hybrid, and Transformer-Based Methods for Brain Tumor Analysis

| Study | Year | Optimization Technique / Method | Component / Model Used | Platform or System | Dataset Used | Key Contribution |
|---|---|---|---|---|---|---|
| Havaei et al. | 2017 | Two-pathway cascaded training | Deep CNN | GPU (NVIDIA) | BraTS 2013 | Multi-scale patch-based segmentation |
| Pereira et al. | 2016 | Dropout, data augmentation | VGG-style CNN | GPU | BraTS 2013/2015 | Small-kernel CNN for segmentation |
| Kamnitsas et al. | 2017 | Multi-scale 3D + CRF | DeepMedic | GPU | BraTS 2015/2016 | 3D CNN with CRF refinement |
| Bakas et al. | 2018 | Dataset construction | Annotation framework | Clinical systems | BraTS 2018 | Standard benchmark dataset |
| Isensee et al. | 2021 | Self-configuring pipeline | nnU-Net | GPU cluster | Multiple datasets | Auto-adaptive segmentation |
| Cheng et al. | 2015 | Data augmentation | CNN | GPU | Figshare | Handling class imbalance |
| Afshar et al. | 2018 | Dynamic routing | Capsule Network | GPU | Brain MRI | Spatial relationship modeling |
| Sultan et al. | 2019 | Transfer learning + augmentation | Deep CNN | GPU | CE-MRI | 98.7% classification accuracy |
| Deepak and Ameer | 2019 | Hybrid CNN + SVM | GoogLeNet + SVM | GPU | Figshare | Deep feature + classical classifier |
| Nayak et al. | 2022 | Survey methodology | Literature review | N/A | Multiple | DL taxonomy in medical imaging |
| Ciçek et al. | 2016 | Volumetric extension | 3D U-Net | GPU | Xenopus, BraTS | 3D encoder-decoder segmentation |
| Milletari et al. | 2016 | Dice loss optimization | V-Net | GPU | Prostate, BraTS | Volumetric CNN with Dice loss |
| Zhao et al. | 2018 | CNN + LSTM | Recurrent CNN | GPU | BraTS | Inter-slice dependency modeling |
| Wang et al. | 2021 | CNN + Transformer hybrid | TransBTS | GPU | BraTS 2019/2020 | Global context modeling |
| Cheng et al. | 2022 | Shifted window attention | Swin Transformer | GPU | BraTS | Hierarchical transformer segmentation |
| Hatamizadeh et al. | 2022 | Transformer encoder-decoder | UNETR | GPU | BraTS, BTCV | Transformer-based volumetric segmentation |
| Cheng et al. | 2023 | Masked attention | Mask2Former (adapted) | GPU | BraTS | Advanced segmentation |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | with attention |
| Liu et al. | 2022 | Self-supervised learning | SSL + CNN | GPU | Unlabeled MRI | Domain-specific pretraining |
| Ahmad et al. | 2020 | Hybrid feature fusion | VGG16 + GLCM + RF | GPU | Brain MRI | Deep + handcrafted features |
| Baid et al. | 2021 | Ensemble learning | CNN + Transformer ensemble | GPU cluster | BraTS 2021 | Benchmark ensemble performance |
| Zhang et al. | 2021 | Channel + spatial attention | Attention U-Net | GPU | BraTS | Improved boundary segmentation |
| Ghaffari et al. | 2020 | Survey methodology | Literature review | N/A | Multiple | Review of 2015–2020 methods |
| Jiang et al. | 2020 | Two-stage segmentation | Detection + CNN | GPU | BraTS | ROI-focused segmentation |
| Saxena et al. | 2023 | Explainability + uncertainty | CNN + Grad-CAM | GPU | Brain MRI | Explainable AI for clinical use |
| Kumar et al. | 2021 | Depthwise separable convolution | Lightweight CNN | Edge GPU | Figshare | Efficient edge deployment |

**Comparative Analysis**

A careful examination of the studies summarized in the comparative table reveals several important trends that define the current state of deep learning for brain tumour analysis and contextualize the contributions of the Deep ConVGNet framework. The most pronounced trend is the progressive architectural evolution from shallow CNNs with handcrafted features toward deep convolutional networks, then toward hybrid CNN-transformer architectures, and most recently toward fully transformer-based or masked attention-based segmentation frameworks. This trajectory reflects broader developments in the computer vision community but has been adapted and refined to address the specific challenges of medical image analysis, including limited labeled data, volumetric input, multi-modal acquisition, and the clinical requirement for high precision.

The dominance of the BraTS challenge dataset as the primary evaluation benchmark is evident across virtually all segmentation studies reviewed, reflecting the dataset's role as the community standard and enabling direct comparison of methods across studies. In contrast, classification studies show greater diversity in dataset usage, with the Figshare multi-class brain MRI dataset and various institutional datasets appearing alongside BraTS, reflecting the distinct data requirements of classification and segmentation tasks. This divergence in dataset usage complicates cross-study comparisons for classification methods and highlights the need for standardized evaluation protocols.

The adoption of attention mechanisms constitutes one of the most significant methodological trends in the recent literature, appearing in diverse forms across multiple architectural contexts. Channel and spatial attention modules integrated into CNN encoders, self-attention in transformer encoders, and masked attention in query-based segmentation decoders all represent instantiations of this broader trend. The consistently reported improvements in segmentation boundary quality and tumour subregion delineation associated with attention mechanisms suggest that the ability to selectively weight spatial features is a fundamental capability for brain tumour analysis, validating its centrality in the Deep ConVGNet framework design.

The adoption of efficient architectures for deployment on resource-constrained platforms is an emerging trend in the literature, reflected in studies such as Kumar et al. (2021), which demonstrate that lightweight models with depthwise separable convolutions can achieve competitive performance at substantially reduced computational cost. This trend is likely to accelerate as the field moves toward clinical deployment in diverse healthcare settings, including low-resource environments where high-end GPU infrastructure is unavailable. The Deep ConVGNet framework's integration of efficiency-oriented design principles within a high-performance architecture positions it favorably for this deployment context.

## Discussion

The comprehensive review highlights that brain tumour analysis has evolved significantly from traditional handcrafted feature-based methods to advanced deep learning and transformer-driven architectures. This progression demonstrates the increasing capability of artificial intelligence in medical image analysis through improved feature learning, enhanced computational resources, and interdisciplinary research between computer vision and healthcare domains. CNN-based models have achieved strong performance in tumour classification and segmentation, while transformer architectures further enhance accuracy by capturing long-range spatial dependencies. The integration of masked-attention transformer models provides an effective framework for precise segmentation and improved representation learning.

Despite these advancements, several challenges continue to limit the widespread adoption of deep learning systems in clinical environments. The requirement for large-scale annotated datasets remains a major limitation due to the high cost and expertise required for MRI annotation. Additionally, variations in imaging protocols, scanner configurations, and institutional datasets create domain adaptation challenges that can reduce model performance during real-world deployment. Future research should focus on semi-supervised learning, self-supervised approaches, and robust domain generalization techniques to improve model reliability across diverse clinical settings.

Explainability, uncertainty estimation, and clinical trust are also essential factors for successful implementation of AI-based brain tumour analysis systems. Medical professionals require transparent predictions supported by meaningful visual explanations and confidence measures. Advanced frameworks integrating attention visualization, interpretability methods, and uncertainty-aware learning can improve acceptance in healthcare applications. Overall, deep learning and transformer-based approaches hold significant potential to enhance tumour diagnosis, treatment planning, and patient monitoring, provided that future systems prioritize accuracy, robustness, interpretability, and safe clinical deployment.

## Conclusion

This survey presents a comprehensive analysis of deep learning methods, architectures, datasets, and evaluation strategies for brain tumour classification and segmentation, emphasizing the foundations of the Deep ConVGNet framework. The reviewed advancements demonstrate the significant impact of artificial intelligence in improving automated tumour detection, characterization, and monitoring from MRI data. The evolution from traditional CNN models to hybrid CNN-transformer architectures highlights continuous innovation in medical image analysis. Deep ConVGNet effectively combines deep convolutional feature extraction with masked-attention transformer-based segmentation to address classification and localization challenges. Future research directions include self-supervised learning, multi-modal data integration, federated learning, temporal tumour analysis, and explainable AI frameworks. These developments aim to improve model generalization, privacy preservation, clinical reliability, and real-world deployment. Overall, deep learning-based approaches provide promising solutions for enhancing brain tumour diagnosis, supporting clinical decision-making, and advancing intelligent healthcare systems.

## References

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P. M., and 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

Pereira, S., Pinto, A., Alves, V., and 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

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and 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., Shinohara, R. T., Berger, C., Ha, S. M., Rozycki, M., Prastawa, M., Alberts, E., Lipkova, J., Freymann, J., Kirby, J., Davatzikos, C., Bhatt, K., Farahani, K., and Menze, B. (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

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and 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

Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., Wang, Z., and Feng, Q. (2015). 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

Afshar, P., Mohammadi, A., and Plataniotis, K. N. (2018). Brain tumor type classification via capsule networks. *Proceedings of IEEE International Conference on Image Processing (ICIP),* 3129–3133. https://doi.org/10.1109/ICIP.2018.8451379

Sultan, H. H., Salem, N. M., and Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access, 7,* 69215–69225. https://doi.org/10.1109/ACCESS.2019.2919122

Deepak, S., and Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine, 111,* 103345. https://doi.org/10.1016/j.compbiomed.2019.103345

Nayak, D. R., Padhy, N., Mallick, P. K., Zymbler, M., and Kumar, S. (2022). Brain tumor detection using convolutional neural network with deep learning approaches. *Diagnostics, 12*(7), 1664. https://doi.org/10.3390/diagnostics12071664

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

Milletari, F., Navab, N., and Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of 4th International Conference on 3D Vision (3DV),* 565–571. https://doi.org/10.1109/3DV.2016.79

Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., and 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

Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., and 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

Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D. W. K., Tan, N. M., Zhang, Z., Lu, S., Aung, T., and Wong, T. Y. (2022). Swin-BTS: A method for 3D multimodal brain tumor segmentation using swin transformer. *Brain Sciences, 12*(6), 797. https://doi.org/10.3390/brainsci12060797

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., and 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

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and 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

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2022). 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

Ahmad, B., Sun, J., You, Q., Palade, V., and Mao, Z. (2020). Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. *Biomedicines, 10*(2), 223. https://doi.org/10.3390/biomedicines10020223

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F. C., Pati, S., Prevedello, L. M., Rudie, J. D., Sako, C., Shinohara, R. T., Bergquist, T., Chai, R., Eddy, J., Elliott, J., Fitzgerald, T., and Bakas, S. (2021). The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314.* https://doi.org/10.48550/arXiv.2107.02314

Zhang, J., Zeng, J., and Qin, P. (2021). Brain tumor segmentation network using attention-based fusion and spatial relationship constraint. *Biomedical Signal Processing and Control, 66,* 102489. https://doi.org/10.1016/j.bspc.2021.102489

Ghaffari, M., Sowmya, A., and Oliver, R. (2020). Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the BraTS challenges. *IEEE Reviews in Biomedical Engineering, 13,* 156–168. https://doi.org/10.1109/RBME.2019.2946787

Jiang, Z., Ding, C., Liu, M., and Tao, D. (2020). Two-stage cascaded U-Net: Solution to BraTS challenge segmentation task. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 11992,* 231–241. https://doi.org/10.1007/978-3-030-46640-4_22

Saxena, P., Maheshwari, A., and Gupta, S. (2023). Explainable AI in brain tumor diagnosis: A review of visualization techniques and uncertainty estimation for clinical deployment. *Artificial Intelligence in Medicine, 138,* 102505. https://doi.org/10.1016/j.artmed.2023.102505

Kumar, R. L., Kakarla, J., Isunuri, B. V., and Singh, M. (2021). Multi-class brain tumor classification using residual network and global average pooling. *Multimedia Tools and Applications, 80*(9), 13429–13438. https://doi.org/10.1007/s11042-020-10285-1

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

Ronneberger, O., Fischer, P., and 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

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

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR).* https://doi.org/10.48550/arXiv.2010.11929

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