https://www.scirp.org/journal/jbise
J. Biomedical Science and Engineering,
2025, Vol. 18, (No. 1), pp: 1-32

From Traditional Methods to 3D U-Net: A Comprehensive Review of Brain Tumour Segmentation Techniques

Mushtaq Mahyoob Saleh, Musab Elkheir Salih, Mohamed A. A. Ahmed, Altahir Mohamed Hussein

Biomedical Engineering Department, College of Engineering, Sudan University of Science and Technology, Khartoum, Sudan

Correspondence to: Mushtaq Mahyoob Saleh, mushtaqbme2022@gmail.com, mmsbme2@gmail.com
Keywords: Brain Tumour, MRI Modalities, Deep Learning, 3D U-Net, BraTS
Received: December 12, 2024 **Accepted:** January 28, 2025 **Published:** January 31, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

ABSTRACT

Accurate brain tumour segmentation is critical for diagnosis and treatment planning, yet challenging due to tumour complexity. Manual segmentation is time-consuming and variable, necessitating automated methods. Deep learning, particularly 3D U-Net architectures, has revolutionised medical image analysis by leveraging volumetric data to capture spatial context, enhancing segmentation accuracy. This paper reviews brain tumour segmentation methods, emphasising 3D U-Net advancements. We analyse contributions from the Brain Tumour Segmentation (BraTS) challenges (2014-2023), highlighting key improvements and persistent challenges, including tumour heterogeneity, limited annotated data, varied imaging protocols, computational constraints, and model generalisation. Unlike previous reviews, we synthesise these challenges, proposing targeted research directions: enhancing model robustness through domain adaptation and multi-institutional data sharing, developing lightweight architectures for clinical deployment, integrating multi-modal and clinical data, and incorporating explainability techniques to build clinician trust. By addressing these challenges, we aim to guide future research toward developing more robust, generalisable, and clinically applicable segmentation models, ultimately improving patient outcomes in neuro-oncology.

1. INTRODUCTION

Brain tumours are among the most dangerous diseases globally, significantly contributing to cancer-related morbidity and mortality. The brain's complex structure and vital nervous system functions make tumours in any part of the brain or skull particularly challenging to diagnose and treat [1]. Gliomas are the most common type of brain tumour, accounting for approximately 78% of malignant brain tumours. They are classified into two categories: Low-Grade Gliomas (LGG), which are less malignant and grow slowly,

and High-Grade Gliomas (HGG), which are highly malignant and proliferate rapidly [2].

Brain tumours can be primary or metastatic (secondary). Primary brain tumours originate in the brain tissue or surrounding areas, consisting of glial and non-glial cells that grow in blood vessels, nerves, and glands. These tumours can be benign or malignant and typically do not spread to other body parts [3]. Benign tumours lack malignant cells but can still impact critical brain functions due to their size and location. Malignant tumours, often referred to as brain cancer, are more dangerous because they can invade nearby brain tissue [4, 5].

The World Health Organization (WHO) classifies brain tumours into four grades based on their histological characteristics and growth rate: Grade I tumours are regular in shape and grow slowly; Grade II tumours appear abnormal and also grow slowly; Grade III tumours grow more rapidly than Grade I and II tumours; and Grade IV tumours proliferate quickly and are highly malignant. Additionally, brain tumours progress through five stages, from stage zero to stage four, with stage four representing the most advanced and aggressive form [6].

Early and accurate diagnosis of brain tumours is crucial for effective treatment planning and improving patient survival rates. Magnetic Resonance Imaging (MRI) is a superior and widely used diagnostic tool for brain imaging, playing a vital role in the identification and localisation of brain tumours due to its high soft tissue contrast and ability to provide detailed anatomical information without ionising radiation [7] MRI typically utilises four input modalities: T1-weighted images without contrast, T1-weighted images with contrast enhancement (T1ce), T2-weighted images, and Fluid-Attenuated Inversion Recovery (FLAIR) images [8, 9]. Each modality provides unique insights into tumour characteristics, aiding in comprehensive assessment and diagnosis [2].

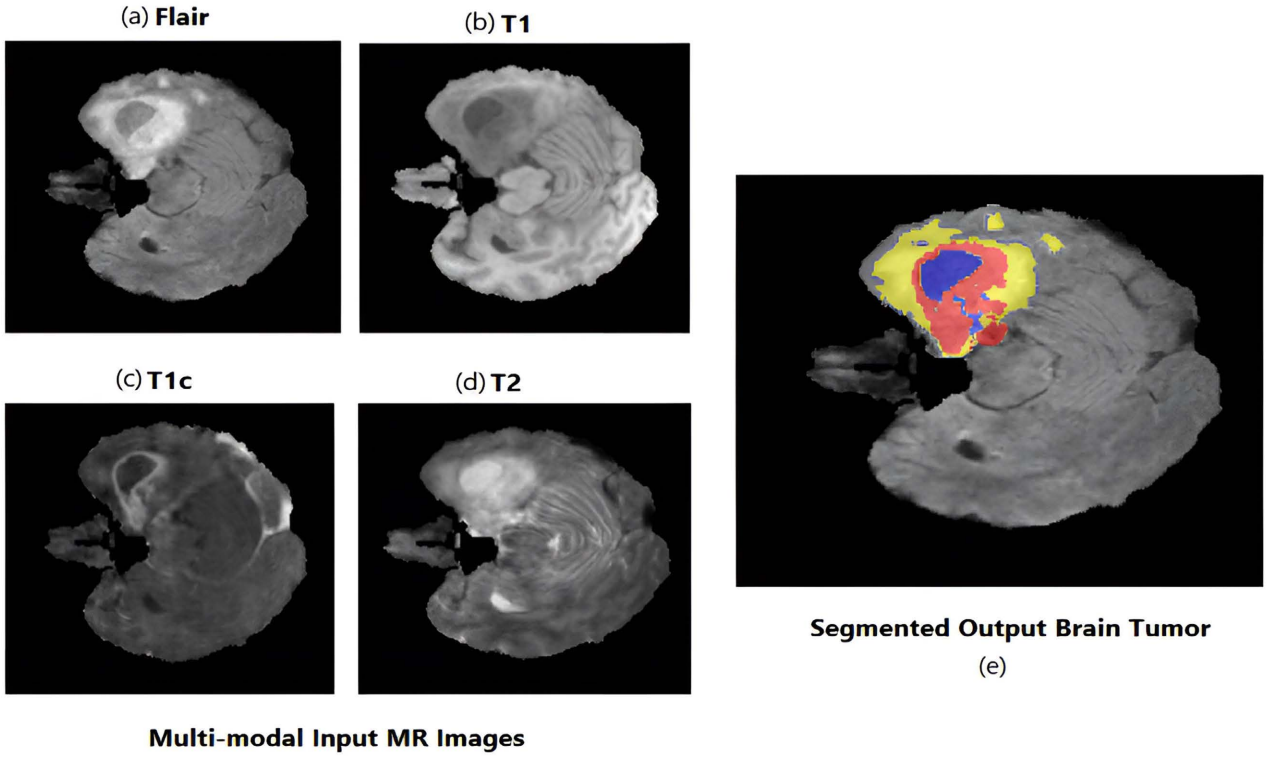

Figure 1. MRI modalities and corresponding tumour regions: (a) Flair image, (b) T1-weighted image without contrast, (c) T1c image with contrast enhancement, (d) T2 image, and (e) Final labels of combined tumor segmentations. Colour overlays denote tumour regions—yellow represents the whole tumour (Edema), brown indicates the enhancing tumour, and blue shows the tumour core (necrosis).

Figure 1 illustrates different MRI modalities used in brain tumour imaging, highlighting tumour regions identified in each modality. The colour overlays denote various tumour components: yellow represents the whole tumour, brown indicates enhancing tumour regions, blue shows tumour core areas. Each MRI modality contributes unique insights into the tumour's structure, essential for comprehensive segmentation and diagnosis. The segmented image shows different tumour regions, such as peritumoral edema (typically bright on FLAIR), enhancing Tumour (Bright on T1c), and necrotic and non-enhancing tumour core (dark on T1c, bright on T2).

Manual segmentation of brain tumours from MRI images poses significant challenges due to the extensive volumetric data, intensity variations, low contrast, and the presence of noise. This process is not only time-consuming but also prone to errors and inter-observer variability [10]. Traditional image segmentation techniques, such as thresholding, edge detection, and region-based methods, often fall short in handling the heterogeneity and complex structures of brain tumour [11, 12].

The advent of artificial intelligence, particularly deep learning techniques, has transformed medical image analysis and processing. Convolutional Neural Networks (CNNs) have shown remarkable success by efficiently extracting relevant features from imaging data and automating the segmentation process [13, 14]. Various popular deep learning (DL) architectures have been applied in medical imaging, including AlexNet [14], VGGNet [15], InceptionNet [16], XceptionNet [17], U-Net [18], ResNet [19], and DenseNet [20]. U-Net Deep learning architectures have demonstrated exceptional performance in biomedical image segmentation tasks due to their ability to capture both local and global context through their encoder-decoder structure with skip connections [18]. The 3D U-Net, an extension of the original 2D U-Net architecture, is specifically designed to process volumetric data, capturing spatial context across multiple slices and improving segmentation accuracy for 3D medical images [21].

Despite significant advancements, several challenges persist in accurately segmenting brain tumours using deep learning techniques. Issues such as tumour heterogeneity, class imbalance in datasets, limited sample sizes, and variations in imaging protocols across different institutions hinder the generalisation of models [22]. Additionally, the computational complexity and memory requirements of training deep learning models on large 3D datasets pose practical constraints, limiting their applicability in resource-constrained environments [23].

Accurate tumour segmentation is not only essential for diagnosis but also forms the foundation for downstream tasks, such as survival prediction. Survival prediction in brain tumour segmentation involves using machine learning techniques, such as Random Forest, to estimate patient survival time based on features extracted from brain tumour regions. These features typically include tumour volumes (e.g., whole tumour, enhancing tumour, and tumour core), shape descriptors (e.g., surface area, sphericity, and compactness), and intensity-based metrics derived from multimodal MRI sequences like T1, T2, and FLAIR. Random Forest operates as a regression model for survival analysis, where it combines decision trees to handle non-linear relationships and interactions between features effectively. Its ensemble approach makes it robust to overfitting, and its ability to rank feature importance provides insights into the most critical predictors of survival. While Random Forest models are relatively easy to train and interpret, their performance heavily relies on the quality of tumour segmentation and the richness of the extracted features, making preprocessing a crucial step in the pipeline.

This review aims to provide a comprehensive overview of current deep learning techniques for brain tumour segmentation, focusing on the advancements and challenges associated with 3D U-Net architectures. We critically analyse the contributions and limitations of state-of-the-art methods, particularly those from the Brain Tumour Segmentation (BraTS) challenges from 2014 to 2023. By synthesising existing knowledge and identifying unresolved issues, we seek to guide future research toward developing more robust, generalizable, and clinically applicable segmentation models. Our unique contribution lies in highlighting the gaps in current methodologies and proposing potential pathways for improvement, including enhancing model robustness, improving cross-domain generalisation, and integrating multi-modal data for comprehensive tumour analysis.

2. MATERIALS AND METHODS

2.1. Review Methodology

This review aims to comprehensively analyse the current state of brain tumour segmentation using deep learning techniques, with a particular focus on 3D U-Net architectures. To ensure a systematic and thorough examination of the literature, a structured search was conducted using electronic databases, including PubMed, IEEE Xplore, Science Direct, and Google Scholar.

Search Strategy: The search terms used included "brain tumour segmentation," "deep learning," "3D U-Net," "MRI," "BraTS challenge," "glioma segmentation," "convolutional neural networks," and "medical image analysis." Studies published between 2014 and October 2023 were considered to capture the most recent advancements, especially those related to the BraTS challenges [24]. Only articles published in English were included.

Inclusion Criteria: The review included peer-reviewed journal articles and conference papers that focused on brain tumour segmentation using deep learning techniques. Research involving 3D U-Net architectures or their variants was prioritised. Papers discussing the BraTS challenges and their contributions and limitations from 2014 to 2023 were also included.

Exclusion Criteria: Studies not related to brain tumour segmentation or those not utilising deep learning methods were excluded. Non-English publications were also excluded.

Data Extraction and Synthesis: Relevant information from the selected studies was extracted, including the proposed methods, datasets used, performance metrics, and key findings. Emphasis was placed on studies that provided critical insights into the advancements, challenges, and future directions of 3D U-Net architectures in brain tumour segmentation.

2.2. Image Segmentation Technique

Image segmentation is a crucial step in medical image analysis, involving the partitioning of an image into meaningful regions to facilitate interpretation and diagnosis. In the context of brain tumour segmentation, the goal is to accurately delineate tumour tissue from healthy brain tissue in MRI scans.

Traditional Segmentation Methods

While traditional segmentation methods like edge detection [11, 12], region-based segmentation [25], thresholding [26], and clustering [27] have been applied to brain tumour segmentation, they often fall short when confronted with the inherent complexity and heterogeneity of these tumours. These methods rely on identifying specific image features to delineate tumour boundaries, but they face several challenges in the context of brain tumour analysis.

Edge-based methods [11, 12] attempt to delineate tumour boundaries by detecting discontinuities in image intensity using operators such as Gradient, Sobel, Prewitt, and Canny. However, the boundaries of brain tumours are often ill-defined and diffuse, making it difficult for these methods to accurately capture the tumour's true extent. Furthermore, noise and artefacts in magnetic resonance imaging (MRI) scans can lead to the detection of spurious edges, reducing the reliability of these techniques [12].

Region-based segmentation groups pixels with similar characteristics like intensity, shape, and texture. The underlying assumption of pixel homogeneity within regions can be violated in brain tumours, which often exhibit heterogeneous internal structures. This heterogeneity leads to under-segmentation or over-segmentation, as the algorithm may fail to separate distinct tumour subregions or may erroneously merge tumour regions with surrounding healthy tissue.

Thresholding-based methods [25], such as Otsu's method [28], classify pixels as tumour or non-tumour based on predefined intensity thresholds. However, the intensity distributions of tumour and healthy tissues often overlap significantly, making it challenging to determine a single optimal threshold [25]. This overlap can result in inaccurate segmentation, especially in cases of low-contrast tumours or those with intensity profiles similar to surrounding structures.

Clustering-based segmentation [29] employs algorithms like K-means to group pixels into clusters

based on feature similarity. While effective in some contexts, K-means clustering assumes spherical and equally sized clusters, which is often not the case for irregularly shaped brain tumours [27]. Moreover, the presence of noise and the variability in tumour appearance across different patients can compromise the accuracy of clustering-based approaches [29].

In summary, traditional segmentation methods often struggle to accurately delineate brain tumours due to the latter's irregular shapes, heterogeneous intensity profiles, and indistinct boundaries. These limitations highlight the need for more sophisticated techniques, such as deep learning, which can learn complex patterns and adapt to the variability inherent in brain tumour imaging data.

2.3. Deep Learning for Brain Tumour Segmentation

Deep learning, a subfield of machine learning based on artificial neural networks with representation learning, has emerged as a powerful tool in medical image analysis [30-32]. Convolutional Neural Networks (CNNs), a specialised deep learning architecture, have been particularly successful in image-related tasks due to their ability to automatically learn hierarchical features directly from raw data [33-35]. This section will focus on CNNs and their application in the context of brain tumour segmentation, highlighting key architectural components, challenges, and evaluation metrics.

2.3.1. Convolutional Neural Networks (CNNs)

CNNs are designed to process data with a grid-like topology, making them particularly well-suited for image analysis. They learn spatial hierarchies of features automatically and adaptively, starting from low-level features like edges and textures to higher-level features representing complex patterns. **Figure 2** illustrates a typical CNN architecture used for brain tumour segmentation (see **Figure 2**). The input MRI image undergoes a series of operations, starting with convolutional layers that extract spatial features using learned filters. These features are then downsampled using pooling layers, reducing dimensionality while retaining essential information. Activation functions introduce non-linearity, enabling the network to learn complex patterns. This process is repeated through multiple layers, progressively extracting higher-level features. Finally, fully connected layers integrate these features to produce a segmentation map, delineating tumour regions from healthy tissue.

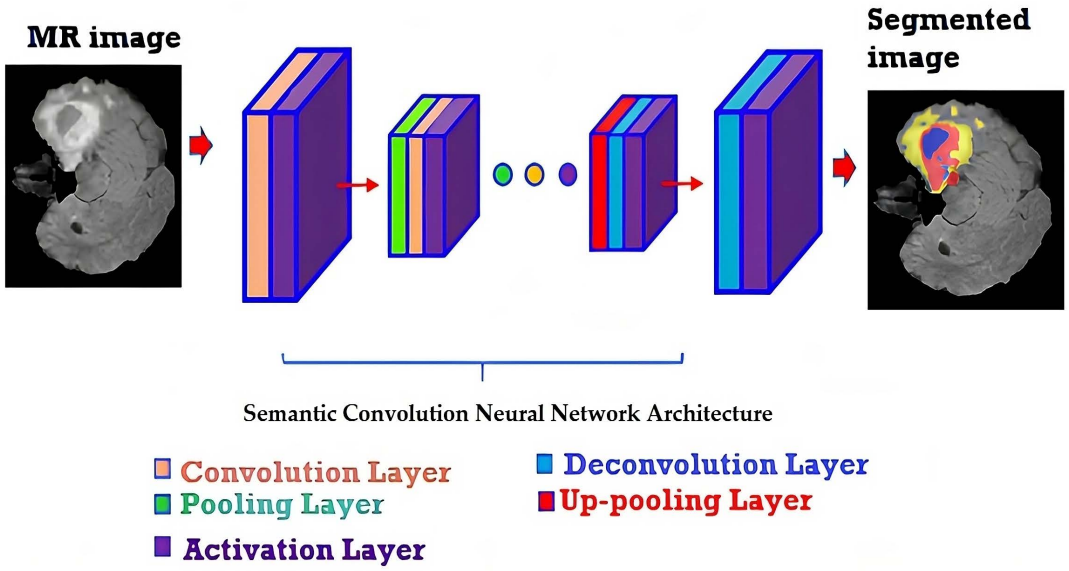

Figure 2. A CNN-based model for brain tumour segmentation. The input MRI image is processed through convolutional layers, pooling layers, and activation functions, resulting in a segmented output that highlights the tumour region separated from healthy tissues.

1) Convolution Operations

Convolutional layers are the core building blocks of CNNs. They apply learned filters to the input image, extracting features such as edges, textures, and patterns (see **Figure 3**). Each filter slides across each channel of the input (e.g., red, green, and blue channels in an RGB image), performing element-wise multiplication and summation to produce a single value in the output feature map. The collection of feature maps generated by different filters represents the learned features at that layer.

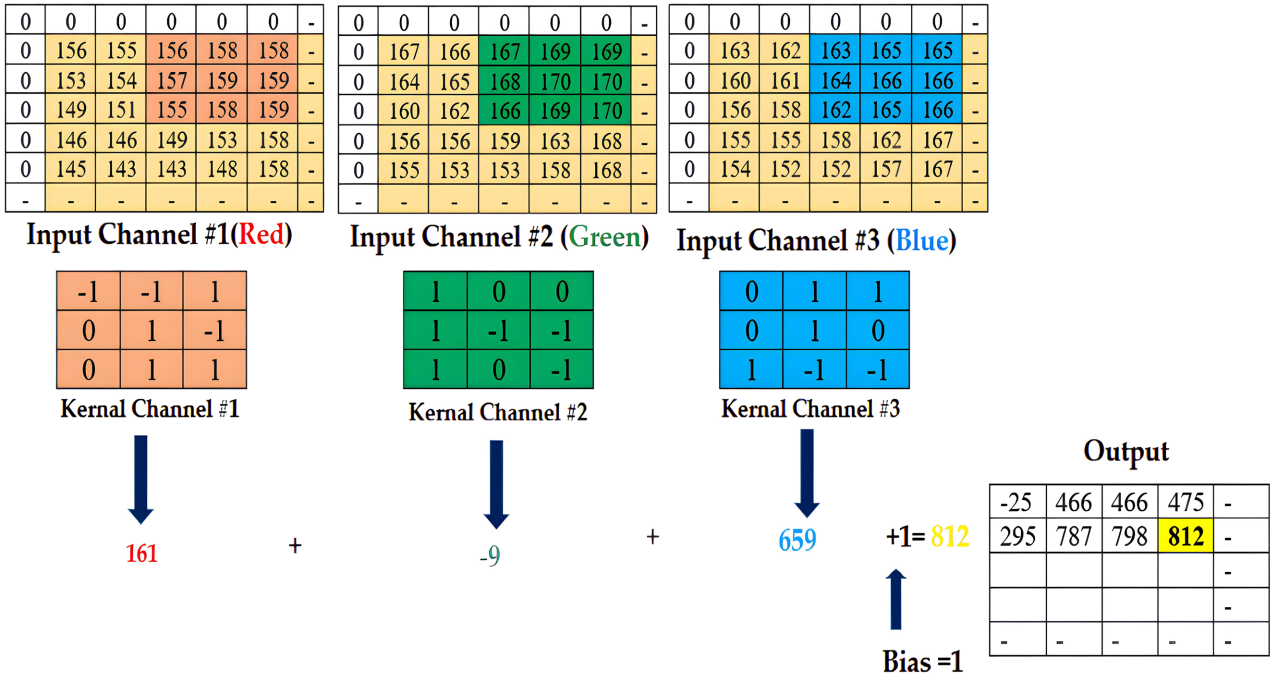

Figure 3. Processing of an RGB colour channel image by a convolution kernel. The convolution operation applies a filter across each colour channel to detect specific features, which are then combined to form a feature map.

2) Challenges in Training Deep Networks

Training deep neural networks presents several challenges, including the vanishing gradient problem, overfitting, high computational cost, and the requirement for large, labelled datasets [36]. The vanishing gradient problem is particularly relevant in deep architectures (see **Figure 4**). During backpropagation, gradients are calculated using the chain rule, and in deep networks, these gradients can become progressively smaller as they are multiplied through multiple layers. This issue is especially pronounced when using activation functions like sigmoid, whose derivatives are always less than or equal to 0.25. As a result, the gradients in the earlier layers become extremely small, hindering effective weight updates and slowing down or even halting the learning process in those layers [37]. This challenge can lead to slow convergence, requiring additional techniques such as batch normalization, residual connections, and careful weight initialization to maintain stable learning.

Figure 4 illustrates the training process of a deep neural network using backpropagation. The process begins with a feedforward calculation, where input neurons pass weighted sums through hidden layers to compute outputs. Each neuron applies an activation function to introduce non-linearity, enhancing learning capability. The output layer produces model predictions, which are compared to observed values to compute the error function. If the error is above a predefined tolerance, backpropagation updates the weights using gradient descent. Gradients are calculated using the chain rule to adjust weights and minimize the loss function iteratively. The weight update process continues for multiple epochs until convergence is achieved. The

training stops when the error falls below the specified threshold, ensuring optimal model performance. However, factors like learning rate selection, regularization, and the choice of optimizer play a crucial role in stabilizing training and preventing divergence.

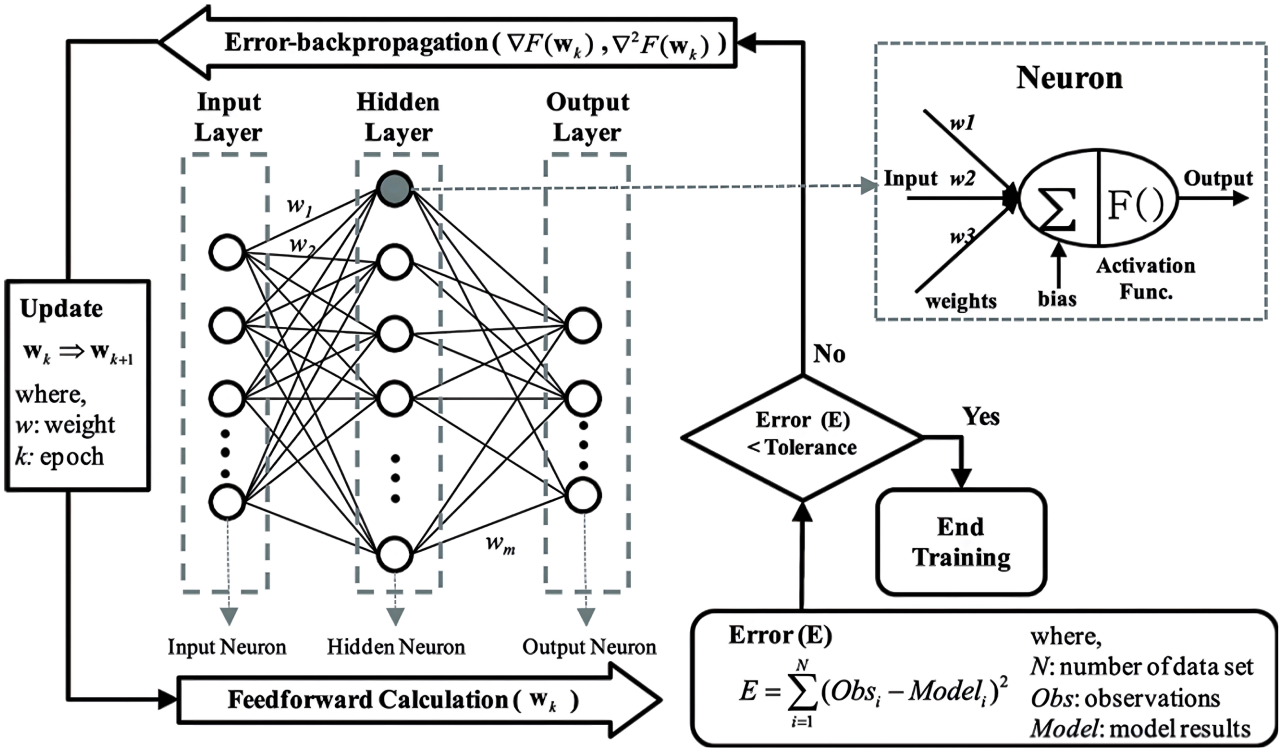

Figure 4. Illustration of the vanishing gradient problem during backpropagation. The diminishing gradients hinder the updating of weights in the earlier layers of deep neural networks, slowing down or halting learning.

3) Activations Functions

Activation functions are crucial components of deep learning models, introducing non-linearity that allows networks to approximate complex, non-linear functions and learn intricate patterns from input data [38-40]. In the context of medical image segmentation, the choice of activation function can significantly influence model performance, convergence speed, and the ability to mitigate issues like the vanishing gradient problem [41-43].

- **Rectified Linear Unit (ReLU):** ReLU is a widely adopted activation function that outputs the input directly if it is positive; otherwise, it outputs zero (Formula: ReLU(z) = max(0, z)). Its simplicity makes it computationally efficient and helps alleviate the vanishing gradient problem for positive inputs. However, ReLU can suffer from the "dying ReLU" problem, where neurons become inactive and output zero for all inputs if their weights are updated in a way that makes them always produce negative outputs [38].
- **Leaky ReLU:** Leaky ReLU addresses the "dying ReLU" problem by allowing a small, non-zero gradient when the input is negative (Formula: Leaky ReLU(z) = z if $z > 0$, else ϵz, where ϵ is a small constant, typically 0.01). This small slope for negative inputs ensures that neurons remain active and contribute to learning, even when their inputs are negative [39].
- **Swish:** Swish is a more recent activation function that exhibits a smooth, non-monotonic curve (Formula: Swish(z) = z*sigmoid(z)). Its non-monotonicity, meaning it doesn't strictly increase or decrease, may allow it to better capture complex patterns in the data. The smooth nature of Swish can lead to

better gradient flow compared to ReLU, potentially improving training performance in deep networks [41].

- **HardELiSH:** HardELiSH is a computationally efficient variant of the ELiSH activation function, using piece-wise linear approximations instead of exponential operations. Formula:

$$\text{HardELiSH}(z)=\begin{cases} z\max\left(0,\min\left(1,\left(\frac{z+1}{2}\right)\right)\right) & \text{if } z\geq 1 \\ \left(e^{z}-1\right)\max\left(0,\min\left(1,\frac{z+1}{2}\right)\right) & \text{if } z<0 \end{cases} \quad (1)$$

The piece-wise linear approximation aims to reduce computational costs without significant performance degradation.

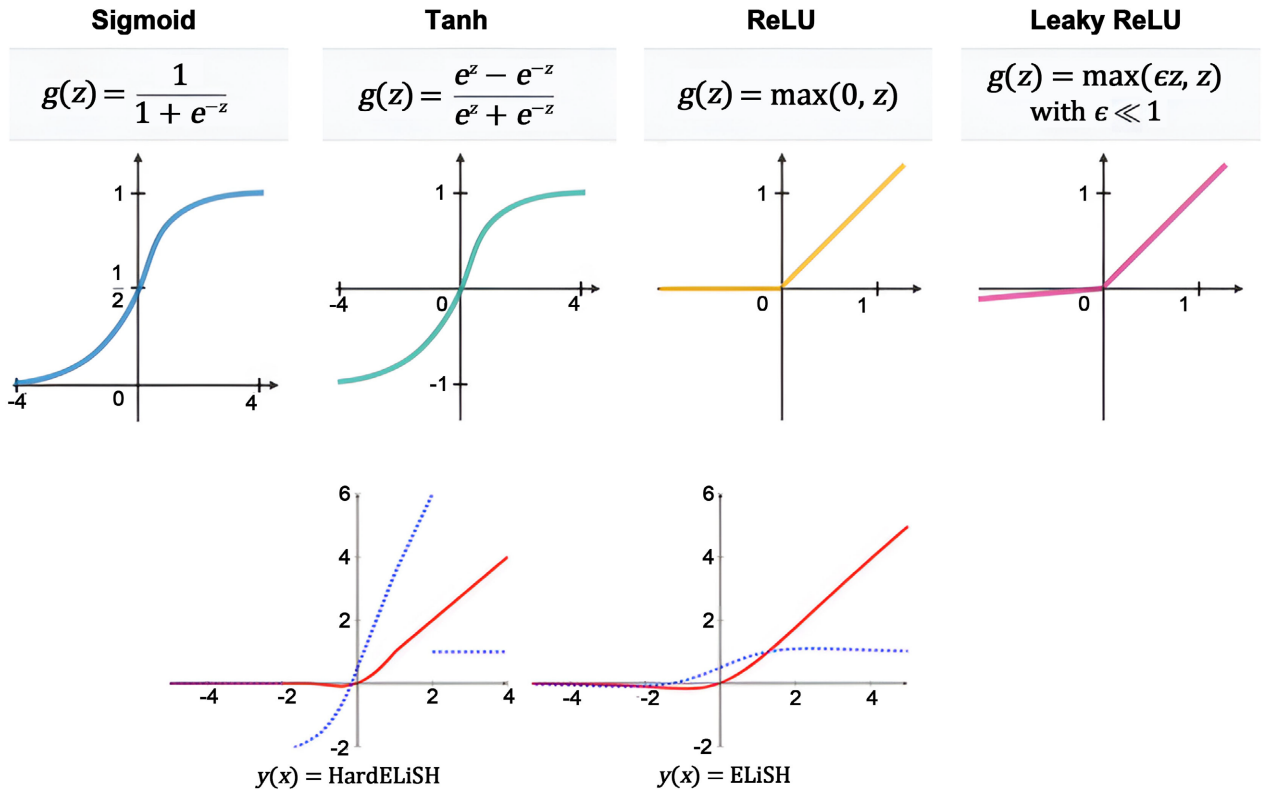

Figure 5. Common activation functions used in neural networks: Sigmoid, Tanh, ReLU (Rectified Linear Unit), Leaky ReLU, swish, and HardEliSH. These functions influence the network's learning dynamics and performance.

Figure 5 illustrates several commonly used activation functions. Sigmoid maps inputs to a range between 0 and 1. ReLU outputs the input directly if positive, otherwise zero. Leaky ReLU introduces a small slope for negative inputs. Sigmoid maps inputs to a range between 0 and 1. Swish exhibits a smooth, non-monotonic curve. HardELiSH offers a computationally efficient approximation of ELiSH using piece-wise linear functions. These functions play a critical role in shaping the learning dynamics and overall performance of neural networks. The HardELiSH is a computationally efficient approximation of the ELiSH (Exponential Linear Squashing) function, employing piece-wise linear functions to retain much of the performance benefits while reducing computational overhead. HardEliSH activation function has shown superior performance in brain tumour segmentation [44]. Its ability to balance computational efficiency with strong

gradient propagation makes it especially well-suited for segmentation tasks involving high-dimensional medical imaging data, such as those in 3D volumetric datasets like BraTS2020. This superior performance can be attributed to HardELiSH robust handling of non-linearities and its capacity to adapt to the complex patterns present in medical images, enabling better delineation of tumour boundaries and improved segmentation accuracy.

These functions are critical for shaping the learning dynamics and overall performance of neural networks, influencing the convergence speed, gradient flow, and ultimately the model's accuracy.

These advanced activation functions enable deep learning models to overcome the limitations of traditional functions, such as Sigmoid or Tanh, which suffer from the vanishing gradient problem in deeper networks. In the context of brain tumour segmentation, they provide essential benefits:

- **Improved Gradient Flow:** By mitigating the vanishing gradient problem, advanced functions ensure that deeper layers in models like 3D U-Net continue learning effectively, resulting in better segmentation performance.
- **Enhanced Feature Extraction:** These functions allow for more nuanced feature mapping, critical for detecting tumour boundaries and distinguishing different tissue types in MRI data.
- **Computational Efficiency:** Functions like ReLU and HardELiSH offer efficiency, enabling models to process high-dimensional MRI data faster, which is crucial for clinical applications.

Incorporating these advanced activation functions in brain tumour segmentation models can lead to more accurate, robust, and clinically viable results, making them a vital component in modern medical imaging techniques. Continued research into novel activation functions remains crucial for further improving the performance and efficiency of deep learning models in medical image analysis.

4) Fully Connected Layers

Following the convolutional and pooling layers, fully connected layers integrate the extracted features to perform classification or regression tasks (see **Figure 6**). In fully connected layers, each neuron is connected to every neuron in the previous layer, allowing the network to learn global patterns and relationships between features [45]. In the context of segmentation, these layers combine the learned features to produce the final segmentation map, assigning each pixel to a specific class (e.g., tumour or healthy tissue) [46].

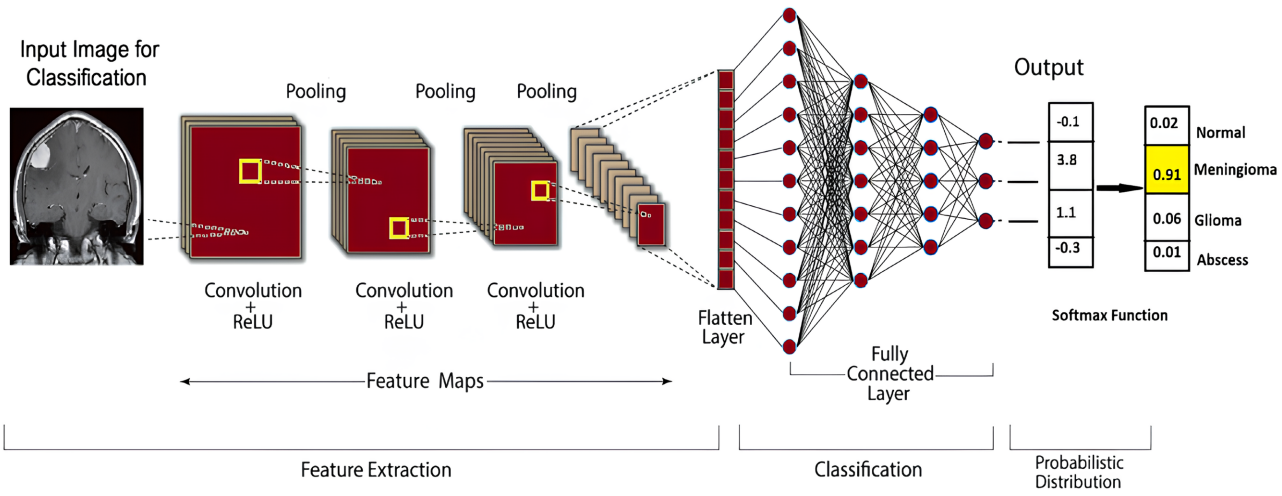

Figure 6. Block diagram of a CNN with fully connected layers. After convolutional and pooling layers extract features, the fully connected layers interpret these features to produce the final output, such as segmentation maps.

2.3.2. 3D U-Net Architecture

The U-Net architecture, originally proposed by Ronneberger *et al.* [18], has become a cornerstone of medical image segmentation. It features an encoder-decoder structure with skip connections. The encoder

path captures contextual information through successive convolutional and pooling layers, while the decoder path recovers spatial information through upsampling and concatenation with corresponding feature maps from the encoder path.

The 3D U-Net architecture extends the original U-Net to process volumetric data, making it particularly suitable for 3D medical imaging modalities like MRI (see Figure 7). The network consists of an encoder path that captures context through a series of 3D convolutional and pooling layers and a decoder path that enables precise localisation using 3D up-convolutional layers. Critically, skip connections between the encoder and decoder paths concatenate feature maps of the same resolution, allowing the network to combine high-level semantic information from the decoder path with fine-grained spatial details from the encoder path. This architecture allows for accurate segmentation by effectively capturing 3D spatial context and preserving detailed spatial information across all three dimensions.

The 3D U-Net architecture also benefits from its ability to process entire volumetric patches, reducing the need for slice-by-slice analysis and ensuring more consistent segmentations across contiguous slices. Its use of 3D convolutions enables the extraction of spatial features from the entire volume, capturing intricate patterns and relationships in medical imaging data. Additionally, the 3D U-Net is highly flexible and can be adapted to different scales and resolutions, making it suitable for tasks involving multi-resolution input data.

The ability of 3D U-Net to effectively capture and integrate hierarchical features makes it highly robust in scenarios involving irregularly shaped anatomical structures or small lesions. Moreover, the symmetrical design of the encoder and decoder paths facilitates efficient learning and optimization, while the skip connections help mitigate the risk of information loss during feature downsampling. As a result, 3D U-Net has become a preferred choice for a wide range of 3D segmentation tasks in medical imaging.

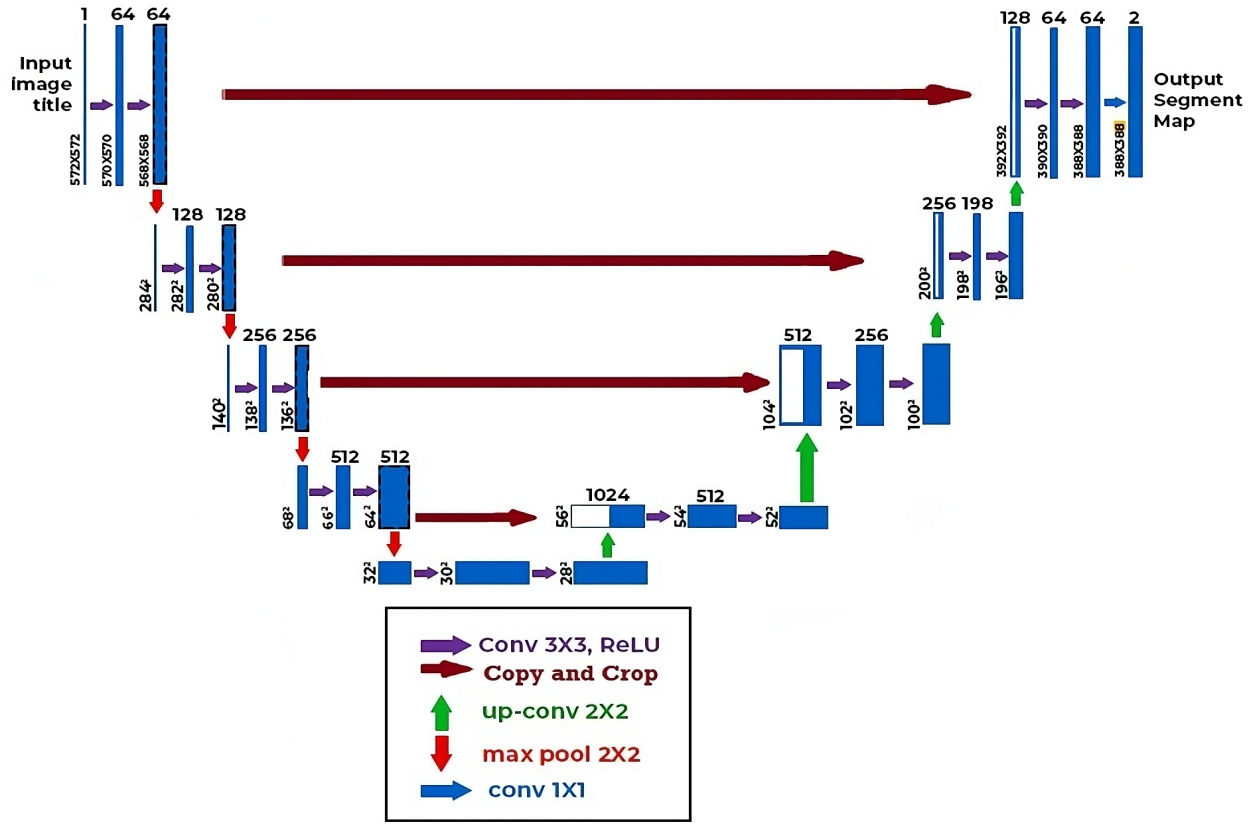

Figure 7. The 3D U-Net architecture for volumetric data segmentation. The network uses 3D convolutions, pooling, and upsampling layers, effectively capturing spatial relationships in all dimensions to improve segmentation accuracy.

2.3.3. Comparison between 2D and 3D U-Net

While 2D U-Net architectures process volumetric data slice-by-slice, 3D U-Net architectures operate directly on the entire volume [47-50]. This difference has significant implications for segmentation accuracy, computational cost, and the ability to capture spatial context. Table 1 summarises the key differences between 2D and 3D U-Net architectures.

Table 1. Differences between 2D U-Net and 3D U-Net architectures.

Features	Comparison between 2D U-Net and 3D U-Net	
	2D U-Net	3D U-Net
Input Data	2D images (slices of volumetric data)	3D volumetric data
Dimensions	Height and Width	Height, width, and depth
Convolution	2D convolution layers	3D convolution layers
Pooling	2D max pooling layers	3D max pooling layers
Up-Convolution	2D up-convolution layers	3D up-convolution layers
Spatial Context	Limited to individual slices	Maintains spatial continuity across slices
Feature Representation	Extracts features based on 2D context	Extracts features based on 3D context
Computational Complexity	Lower computational power and memory	Higher computational power and memory
Training and Accuracy	Faster and efficient	Higher accuracy and precision
Output	2D segmentation maps	3D segmentation maps

Table 1 summarizes the key differences between 2D U-Net and 3D U-Net architectures, highlighting aspects such as input data, dimensions, layer types (convolution, pooling, up-convolution), spatial context, feature representation, computational complexity, training efficiency, and the format of output segmentation maps.

2.3.4. Evaluation Metrics

To thoroughly evaluate the performance of segmentation models, it is crucial to employ a variety of metrics that assess different aspects of accuracy and overlap with the ground truth [51]. Each metric provides unique insights into the model's ability to accurately delineate tumour regions from healthy tissue.

- **True Positive (T_P):** Instances where the model correctly identifies a pixel as belonging to the tumour.
- **True Negative (T_N):** Instances where the model correctly identifies a pixel as belonging to healthy tissue (not part of the tumour).
- **False Positive (F_P):** Instances where the model incorrectly identifies a healthy pixel as belonging to the tumour.
- **False Negative (F_N):** Instances where the model incorrectly identifies a tumour pixel as belonging to healthy tissue.

1) Equations for Evaluations Metrics

- **Accuracy**

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \tag{2}$$

Accuracy measures the overall correctness of the model's predictions, representing the proportion of correctly classified pixels (both tumour and non-tumour) out of the total number of pixels However, accuracy can be misleading in cases of class imbalance, where the number of non-tumour pixels significantly

exceeds the number of tumour pixels [52]. In medical imaging, where tumour regions are often small compared to the background, accuracy alone may not adequately reflect the model's ability to detect pathological regions.

- **Precision**

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{3}$$

Precision quantifies the proportion of true positive predictions among all positive predictions made by the model. It reflects the model's ability to avoid false positives and is particularly important in medical applications where false positives can lead to unnecessary treatments [53]. However, high precision alone does not guarantee that the model is effective at identifying all tumour regions, as it does not account for false negatives.

- **Recall (Sensitivity)**

$$\text{Recall} = \frac{T_p}{T_P + F_N} \tag{4}$$

Recall, also known as sensitivity or the true positive rate, measures the proportion of actual tumour pixels that are correctly identified by the model. It reflects the model's ability to detect all positive instances and is crucial in medical diagnosis to minimise missed cases [53]. However, high recall may come at the cost of increased false positives, which can reduce precision.

- **Specificity**

$$\text{Specificity} = \frac{T_N}{F_P + T_N} \tag{5}$$

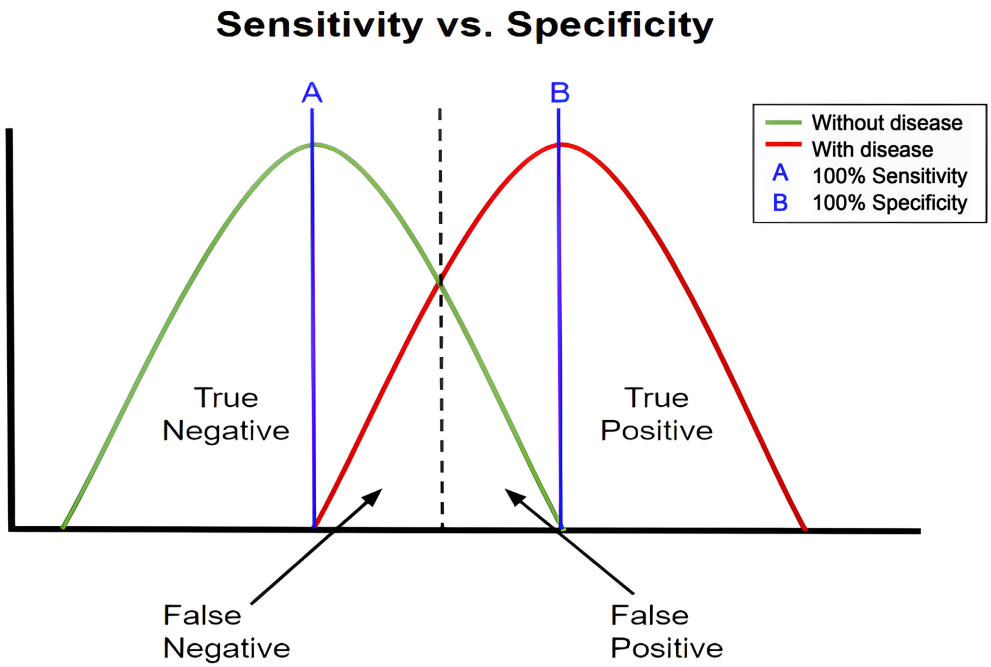

Figure 8. Graphical representation of sensitivity and specificity. Sensitivity (true positive rate) measures the proportion of actual positives correctly identified, while specificity (true negative rate) measures the proportion of actual negatives correctly identified. These metrics are essential for assessing the accuracy of segmentation models in detecting tumour presence [52].

Specificity, also known as the true negative rate, measures the proportion of actual non-tumour pixels that are correctly identified by the model. It reflects the model's ability to correctly identify negative instances [52]. Which is important for ensuring that healthy regions are not misclassified as pathological. However, specificity alone does not provide information about the model's ability to detect tumour regions.

Figure 8 provides a graphical representation of sensitivity and specificity. Sensitivity, also known as the true positive rate, measures the proportion of actual positives (tumour pixels) that are correctly identified by the model. Specificity, also known as the true negative rate, measures the proportion of actual negatives (non-tumour pixels) that are correctly identified. These metrics are essential for evaluating the accuracy of segmentation models in detecting the presence and absence of tumours [54].

- **F1-score (Dice Similarity Coefficient)**

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

The F1-score, also known as the Dice Similarity Coefficient (DSC) in the context of image segmentation, is the harmonic mean of precision and recall. It provides a balanced measure of the model's performance, which is particularly useful when both false positives and false negatives are of concern [55].

The Dice score is widely used in medical image segmentation due to its ability to handle class imbalance and provide a single metric that reflects both precision and recall. However, it may not fully capture the spatial accuracy of segmentation boundaries, which can be critical in clinical applications.

- **Jaccard Index**

$$\text{Jaccard Index} = \frac{T_P}{T_P + F_P + F_N} \tag{7}$$

The Jaccard Index, also known as the Intersection over Union (IoU), measures the overlap between the predicted segmentation and the ground truth segmentation. It is similar to the Dice score but is more sensitive to false positives [55]. It is particularly useful for evaluating segmentation performance in scenarios where precise boundary delineation is critical, such as surgical planning or radiation therapy. However, like the Dice score, it may not fully account for the clinical significance of segmentation errors.

- **Hausdorff Distance (HD)**

The Hausdorff Distance (HD) measures the maximum distance between the boundaries of the predicted segmentation and the ground truth. It is particularly useful for evaluating the spatial accuracy of segmentation boundaries, which is critical in applications such as radiotherapy planning, where precise tumour localization is essential. However, HD can be sensitive to outliers and may not reflect overall segmentation quality [56].

$$H(A,B) = \max\left(\max_{a\in A}\min_{b\in B}\|a-b\|, \max_{b\in B}\min_{a\in A}\|a-b\|\right) \tag{8}$$

- **Intersection over Union (IOU)**

Intersection over Union is a key metric in deep learning for evaluating segmentation and object detection models. It measures the overlap between the predicted and ground truth regions. IoU is calculated as:

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}} = \frac{A_{pred} \cap A_{gt}}{A_{pred} \cup A_{gt}} \tag{9}$$

where A_{pred} is the predicted region and A_{gt} is the ground truth. IoU values range from 0 to 1, where 1 indicates perfect overlap. Higher IoU means better model performance. In semantic segmentation, IoU is computed for each class and averaged (Mean IoU). A common threshold (e.g., 0.5) is used in object detection to determine true positives. IoU helps in model optimization by assessing prediction accuracy effectively.

- **Clinical Relevance and Limitations**

Each of these metrics provides unique insights into the performance of segmentation models, but they also have limitations that must be considered in the context of medical imaging:

- Accuracy is less informative in imbalanced datasets, where tumour regions are small compared to the background.
- Precision is critical for minimizing false positives, which can lead to unnecessary clinical interventions.
- Recall is essential for ensuring that all tumour regions are detected, reducing the risk of missed diagnoses.
- Specificity is important for correctly identifying healthy regions, but it does not directly reflect tumour detection performance.
- Dice Score and Jaccard Index provide balanced measures of segmentation performance but may not fully capture boundary accuracy or clinical significance.
- Hausdorff Distance is useful for evaluating boundary precision but can be sensitive to outliers.
- IoU measures the overlap between predicted and ground truth regions but is less sensitive to small structural differences and imbalanced datasets.

In clinical practice, a combination of these metrics is often used to comprehensively evaluate segmentation performance. For example, the Dice score and Hausdorff Distance are commonly used together to assess both overlap and boundary accuracy. Additionally, visual inspection by clinicians remains a critical step in validating segmentation results, as it provides context-specific insights that quantitative metrics may not capture.

Table 2. Summary of evaluation metrics.

Metrics	Evaluation Metrics	
	Formula	Description
Accuracy	$\frac{T_P + T_N}{T_P + F_P + F_N + T_N}$	Measures overall correctness of predictions.
Precision	$\frac{T_P}{T_P + F_P}$	Proportion of true positives among all positive predictions.
Recall (Sensitivity)	$\frac{T_p}{T_P + F_N}$	Proportion of actual positives correctly identified.
Specificity	$\frac{T_N}{F_P + T_N}$	Proportion of actual negatives correctly identified.
F1-score (Dice Similarity Coefficient)	$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall.
Jaccard Index	$\frac{T_P}{T_P + F_P + F_N}$	Measures overlap between predicted and ground truth segmentation.
Hausdorff Distance	$H(A,B) = \max\left(\max_{a\in A}\min_{b\in B}\lVert a-b\rVert, \max_{b\in B}\min_{a\in A}\lVert a-b\rVert\right)$	Measures maximum boundary deviation between prediction and ground truth.

Table 2 shows the evaluation metrics commonly used in segmentation tasks, including their formulas and descriptions. These metrics assess various aspects of model performance, such as overall accuracy, precision, recall (sensitivity), specificity, and overlap measures like F1-score (Dice Similarity Coefficient) and Jaccard Index. Additionally, the Hausdorff Distance evaluates the maximum boundary deviation between predictions and ground truth, providing insights into segmentation quality.

2.4. Overview of Brain Tumour Segmentation Using Deep Learning

Deep learning techniques have revolutionised the field of brain tumour segmentation by enabling models to learn intricate features directly from data, leading to significant improvements in accuracy and efficiency. This section provides an overview of key contributions in this area, highlighting both general studies and notable advancements from the annual Brain Tumour Segmentation (BraTS) challenges. **Figure 1** illustrates different MRI modalities and their corresponding segmented brain tumours, showcasing the ability of deep learning models to delineate complex tumour regions.

2.4.1. Contributions and Limitations from General Studies

Several studies have explored diverse deep-learning approaches for brain tumour segmentation, each with its strengths and limitations:

Table 3. **Contributions and limitations from general studies.**

Author	General Studies		
	Year	Contributions	Limitations
Fernando and Tsokos [57]	2023	The study Integrated statistical methods with deep learning for 3D MRI brain tumour segmentation. Their approach showed enhanced accuracy compared to traditional methods	The proposed study suffered from slow convergence and susceptibility to local optima during training.
Montaha *et al.* [58]	2023	This study explored a standard U-Net architecture for brain tumour segmentation, demonstrating competitive performance.	Noted the computational complexity of the model and the potential for dataset-specific limitations in its generalizability.
Ullah *et al.* [59]	2021	Introduced a brain MR image enhancement method as a pre-processing step to improve segmentation using a 3D U-Net. They achieved high Dice scores	They did not outperform existing state-of-the-art techniques.
Feng *et al.* [60]	2020	Proposed an ensemble of 3D U-Nets combined with a multivariate linear regression model. While their method improved segmentation accuracy	They raised concerns about overfitting due to the limited size of the datasets used.
Henry *et al.* [61]	2020	Presented a 3D U-Net-based solution that incorporated self-ensembling and deep supervision. Their method achieved robust segmentation results.	This study was limited by the diversity of the datasets and high computational demands.
Ballester and Vilaplana [62]	2020	Investigated 3D U-Net architectures with patch-based techniques for MRI brain tumour segmentation. They proposed using model ensembles to improve performance.	The presented study Acknowledged the potential for increased false positive detections, particularly in the enhancing tumor region.

Continued

Wang *et al.* [63]	2020	Developed a 3D U-Net-based method for both segmentation and survival prediction. Their approach enhanced segmentation accuracy	Exhibited modest performance in survival prediction and faced computational challenges.
Sun *et al.* [64]	2019	Developed an automated framework that combined deep CNNs with Radiomic features extracted from MRI data. Their approach improved segmentation performance	Showed limitations in accurately predicting patient survival.

Table 3 summarizes the contributions and limitations of various general studies on brain tumor segmentation using 3D medical imaging. The studies span a range of methodologies, including statistical integration with deep learning, U-Net architectures, preprocessing enhancements, and ensemble models. Contributions include improved segmentation accuracy, robust frameworks, and innovative preprocessing techniques. However, limitations such as computational complexity, dataset-specific challenges, overfitting, slow convergence, and modest survival prediction performance highlight the need for further advancements in these approaches.

2.4.2. BraTS Challenges Contributions and Limitations and Research Gaps (2014-2023)

The Multimodal Brain Tumour Segmentation (BraTS) challenge has played a crucial role in advancing the field by providing standardised datasets, evaluation metrics, and a common platform for comparing different methodologies [24]. Each year, the challenge has spurred the development of novel techniques and pushed the boundaries of segmentation accuracy. The following is a summary of notable contributions and achievements from each year of the BraTS challenge, showcasing the evolution of segmentation techniques.

- **2014—Establishment of the BraTS Benchmark:**
  - **Contribution:**

Menze *et al.* [24] established the initial BraTS dataset and benchmark, providing a crucial resource for evaluating and comparing brain tumour segmentation models. The dataset included multi-contrast MRI scans (T1, T1ce, T2, FLAIR) with expert annotations of tumour regions (whole tumour, tumour core, enhancing tumour).

The study's contributions include organizing the BRATS 2012 and 2013 challenges, which created a unique public dataset of MR scans for brain tumour segmentation with multiple expert annotations. The authors established the annotation protocol, acquired and annotated clinical images, generated synthetic datasets, and developed pre-processing methods and evaluation scripts. They also maintained online validation tools for community use. Notably, the study highlights that fusing different segmentation algorithms can significantly improve performance, suggesting that future advancements may stem from exploring various fusion strategies, thereby providing a valuable resource for researchers in medical imaging and brain tumour segmentation.

  - **Limitations:**

The study on the BRATS benchmark identifies several limitations, including the high variability in brain tumours, which complicates algorithm rankings and comparisons. Challenges arise from differing test settings across workshop years, making direct comparisons difficult. Additionally, reliance on a single set of annotations for test data may not adequately capture expert variability, potentially skewing results. The complexity of algorithms further obscures the understanding of performance discrepancies, while the choice of evaluation metrics can significantly influence rankings. Furthermore, the lack of longitudinal analysis for automated routines highlights a critical gap in assessing their long-term reliability and effectiveness. These findings underscore the need for future research and benchmarks to address these limitations and advance the field of brain tumour segmentation.

➢ **Research Gap and Future Directions:**

The study identifies several research gaps in brain tumour segmentation, including the high variability of brain tumours, which complicates the evaluation of segmentation algorithms, and the reliance on a single set of annotations that may not fully represent expert opinions. To address these challenges, future improvements could involve expanding datasets to better capture tumour variability, incorporating multiple expert annotations for more reliable comparisons, and conducting longitudinal studies to assess algorithm performance over time. Additionally, enhancing algorithmic transparency through detailed analysis of segmentation pipelines, exploring diverse evaluation metrics for comprehensive assessments, and investigating fusion strategies to combine multiple algorithms for improved performance are recommended. Addressing these areas can significantly enhance the effectiveness and reliability of brain tumour segmentation algorithms.

- **2016—Introduction of 3D Convolutional Neural Networks:**

➢ **Contribution**:

Kamnitsas *et al.* [65] introduced a 3D CNN architecture called DeepMedic for brain tumour segmentation, demonstrating the advantages of using volumetric data over 2D slice-based methods.

The study makes several significant contributions to the field of brain tumour segmentation by enhancing the DeepMedic architecture with residual connections. A key contribution is the demonstration of how residual connections improve the performance of a 3D CNN model, resulting in higher sensitivity and accuracy in segmenting brain tumours. The study includes a comprehensive evaluation of the model on the BRATS 2016 challenge, where it achieves competitive performance compared to other state-of-the-art methods. Furthermore, the research underscores the critical role of data augmentation and normalization techniques in enhancing model robustness, particularly when applied to heterogeneous testing scenarios. These findings provide valuable insights for advancing brain tumour segmentation methodologies.

➢ **Limitations:**

The study acknowledges several limitations that warrant consideration. A significant limitation is the model's reliance on high-quality training data, as its performance can degrade when there is a mismatch between the training and testing distributions. While the DeepMedic model demonstrates strong overall performance, it faces challenges in accurately segmenting fine substructures, such as necrosis and non-enhancing tumours. Additionally, the experiments were primarily conducted under conditions where the training and testing distributions were similar, leaving a gap in understanding how the model generalizes to more diverse and heterogeneous datasets. These limitations highlight areas for future research to improve the robustness and applicability of the model in real-world clinical settings.

➢ **Research Gap and Future Directions:**

The study identifies several research gaps and outlines promising future directions. It suggests exploring the optimal amount of external data required to enhance model generalization to new and diverse distributions, as well as investigating the relationship between network capacity and the volume of available training data. Additionally, the authors propose further research into the impact of different architectural choices and training strategies on segmentation performance, particularly for challenging substructures such as necrosis and non-enhancing tumours. Addressing these areas could lead to the development of more robust, efficient, and adaptable models, ultimately improving clinical outcomes in brain tumour diagnosis and treatment.

- **2017—Ensembles of Multiple Architectures:**

➢ **Contribution:**

Kamnitsas *et al.* [66] further developed the Ensemble of Multiple Models and Architectures (EMMA), combining DeepMedic, FCN, and U-Net to boost performance and reduce overfitting.

The study makes significant contributions to medical image analysis, particularly in brain tumour segmentation. The authors introduce EMMA, an ensemble method that integrates multiple models and architectures to improve segmentation accuracy. By combining diverse neural network configurations and training strategies, the study demonstrates that ensembling effectively reduces biases and variances inherent in individual models, resulting in enhanced robustness for segmenting both high-grade and low-grade gliomas demonstrating the benefits of ensemble methods in increasing robustness and accuracy. The approach is

validated using a comprehensive dataset from the Brain Tumour Segmentation Challenge 2017, underscoring its effectiveness.

- **Limitations:**

The study also acknowledges several limitations. A key challenge is the reliance on the quality and diversity of training data, which can affect the generalizability of the models. Additionally, the complexity of the ensemble method introduces computational overhead, potentially limiting its feasibility for real-time clinical applications. While the approach improves segmentation performance, it does not fully address the issue of model interpretability, which remains a critical concern in medical applications where understanding the decision-making process is essential.

- **Research Gap and Future Directions:**

The study identifies several research gaps and future directions. These include exploring unsupervised and semi-supervised learning techniques to enhance model training with limited labeled data, as well as integrating multi-modal imaging data to further improve segmentation accuracy and robustness. Future work could also focus on developing more efficient ensemble methods that reduce computational demands without compromising performance. Finally, advancing model interpretability through explainable AI techniques could facilitate the practical adoption of these methods in clinical environments, bridging the gap between cutting-edge research and real-world applications.

- **2018—Encoder-Decoder Architectures with Autoencoder Regularization:**
  - **Contribution:**

Myronenko [67] proposed a 3D encoder-decoder architecture with an additional branch for variational autoencoder (VAE) regularisation. This approach aimed to improve segmentation accuracy by incorporating a generative model to learn a latent representation of the input data.

The study presents significant contributions to the field of brain tumour segmentation through the development of an innovative autoencoder-based approach. The authors introduce a semantic segmentation network that utilizes an encoder-decoder architecture, which effectively addresses the challenges of limited training data. By employing test time augmentation and ensembling multiple models, the approach enhances segmentation accuracy across three nested tumour subregions: whole tumour, tumour core, and enhancing tumour. The results achieved in the BraTS 2018 challenge demonstrate the method's effectiveness, positioning it among the top-performing submissions and showcasing its potential for clinical applications in neuro-oncology.

  - **Limitations:**

Despite these advancements, the study acknowledges certain limitations. One notable constraint is the reliance on the availability of high-quality annotated datasets, which can be scarce in medical imaging. Additionally, while the proposed method shows promising results, it may still struggle with variability in tumour presentations and imaging artifacts that can affect segmentation performance. The authors also highlight the computational demands of training and deploying deep learning models, which may limit accessibility for some research and clinical settings.

  - **Research Gap and Future Directions:**

This proposed study identifies several research gaps and future directions. There is a need for further exploration of unsupervised and semi-supervised learning techniques to mitigate the dependency on annotated data. Additionally, the authors suggest investigating the integration of multi-modal imaging data to improve segmentation robustness and accuracy. Future work could also focus on real-time segmentation applications in clinical environments, as well as the development of user-friendly tools that facilitate the adoption of these advanced techniques by healthcare professionals. Overall, the study lays a foundation for ongoing research in automated brain tumour segmentation, emphasizing the importance of collaboration between machine learning and clinical expertise.

- **2019—Cascaded U-Net Approach:**
  - **Contribution:**

Jiang *et al.* [68] introduced a Two-Stage Cascaded U-Net architecture that refined segmentation predictions in a coarse-to-fine manner. The first stage produced a preliminary segmentation, which was then

used as input to the second stage for refinement.

The paper makes several significant contributions to the field of brain tumour segmentation. It introduces a novel two-stage cascaded U-Net architecture designed to segment tumour substructures from coarse to fine detail. Trained end-to-end on the BraTS 2019 dataset, which includes diverse high-grade and low-grade gliomas, the model achieved state-of-the-art performance, securing first place in the BraTS 2019 challenge. The method demonstrated exceptional robustness for different tumour regions.

➢ **Limitations:**

A key limitation is the reliance on the BraTS 2019 dataset, which may not fully capture the variability encountered in real-world clinical scenarios. Additionally, the model's performance may be influenced by the quality and consistency of input images, and its complex architecture raises concerns about potential overfitting. The authors also note that variability in segmentation performance across different models suggests a need for further refinement and validation to ensure generalizability.

➢ **Research Gap and Future Directions:**

The research gaps and future directions in this study include exploring more diverse datasets to enhance the model's robustness and applicability across clinical settings. The authors also propose investigating the integration of additional modalities and advanced techniques, such as attention mechanisms or unsupervised learning, to further improve segmentation accuracy. Future work could focus on enabling real-time segmentation capabilities and developing user-friendly tools for clinicians, ultimately aiming to improve patient outcomes through more efficient and accurate tumour analysis and monitoring.

- **2020—nnU-Net Architecture and Batch Normalization:**

➢ **Contribution:**

Isensee *et al.* [69] proposed the nnU-Net, a self-configuring framework that automatically adapted to new datasets by adjusting various parameters, including network topology, pre-processing, and training details. They replaced instance normalisation with batch normalisation.

The paper highlights significant contributions of the nnU-Net to brain tumour segmentation. The authors demonstrated the generalizability of nnU-Net by achieving high segmentation accuracy without requiring extensive modifications. By implementing specific enhancements tailored for the BraTS 2020 challenge, the model outperformed the baseline on the validation set. The final ensemble model, selected based on the best-performing configurations, achieved impressive Dice scores and HD95 values for whole tumour, tumour core, and enhancing tumour segmentation, ultimately securing first place in the BraTS 2020 competition. This work underscores the effectiveness of nnU-Net in medical image segmentation, particularly for brain tumours.

➢ **Limitations:**

The study acknowledges notable limitations. The study covers only a limited number of modifications and lacks extensive experimental validation, raising concerns about the robustness and reproducibility of the results. The performance improvements observed may not generalize across different datasets or clinical scenarios. Additionally, reliance on specific validation metrics, such as Dice scores and HD95, may not fully capture the clinical relevance of the segmentation results, potentially limiting the applicability of the findings in real-world settings.

➢ **Research Gap and Future Directions:**

Several research gaps and future directions can be explored from this study. These include investigating more extensive modifications and optimizations of the nnU-Net architecture, such as integrating attention mechanisms or multi-task learning. Comprehensive validation studies across diverse datasets would further enhance understanding of the model's generalizability and robustness. Additionally, addressing class imbalance in brain tumour segmentation through novel loss functions or data augmentation strategies could improve performance. These efforts have the potential to advance segmentation accuracy and contribute to better clinical outcomes for patients with brain tumours.

- **2021—Enhanced nnU-Net with Group Normalization:**

➢ **Contribution:**

Luu and Park [70] extended the nnU-Net by doubling the filter sizes in the initial layers and implementing

group normalisation instead of batch normalisation.

The proposed study makes significant contributions to brain tumour segmentation by extending the nnU-Net framework, a well-established method in medical image analysis. The authors introduced several key modifications, including a larger network architecture, the replacement of batch normalization with group normalization, and the incorporation of axial attention in the decoder. These enhancements led to improved performance in the Brain Tumour Segmentation Challenge (BraTS) 2021, where their models secured first place in the final ranking on unseen test data. Additionally, the authors made their codes, pre-trained weights, and Docker image publicly available, promoting transparency and enabling further research in the field.

- **Limitations:**

Despite these advancements, the study acknowledges certain limitations. A notable challenge is the model's dependency on high-quality MR images, as segmentation performance can degrade with poor image quality or artifacts. This underscores the importance of data integrity in training robust models. Furthermore, while the modifications to nnU-Net yielded incremental improvements over the baseline, the authors suggest that further experimentation is needed to fully explore the potential of these enhancements and their impact on segmentation accuracy across diverse datasets.

- **Research Gap and Future Directions:**

The research gap identified in the study highlights the need for more comprehensive datasets that capture a wider variety of tumour characteristics and imaging conditions. Future directions include exploring diverse data acquisition methods and advanced data augmentation techniques to address edge cases where segmentation performance may falter. The authors also emphasize the importance of analysing failure cases to refine model behaviour and improve overall accuracy. These efforts could lead to the development of more resilient algorithms capable of handling the complexities of brain tumour imaging in clinical settings.

- **2022—Multi-Framework Ensembles with DeepSeg and nnU-Net:**
  - **Contribution**:

Zeineldin *et al.* [71] developed an ensemble approach combining multiple frameworks, including DeepSeg, nnU-Net [69], and DeepSCAN, for automatic glioma segmentation in pre-operative MRIs.

The study makes significant contributions to brain tumour segmentation by developing an ensemble model that integrates multiple state-of-the-art U-Net variants, including DeepSeg, nnU-Net, and DeepSCAN. This ensemble approach achieved exceptional performance in the BraTS-CE 2022 challenge, delivering high Dice Similarity Coefficients (DSC) and low Hausdorff distances (HD95) for the enhancing tumour (ET), tumour core (TC), and whole tumour (WT). The method secured first place in the competition and demonstrated strong generalizability on unseen test datasets, highlighting its potential for clinical applications in automatic glioma boundary detection using pre-operative MRI scans.

- **Limitations:**

The study acknowledges certain limitations, particularly in the context of pediatric brain tumour segmentation. The inherent variability and complexity of pediatric tumours, which often exhibit heterogeneous tissue types, pose significant challenges for accurate segmentation. The model's performance on the pediatric test dataset was less robust compared to adult datasets, underscoring the need for specialized techniques and larger annotated datasets tailored to paediatric cases. This limitation highlights the importance of further research to address the unique characteristics of paediatric brain tumours.

- **Research Gap and Future Directions:**

This study identifies several research gaps and future directions. These include developing more sophisticated models capable of handling the complexities of paediatric brain tumour segmentation, integrating multimodal imaging data, and exploring advanced post-processing techniques to enhance accuracy. Additionally, expanding dataset diversity and size could improve model generalization and robustness across different patient populations. Addressing these gaps could lead to more effective and reliable tools for brain tumour diagnosis and treatment planning, ultimately benefiting both adult and pediatric patients.

- **2023—Enhanced Synthetic Data Augmentation and GANs:**
  - **Contribution:**

Ferreira *et al.* [72] employed synthetic data augmentation through Generative Adversarial Networks (GANs) and integrated multiple deep learning models, including nnU-Net and Swin UNETR.

The study makes significant contributions to brain tumour segmentation, particularly in the context of the BraTS 2023 Adult Glioma challenge. The authors introduced an innovative approach combining Enhanced Synthetic Data Augmentation with Model Ensemble techniques. By leveraging generative adversarial networks (GANs) and advanced registration methods, they increased the volume and quality of training data, significantly improving the performance of deep learning models in segmenting gliomas. This work not only advances the state-of-the-art in brain tumour segmentation but also provides a framework for future research in synthetic data generation and model optimization.

- **Limitations:**

The study acknowledges certain limitations. A key challenge is the reliance on high-quality annotated datasets, which are often scarce and time-consuming to obtain. Additionally, while the proposed methods show promising results, they may not generalize well across diverse patient populations or different imaging modalities. The authors also highlight the computational demands of their approach, which could limit its applicability in real-time clinical settings. These limitations suggest that further refinement and validation are needed to ensure robustness and practicality.

- **Research Gap and Future Directions:**

The study identifies several research gaps and future directions. These include exploring the generalizability of the proposed methods across diverse demographics and imaging techniques, as well as investigating the integration of multimodal data to further enhance segmentation accuracy. Future work could also focus on developing more efficient algorithms to reduce computational costs while maintaining high performance. Overall, the findings emphasize the importance of continued innovation in synthetic data generation and deep learning methodologies to address the evolving challenges in medical imaging and tumour segmentation.

Table 4 summarizes the methods and results from the BraTS Challenges between 2014 and 2023, showcasing advancements in segmentation techniques for whole tumor (WT), core tumor (CT), and enhanced tumor (ET). Early approaches, such as the 2014 benchmark by Menze *et al.*, established foundational metrics with Dice scores of >0.80 (WT), 0.70 (CT), and 0.60 (ET). Subsequent methods, including 3D CNNs, ensembles, and advanced U-Net architectures, demonstrated significant improvements in segmentation accuracy. Notably, Luu and Park's enhanced nnU-Net achieved the highest Dice scores in 2021, with 0.938 (WT), 0.923 (CT), and 0.882 (ET). Recent efforts in 2023, incorporating synthetic data augmentation and GANs, further refined performance, emphasizing the growing sophistication of deep learning techniques in medical imaging.

Table 4. Summary of BraTS challenges methods and results [2014-2023].

Author	Year	Dice Score Results (WT, CT, and ET)			
		Method	Whole Tumour	Core Tumour	Enhanced Tumour
Menze *et al.* [24]	2014	Established the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)	>0.80	0.70	0.60
Kamnitsas *et al.* [65]	2016	3D Convolutional Neural Network (CNN)	0.896	0.754	0.718
Kamnitsas *et al.* [66]	2017	Ensembles of Multiple Models and Architectures (EMMA). DeepMedic, FCN, and U-Net.	0.901	0.797	0.738

Continued

Myronenko [67]	2018	Encoder-Decoder Architectures with Autoencoder Regularization	0.910	0.867	0.823
Jiang *et al.* [68]	2019	Cascaded U-Net Approach for refinement of segmentation	0.909	0.864	0.802
Isensee *et al.* [69]	2020	nnU-Net Architecture and Batch Normalization	0.912	0.850	0.798
Luu and Park [70]	2021	Enhanced nnU-Net with Group Normalization	0.938	0.923	0.882
Zeineldin *et al.* [71]	2022	Multi-Framework Ensembles with DeepSeg and nnU-Net, and DeepSCAN	0.927	0.875	0.843
Ferreira *et al.* [72]	2023	Enhanced Synthetic Data Augmentation and GANs and integration of multiple deep learning models (nnU-Net, Swin UNETR)	0.905	0.867	0.850

3. DISSCUSIONS

Advancements in artificial intelligence (AI), machine learning (ML) and particularly deep learning (DL) have ushered in a new era for medical imaging diagnostics, with brain tumour segmentation being a key area of impact. Deep learning models, notably Convolutional Neural Networks (CNNs) and U-Net architectures, have demonstrated remarkable success in automating the segmentation process by efficiently extracting relevant features from imaging data. The evolution towards 3D U-Net architectures has further leveraged the richness of volumetric data to enhance segmentation accuracy by capturing spatial context across all dimensions [21]. This review has highlighted the significant progress made in this field, as evidenced by the increasing sophistication and performance of models developed over the past decade. However, despite these advancements, several challenges continue to hinder the widespread clinical applicability of these models. This discussion critically analyses the current state of the field, synthesises key trends, identifies existing limitations, and explores practical implications and future research directions for both clinicians and researchers.

3.1. Challenges in Deep Learning-Based Brain Tumour Segmentation

3.1.1. Tumour Heterogeneity

A major challenge consistently identified in the literature is the significant variability in brain tumours' size, shape, location, and appearance across patients [21, 22, 24]. This heterogeneity complicates model generalisation, as models trained on datasets with limited tumour characteristics may not perform well on the diverse range of tumours encountered in clinical practice. The infiltrative growth patterns and indistinct boundaries of gliomas further exacerbate the challenges of accurate segmentation.

3.1.2. Limited Data and Class Imbalance

The scarcity of high-quality, annotated medical imaging data remains a significant bottleneck [57, 64]. This scarcity stems from privacy concerns, the labour-intensive nature of manual annotation, and the relatively low prevalence of certain tumour types. Moreover, the inherent class imbalance between the vast non-tumour regions and the relatively small tumour regions can bias models towards the majority class, leading

to reduced segmentation performance, particularly for smaller tumours or tumour subregion. This is particularly concerning in the medical domain, where false negatives can have severe clinical consequences.

Differences in MRI scanners, acquisition parameters, and imaging protocols across institutions introduce considerable variability that significantly affects model performance and generalisation. Models trained on data from one institution may not perform adequately on data acquired from another due to differences in image quality, contrast settings, and noise levels. Standardisation of imaging protocols is often proposed as a solution, but it faces practical challenges in implementation.

3.1.3. Computational Complexity

The literature consistently highlights the substantial computational resources and memory required to train deep 3D models, such as the 3D U-Net, due to the high dimensionality of volumetric data [58, 61]. This poses a significant challenge for researchers and clinical institutions with limited access to high-performance computing infrastructure, hindering widespread adoption and limiting the development of more complex, potentially more accurate models.

3.1.4. Overfitting and Generalization

Many studies report challenges related to overfitting, where models trained on limited datasets capture noise and artefacts instead of the underlying patterns of tumour morphology [60, 64]. This leads to poor generalisation performance on external, unseen datasets, severely limiting the clinical utility of such models. Ensuring robust generalisation to new data is, therefore, a critical challenge that needs to be addressed for reliable clinical deployment.

3.2. Comparative Analysis of Methods

The evolution of brain tumour segmentation techniques, particularly evident in the contributions to the BraTS challenges from 2014 to 2023, reflects a clear trend towards increasingly sophisticated models. Early CNN-based methods achieved baseline performance but struggled with generalisation due to limited data diversity [22]. The subsequent introduction of 3D CNNs capitalised on volumetric data, improving spatial feature representation and leading to significant performance gains. For example, the average Dice scores for whole tumour segmentation in BraTS improved from around 0.8 in the early years to over 0.90 with the adoption of 3D CNNs.

Ensemble methods, which combine multiple architectures to enhance robustness and mitigate individual model weaknesses, have also gained traction. These methods have been shown to further improve segmentation accuracy, often achieving Dice scores of around 0.88 in BraTS challenges. More recently, the nnU-Net framework [69], which dynamically adapts models to dataset characteristics through normalization techniques and architectural refinements, has emerged as a state-of-the-art approach, achieving Dice scores exceeding 0.9 for certain tumour subregion in recent BraTS challenges.

The incorporation of synthetic data augmentation using Generative Adversarial Networks (GANs) represents another notable trend aimed at addressing data scarcity and diversity issues [73]. GANs, such as CycleGAN and StyleGAN, have been widely adopted in medical imaging for generating realistic synthetic data [74]. For instance, CycleGAN has been used for domain adaptation, enabling the translation of images from one modality (e.g., MRI) to another (e.g., CT) while preserving anatomical structures. Similarly, StyleGAN has shown promise in generating high-resolution, diverse tumour images that enhance the generalizability of segmentation models. However, these techniques require careful validation to ensure the generated data are realistic and do not introduce biases that could negatively impact model training. Challenges include mode collapse, where the generator produces limited varieties of images, and the difficulty of ensuring that synthetic images retain clinically relevant features [75].

In addition to GANs, domain adaptation methods such as unsupervised domain adaptation (UDA) and adversarial domain adaptation have been employed to bridge the gap between source and target domains in medical imaging. For example, UDA techniques like Domain-Adversarial Neural Networks (DANN) have been used to align feature distributions across different imaging modalities, improving model

performance on unseen datasets. While these methods reduce the need for extensive labeled data in the target domain, they often struggle with large domain shifts and may require additional fine-tuning for specific applications [76].

Despite these advancements, the reviewed literature indicates that significant challenges remain in handling the full spectrum of tumour heterogeneity and reducing the computational demands of these models. For instance, GAN-based approaches are computationally intensive and may require significant resources for training and validation [77]. Furthermore, the lack of standardized evaluation metrics for synthetic data quality poses a barrier to widespread adoption in clinical settings.

3.3. Limitations of BraTS Challenges

The BraTS dataset, while comprehensive, has several limitations that impact its applicability to real-world clinical scenarios. Firstly, the dataset primarily focuses on adult gliomas and may not fully capture the variability seen in paediatric brain tumours or rare tumour types. This limits the generalizability of models trained on BraTS data. Additionally, the reliance on pre-operative MRI scans excludes post-operative or follow-up imaging, which are critical for longitudinal studies and treatment monitoring in clinical practice. Another limitation is the annotation process, which relies on a limited number of experts. This may not fully represent the variability in clinical interpretations, potentially introducing biases in model training and evaluation. Furthermore, the lack of multi-institutional annotations or consensus-based ground truth may restrict the dataset's applicability to diverse clinical settings. Finally, while metrics like Dice scores and Hausdorff distances are widely used, they may not fully capture the clinical relevance of segmentation results. For example, small errors in critical regions, such as tumour boundaries near vital structures, may have significant clinical implications but are not adequately reflected in these metrics.

3.4. Suggestions for Improving BraTS Challenges Future Iterations

To address these limitations and better align the BraTS challenge with real-world clinical needs, we propose several improvements for future iterations.

- Expanding the dataset to include a broader range of tumour types, such as pediatric gliomas, metastatic tumours, and rare subtypes, would better reflect real-world clinical diversity. Incorporating post-operative and follow-up imaging data would also enable the development of models for longitudinal analysis and treatment monitoring.
- Enhancing annotation quality through multi-institutional annotations and consensus-based ground truth could improve the robustness and representativeness of the dataset. Including annotations for clinically significant subregion, such as peritumoral edema or infiltrative tumour margins, would further enhance the dataset's utility.
- Refining evaluation metrics to account for the spatial accuracy of critical regions and incorporating metrics that evaluate model uncertainty and interpretability would provide a more comprehensive assessment of model performance.
- Addressing computational and practical challenges by encouraging the development of efficient models suitable for real-time applications and providing guidelines for robustness to imaging artifacts would better align the challenge with clinical needs.

3.5. Bridging the Gap between BraTS Performance and Clinical Applications

Despite the impressive performance of models trained on BraTS data, there remains a gap between dataset performance and practical clinical applications. Models may struggle with real-world challenges such as poor image quality, imaging artifacts, or heterogeneous tumour presentations. To bridge this gap, we propose robustness testing using datasets with intentionally degraded images or artifacts, clinical validation studies to assess model performance in real-world settings, and the development of user-friendly tools that facilitate the integration of segmentation models into clinical workflows. By addressing these limitations and incorporating these suggestions, future iterations of the BraTS challenge can better reflect real-world

clinical needs and contribute to the development of more effective and reliable tools for brain tumour diagnosis and treatment.

3.6. Practical Implications for Clinicians and Researchers

For clinicians, the integration of reliable and accurate segmentation models holds the potential to revolutionise patient care by enabling precise tumour delineation, which is crucial for treatment planning, surgical guidance, and monitoring treatment response. Studies have shown that automated segmentation can reduce the time required for tumour delineation significantly compared to manual segmentation, potentially decreasing clinician workload and minimising the risk of human error. This could lead to faster treatment planning and, potentially, improved patient outcomes. However, several practical hurdles need to be addressed for widespread clinical adoption. These include the need for seamless integration with existing hospital information systems, the development of user-friendly interfaces that do not disrupt established clinical workflows and obtaining necessary regulatory approvals. Moreover, clinicians require models that are not only accurate but also interpretable and trustworthy to confidently integrate them into their decision-making processes.

For researchers, the reviewed literature underscores the need to address the challenges of data scarcity, model generalisation, and computational efficiency. Developing models that perform consistently across diverse patient populations and imaging protocols is essential for facilitating clinical translation. Collaborative efforts between institutions to create larger, more diverse datasets are frequently proposed as a key strategy to improve model robustness and generalizability.

3.7. Future Research Directions

Based on the comprehensive review of the current literature, future research should prioritise the following areas to overcome the identified challenges and accelerate the clinical translation of deep learning-based brain tumour segmentation:

- **Enhancing Model Robustness and Generalization**
  - Domain Adaptation Techniques: The development and implementation of unsupervised or semi-supervised learning techniques to improve model robustness against variations in imaging protocols are crucial. These techniques can help models adapt to new, unseen data distributions without requiring extensive re-training.
  - Multi-Institutional Datasets: A concerted effort towards creating and utilising large, diverse, multi-institutional datasets is essential to train models that generalise well across different imaging settings and patient populations.
- **Leveraging Synthetic Data and Advanced Data Augmentation**
  - Synthetic Data Generation: Further research on employing GANs to create realistic synthetic images is needed to address data scarcity and diversity [78]. Rigorous validation and quality control of synthetic data are crucial to ensure its fidelity and avoid introducing biases.
  - Advanced Augmentation: Applying more sophisticated transformations such as non-linear deformations intensity variations and incorporating prior knowledge about tumour growth patterns can enhance data diversity and improve model robustness.
- **Developing Efficient and Lightweight Architectures**
  - Model Optimization: Exploring and implementing techniques like model pruning, quantisation, and knowledge distillation to reduce computational demands are essential for the practical deployment.
  - Adaptation of Efficient Models: Adapting lightweight architectures like MobileNet [79] or SqueezeNet [80] for 3D data can enable deployment in resource-constrained environments and on-edge devices.
- **Integrating Multi-Modal Data and Clinical Information**
  - Data Fusion Techniques: Combining different MRI modalities (e.g., T1, T2, FLAIR, T1ce) with other imaging data (e.g., PET, CT) and clinical information (e.g., age, genetic markers) can potentially improve segmentation accuracy and provide a more comprehensive understanding of the tumour.

- Personalised Medicine**:** Incorporating genetic markers and patient demographics to tailor models to individual patients represents a promising avenue for developing personalised medicine approaches and potentially improving treatment outcomes.

- **Addressing Class Imbalance**
  - Advanced Loss Functions**:** Implementing and evaluating advanced loss functions, such as focal loss or Dice loss, to focus training on minority classes (e.g., small tumours, tumour subregion) is crucial for improving segmentation performance in these challenging cases.
  - Data Sampling Techniques: Utilizing oversampling or Synthetic Minority Oversampling Technique (SMOTE) can help balance datasets and improve model performance on underrepresented classes.
- **Enhancing Model Explainability and Trustworthiness**
  - Explainable AI Frameworks: Developing methods to visualise and understand model predictions is essential for aiding clinician validation, building trust in the model's outputs, and facilitating clinical adoption.
  - Interpretability Techniques: Incorporating attention mechanisms and saliency maps can make the decision-making processes of deep learning models more transparent and understandable to clinicians
- **Promoting Interdisciplinary Collaboration and Compliance**
  - Collaborative Research: Fostering stronger partnerships between clinicians, researchers, and industry is crucial to align model development with clinical needs, ensure the practical utility of developed solutions, and accelerate the translation of research findings into clinical practice.
  - Ethical and Regulatory Considerations: Addressing patient privacy, data security, and compliance with healthcare regulations is paramount for facilitating clinical integration and ensuring the responsible use of AI in the healthcare.

4. CONCLUSIONS

This review has highlighted the transformative impact of deep learning, particularly 3D U-Net architectures, on brain tumour segmentation. The ability of 3D U-Nets to leverage volumetric data has driven significant advancements, with Dice scores improving from approximately 0.7 in early BraTS challenges to over 0.9 in state-of-the-art methods like nnU-Net. These advancements underscore the critical role of collaborative initiatives, such as the BraTS challenges, in fostering innovation and pushing the boundaries of segmentation accuracy.

Despite these advancements, several critical challenges persist, including tumour heterogeneity, limited access to large, diverse, annotated datasets, variability in imaging protocols, high computational demands, and the risks of overfitting and poor generalization. Addressing these challenges is essential to bridge the gap between technological advancements and their widespread clinical adoption.

The clinical impact of overcoming these obstacles is profound. Accurate segmentation models can enable earlier, more precise diagnoses, facilitate personalized treatment planning, and improve disease monitoring. These advancements hold the potential to enhance survival rates and significantly improve the quality of life for patients battling brain tumours.

In conclusion, while deep learning, particularly through architectures like the 3D U-Net, has revolutionised brain tumour segmentation, the field stands at a critical juncture. To fully realise the transformative potential of AI in neuro-oncology, to accelerate progress, we recommend the following actionable directions for researchers and clinicians:

- Develop robust, generalisable models: Focus on domain adaptation techniques and foster the creation of large, multi-institutional datasets to improve generalisation across diverse populations and imaging protocols.
- Design computationally efficient models: Explore model compression techniques and lightweight architectures to enable deployment in resource-constrained clinical settings.
- Enhance model interpretability and trustworthiness: Integrate attention mechanisms and explainable

AI frameworks to build clinician confidence and encourage adoption in practice.

By addressing these priorities, the research community can foster interdisciplinary collaboration and drive the integration of AI solutions into routine neuro-oncological care. The ultimate goal is to translate these advancements into transformative solutions that redefine brain tumour diagnosis and treatment, improving both survival outcomes and patient quality of life.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this paper.

REFERENCES

1. J.H. Medicine. Brain Tumors and Brain Cancer. https://www.hopkinsmedicine.org/health/conditions-and-diseases/brain-tumor
2. Tie, J., Peng, H. and Zhou, J. (2021) MRI Brain Tumor Segmentation Using 3D U-Net with Dense Encoder Blocks and Residual Decoder Blocks. *Computer Modeling in Engineering & Sciences*, **128**, 427-445. https://doi.org/10.32604/cmes.2021.014107
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., *et al.* (2017) Advancing the Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features. *Scientific Data*, **4**, Article ID: 170117. https://doi.org/10.1038/sdata.2017.117
4. M. B. & Spine. Brain Tumor Diagnosis and Treatment Options. Cincinnati, OH Mayfield Brain & Spine. https://mayfieldclinic.com/pe-braintumor.htm
5. A. A. of N. Surgeons. Brain Tumors. AANS. https://www.aans.org/patients/conditions-treatments/brain-tumors/
6. Kaifi, R. (2023) A Review of Recent Advances in Brain Tumor Diagnosis Based on Ai-Based Classification. *Diagnostics*, **13**, Article No. 3007. https://doi.org/10.3390/diagnostics13183007
7. Martucci, M., Russo, R., Schimperna, F., D'Apolito, G., Panfili, M., Grimaldi, A., *et al.* (2023) Magnetic Resonance Imaging of Primary Adult Brain Tumors: State of the Art and Future Perspectives. *Biomedicines*, **11**, Article No. 364. https://doi.org/10.3390/biomedicines11020364
8. Sabeghi, P., Zarand, P., Zargham, S., Golestany, B., Shariat, A., Chang, M., *et al.* (2024) Advances in Neuro-Oncological Imaging: An Update on Diagnostic Approach to Brain Tumors. *Cancers*, **16**, Article No. 576. https://doi.org/10.3390/cancers16030576
9. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., *et al.* (2016) GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation. In: Crimi, A., *et al.*, Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, 144-155. https://doi.org/10.1007/978-3-319-30858-6_13
10. Ahamed, M.F., Hossain, M.M., Nahiduzzaman, M., Islam, M.R., Islam, M.R., Ahsan, M., *et al.* (2023) A Review on Brain Tumor Segmentation Based on Deep Learning Methods with Federated Learning Techniques. *Computerized Medical Imaging and Graphics*, **110**, Article ID: 102313. https://doi.org/10.1016/j.compmedimag.2023.102313
11. Kaur, D. and Kaur, Y. (2014) Various Image Segmentation Techniques: A Review.
12. Aslam, A., Khan, E. and Beg, M.M.S. (2015) Improved Edge Detection Algorithm for Brain Tumor Segmentation. *Procedia Computer Science*, **58**, 430-437. https://doi.org/10.1016/j.procs.2015.08.057
13. Archana, R. and Jeevaraj, P.S.E. (2024) Deep Learning Models for Digital Image Processing: A Review. *Artificial Intelligence Review*, **57**, Article No. 11. https://doi.org/10.1007/s10462-023-10631-z

14. Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**. http://code.google.com/p/cuda-convnet/

15. Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. 3*rd International Conference on Learning Representations*, *ICLR* 2015, San Diego, 7-9 May 2015. https://arxiv.org/abs/1409.1556v6

16. Szegedy, C., Liu, W., Jia, Y.Q., Sermanet, P., Reed, S., Anguelov, D., *et al.* (2015) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 1-9. https://doi.org/10.1109/cvpr.2015.7298594

17. Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1800-1807. https://doi.org/10.1109/cvpr.2017.195

18. Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI* 2015, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

19. He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

20. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2261-2269. https://doi.org/10.1109/cvpr.2017.243

21. Nguyen, T.T., Nguyen, H., Bui, T., Nguyen-Tat, T.B. and Ngo, V.M. (2024) Brain Tumor Segmentation in MRI Images with 3D U-Net and Contextual Transformer. 2024 *International Conference on Multimedia Analysis and Pattern Recognition* (*MAPR*), Da Nang, 15-16 August 2024, 1-6. https://doi.org/10.1109/mapr63514.2024.10660920

22. Khaliki, M.Z. and Başarslan, M.S. (2024) Brain Tumor Detection from Images and Comparison with Transfer Learning Methods and 3-Layer CNN. *Scientific Reports*, **14**, Article No. 2664. https://doi.org/10.1038/s41598-024-52823-9

23. Ahmed, S.F., Alam, M.S.B., Hassan, M., Rozbu, M.R., Ishtiak, T., Rafa, N., *et al.* (2023) Deep Learning Modelling Techniques: Current Progress, Applications, Advantages, and Challenges. *Artificial Intelligence Review*, **56**, 13521-13617. https://doi.org/10.1007/s10462-023-10466-8

24. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., *et al.* (2015) The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, **34**, 1993-2024. https://doi.org/10.1109/tmi.2014.2377694

25. Pohle, R. and Toennies, K.D. (2001) Segmentation of Medical Images Using Adaptive Region Growing. *SPIE Proceedings*, **4322**, 1337-1346. https://doi.org/10.1117/12.431013

26. Shanthi, K.J. and Kumar, M.S. (2007) Skull Stripping and Automatic Segmentation of Brain MRI Using Seed Growth and Threshold Techniques. 2007 *International Conference on Intelligent and Advanced Systems*, Kuala Lumpur, 25-28 November 2007, 422-426. https://doi.org/10.1109/icias.2007.4658421

27. Manikandan, R., Monolisa, G.S. and Saranya, K. (2013) A Cluster Based Segmentation of Magnetic Resonance Images for Brain Tumor Detection. *Middle-East Journal of Scientific Research*, **14**, 669-672.

28. Abbasi, S. and Tajeripour, F. (2017) Detection of Brain Tumor in 3D MRI Images Using Local Binary Patterns and Histogram Orientation Gradient. *Neurocomputing*, **219**, 526-535. https://doi.org/10.1016/j.neucom.2016.09.051

29. Abdel-Maksoud, E., Elmogy, M. and Al-Awadi, R. (2015) Brain Tumor Segmentation Based on a Hybrid Clustering Technique. *Egyptian Informatics Journal*, **16**, 71-81. https://doi.org/10.1016/j.eij.2015.01.003

30. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., *et al.* (2021) Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *Journal of Big Data*, **8**, 1-74. https://doi.org/10.1186/s40537-021-00444-8

31. Mall, P.K., Singh, P.K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., *et al.* (2023) A Comprehensive Review of Deep Neural Networks for Medical Image Processing: Recent Developments and Future Opportunities. *Healthcare Analytics*, **4**, Article ID: 100216. https://doi.org/10.1016/j.health.2023.100216

32. Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M. and Parmar, M. (2024) A Review of Convolutional Neural Networks in Computer Vision. *Artificial Intelligence Review*, **57**, 1-43. https://doi.org/10.1007/s10462-024-10721-6

33. 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**, 1240-1251. https://doi.org/10.1109/tmi.2016.2538465

34. Krichen, M. (2023) Convolutional Neural Networks: A Survey. *Computers*, **12**, Article No. 151. https://doi.org/10.3390/computers12080151

35. Rosebrock, A. (2021) Convolutional Neural Networks (CNNs) and Layer Types. https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/

36. Tan, H.H. and Lim, K.H. (2019) Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization. 2019 7*th International Conference on Smart Computing & Communications* (*ICSCC*), Sarawak, 28-30 June 2019, 1-4. https://doi.org/10.1109/icscc.2019.8843652

37. Yash (2021) The Challenge of Vanishing/Exploding Gradients in Deep Neural Networks. https://www.analyticsvidhya.com/blog/2021/06/the-challenge-of-vanishing-exploding-gradients-in-deep-neural-networks/

38. Fred, A. and Agarap, M. (2018) Deep Learning Using Rectified Linear Units (RELU). 1-6.

39. Xu, J., Li, Z., Du, B., Zhang, M. and Liu, J. (2020) Reluplex Made More Practical: Leaky ReLU. 2020 *IEEE Symposium on Computers and Communications* (*ISCC*), Rennes, 7-10 July 2020, 1-7. https://doi.org/10.1109/iscc50000.2020.9219587

40. Lederer, J. (2021) Activation Functions in Artificial Neural Networks: A Systematic Overview. https://arxiv.org/abs/2101.09957v1

41. Ramachandran, P., Zoph, B. and Le, Q.V. (2017) Searching for Activation Functions.

42. Basirat, M. and Roth, P.M. (2018) The Quest for the Golden Activation Function. http://arxiv.org/abs/1808.00783v1

43. Pratiwi, H., Windarto, A.P., Susliansyah, S., Aria, R.R., Susilowati, S., Rahayu, L.K., *et al.* (2020) Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks. *Journal of Physics*: *Conference Series*, **1471**, Article ID: 012010. https://doi.org/10.1088/1742-6596/1471/1/012010

44. Salih, M.M., Salih, M.E. and Ahmed, M.A.A. (2019) Enhancement of U-Net Performance in MRI Brain Tumour Segmentation Using HardELiSH Activation Function. 2019 *International Conference on Computer*, *Control*, *Electrical*, *and Electronics Engineering* (ICCCEEE), Khartoum, 21-23 September 2019, 1-5. https://doi.org/10.1109/iccceee46830.2019.9071235

45. Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 431-440. https://doi.org/10.1109/cvpr.2015.7298965

46. Rguibi, Z., Hajami, A., Zitouni, D., Elqaraoui, A. and Bedraoui, A. (2022) CXAI: Explaining Convolutional Neural Networks for Medical Imaging Diagnostic. *Electronics*, **11**, Article No. 1775. https://doi.org/10.3390/electronics11111775

47. Avesta, A., Hossain, S., Lin, M., Aboian, M., Krumholz, H.M. and Aneja, S. (2023) Comparing 3D, 2.5D, and 2D Approaches to Brain Image Auto-Segmentation. *Bioengineering*, **10**, Article No. 181. https://doi.org/10.3390/bioengineering10020181

48. Zhang, Y., Liao, Q., Ding, L. and Zhang, J. (2022) Bridging 2D and 3D Segmentation Networks for Computation-Efficient Volumetric Medical Image Segmentation: An Empirical Study of 2.5D Solutions. *Computerized Medical Imaging and Graphics*, **99**, Article ID: 102088. https://doi.org/10.1016/j.compmedimag.2022.102088

49. Yu, Q., Xia, Y., Xie, L., Fishman, E.K. and Yuille, A.L. (2019) Thickened 2D Networks for Efficient 3D Medical Image Segmentation. http://arxiv.org/abs/1904.01150

50. Sun, Y., Hsieh, A., Fang, S., Wu, H., Kao, L., Chung, W., *et al.* (2021) Can 3D Artificial Intelligence Models Outshine 2D Ones in the Detection of Intracranial Metastatic Tumors on Magnetic Resonance Images? *Journal of the Chinese Medical Association*, **84**, 956-962. https://doi.org/10.1097/jcma.0000000000000614

51. Taha, A.A. and Hanbury, A. (2015) Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Medical Imaging*, **15**, Article No. 29. https://doi.org/10.1186/s12880-015-0068-x

52. Sensitivity and Specificity—Wikipedia. https://en.wikipedia.org/wiki/Sensitivity_and_specificity

53. Precision and Recall—Wikipedia. https://en.wikipedia.org/wiki/Precision_and_recall

54. 10.1 Sensitivity and Specificity—Foundations of Biomedical Science: Quantitative Literacy: Theory and Problems. https://oercollective.caul.edu.au/foundations-of-biomedical-science/chapter/10-1-sensitivity-and-specificity/

55. Thada, V. and Jaglan, V. (2013) Comparison of Jaccard, Dice, Cosine Similarity Coefficient to Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm. *International Journal of Innovations in Engineering and Technology*, **2**, 202-205.

56. Aydin, O.U., Taha, A.A., Hilbert, A., Khalil, A.A., Galinovic, I., Fiebach, J.B., *et al.* (2021) On the Usage of Average Hausdorff Distance for Segmentation Performance Assessment: Hidden Error When Used for Ranking. *European Radiology Experimental*, **5**, Article No. 4. https://doi.org/10.1186/s41747-020-00200-2

57. Fernando, K.R.M. and Tsokos, C.P. (2023) Deep and Statistical Learning in Biomedical Imaging: State of the Art in 3D MRI Brain Tumor Segmentation. *Information Fusion*, **92**, 450-465. https://doi.org/10.1016/j.inffus.2022.12.013

58. Montaha, S., Azam, S., Rakibul Haque Rafid, A.K.M., Hasan, M.Z. and Karim, A. (2023) Brain Tumor Segmentation from 3D MRI Scans Using U-Net. *SN Computer Science*, **4**, Article No. 386. https://doi.org/10.1007/s42979-023-01854-6

59. Ullah, F., Ansari, S.U., Hanif, M., Ayari, M.A., Chowdhury, M.E.H., Khandakar, A.A., *et al.* (2021) Brain MR Image Enhancement for Tumor Segmentation Using 3D U-Net. *Sensors*, **21**, Article No. 7528. https://doi.org/10.3390/s21227528

60. Feng, X., Tustison, N.J., Patel, S.H. and Meyer, C.H. (2020) Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features. *Frontiers in Computational Neuroscience*, **14**, Article ID: 488825. https://doi.org/10.3389/fncom.2020.00025

61. Henry, T., Carré, A., Lerousseau, M., Estienne, T., Robert, C., Paragios, N., *et al.* (2021) Brain Tumor Segmentation with Self-Ensembled, Deeply-Supervised 3D U-Net Neural Networks: A Brats 2020 Challenge Solution. In: Crimi, A. and Bakas, S., Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, 327-339. https://doi.org/10.1007/978-3-030-72084-1_30

62. Ballestar, L.M. and Vilaplana, V. (2021) MRI Brain Tumor Segmentation and Uncertainty Estimation Using 3D-

Unet Architectures. In: Crimi, A. and Bakas, S., Eds., *Brainlesion*: *Glioma*, *Multiple Sclerosis*, *Stroke and Traumatic Brain Injuries*, Springer International Publishing, 376-390. https://doi.org/10.1007/978-3-030-72084-1_34

63. Wang, F., Jiang, R., Zheng, L., Meng, C. and Biswal, B. (2020) 3D U-Net Based Brain Tumor Segmentation and Survival Days Prediction. In: Crimi, A. and Bakas, S., Eds., *Brainlesion*: *Glioma*, *Multiple Sclerosis*, *Stroke and Traumatic Brain Injuries*, Springer International Publishing, 131-141. https://doi.org/10.1007/978-3-030-46640-4_13
64. Sun, L., Zhang, S., Chen, H. and Luo, L. (2019) Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans with Deep Learning. *Frontiers in Neuroscience*, **13**, Article ID: 468066. https://doi.org/10.3389/fnins.2019.00810
65. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., *et al*. (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
66. Kamnitsas, K., *et al*. (2017) Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. http://arxiv.org/abs/1711.01468
67. Myronenko, A. (2019) 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi, A., *et al*., Eds., *Brainlesion*: *Glioma*, *Multiple Sclerosis*, *Stroke and Traumatic Brain Injuries*, Springer International Publishing, 311-320. https://doi.org/10.1007/978-3-030-11726-9_28
68. Jiang, Z., Ding, C., Liu, M. and Tao, D. (2020) Two-Stage Cascaded U-Net: 1st Place Solution to Brats Challenge 2019 Segmentation Task. In: Crimi, A. and Bakas, S., Eds., *Brainlesion*: *Glioma*, *Multiple Sclerosis*, *Stroke and Traumatic Brain Injuries*, Springer International Publishing, 231-241. https://doi.org/10.1007/978-3-030-46640-4_22
69. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P. and Maier-Hein, K.H. (2021) nnU-Net for Brain Tumor Segmentation. In: Crimi, A. and Bakas, S., Eds., *Brainlesion*: *Glioma*, *Multiple Sclerosis*, *Stroke and Traumatic Brain Injuries*, Springer International Publishing, 118-132. https://doi.org/10.1007/978-3-030-72087-2_11
70. Luu, H.M. and Park, S. (2022) Extending nn-UNet for Brain Tumor Segmentation. In: Crimi, A. and Bakas, S., Eds., *Brainlesion*: *Glioma*, *Multiple Sclerosis*, *Stroke and Traumatic Brain Injuries*, Springer International Publishing, 173-186. https://doi.org/10.1007/978-3-031-09002-8_16
71. Zeineldin, R.A., Karar, M.E., Coburger, J., Wirtz, C.R. and Burgert, O. (2020) Deepseg: Deep Neural Network Framework for Automatic Brain Tumor Segmentation Using Magnetic Resonance FLAIR Images. *International Journal of Computer Assisted Radiology and Surgery*, **15**, 909-920. https://doi.org/10.1007/s11548-020-02186-z
72. Ferreira, A., *et al*. (2024) How We Won BraTS 2023 Adult Glioma Challenge? Just Faking It! Enhanced Synthetic Data Augmentation and Model Ensemble for Brain Tumour Segmentation. https://arxiv.org/abs/2402.17317v2
73. Biswas, A., Md Abdullah Al, N., Imran, A., Sejuty, A.T., Fairooz, F., Puppala, S., *et al*. (2023) Generative Adversarial Networks for Data Augmentation. In: Zheng, B., *et al*., Eds., *Data Driven Approaches on Medical Imaging*, Springer Nature, 159-177. https://doi.org/10.1007/978-3-031-47772-0_8
74. Types of GANs for Medical Imaging. https://www.researchgate.net/publication/386451967_Types_of_GANs_for_Medical_Imaging
75. MRI Cross-Modality Image-to-Image Translation with CycleGAN|by Aditya Kakde|Medium. https://adityakakde.medium.com/mri-cross-modality-image-to-image-translation-with-cyclegan-7720a4c786b2
76. Hu, Q., Wei, Y., Pang, J. and Liang, M. (2024) Unsupervised Domain Adaptation for Brain Structure Segmentation via Mutual Information Maximization Alignment. *Biomedical Signal Processing and Control*, **90**, Article ID: 105784. https://doi.org/10.1016/j.bspc.2023.105784
77. Huang, L., Zhu, E., Chen, L., Wang, Z., Chai, S. and Zhang, B. (2022) A Transformer-Based Generative

Adversarial Network for Brain Tumor Segmentation. *Frontiers in Neuroscience*, **16**, Article ID: 1054948. https://doi.org/10.3389/fnins.2022.1054948

78. Sille, R., Choudhury, T., Sharma, A., Chauhan, P., Tomar, R. and Sharma, D. (2023) A Novel Generative Adversarial Network-Based Approach for Automated Brain Tumour Segmentation. *Medicina*, **59**, Article No. 119. https://doi.org/10.3390/medicina59010119
79. Kumar, M.P., Hasmitha, D., Usha, B., Jyothsna, B. and Sravya, D. (2024) Brain Tumor Classification Using MobileNet. 2024 *International Conference on Integrated Circuits and Communication Systems* (*ICICACS*), Raichur, 23-24 February 2024, 1-7. https://doi.org/10.1109/icicacs60521.2024.10499117
80. Baig, M.D., Haq, H.B.U., Akram, W. and Awan, A.M. (2024) Brain Tumor Detection Enhanced with Transfer Learning Using SqueezeNet. *Decision Making Advances*, **2**, 129-141. https://doi.org/10.31181/dma21202432