

Article

Multimodal Stereotactic Brain Tumor Segmentation Using 3D-Znet

Mohammad Ashraf Ottom ^{1,2,*} , Hanif Abdul Rahman ^{1,3} , Iyad M. Alazzam ² and Ivo D. Dinov ^{1,*} 

¹ Statistics Online Computational Resource, University of Michigan, Ann Arbor, MI 48104, USA

² Department of Information Systems, Yarmouk University, Irbid 21163, Jordan

³ PAPRSB Institute of Health Sciences, Universiti Brunei Darussalam, Gadong BE1410, Brunei

* Correspondence: ottom.ma@yu.edu.jo (M.A.O.); statistics@umich.edu (I.D.D.)

Abstract: Stereotactic brain tumor segmentation based on 3D neuroimaging data is a challenging task due to the complexity of the brain architecture, extreme heterogeneity of tumor malformations, and the extreme variability of intensity signal and noise distributions. Early tumor diagnosis can help medical professionals to select optimal medical treatment plans that can potentially save lives. Artificial intelligence (AI) has previously been used for automated tumor diagnostics and segmentation models. However, the model development, validation, and reproducibility processes are challenging. Often, cumulative efforts are required to produce a fully automated and reliable computer-aided diagnostic system for tumor segmentation. This study proposes an enhanced deep neural network approach, the 3D-Znet model, based on the variational autoencoder–autodecoder Znet method, for segmenting 3D MR (magnetic resonance) volumes. The 3D-Znet artificial neural network architecture relies on fully dense connections to enable the reuse of features on multiple levels to improve model performance. It consists of four encoders and four decoders along with the initial input and the final output blocks. Encoder–decoder blocks in the network include double convolutional 3D layers, 3D batch normalization, and an activation function. These are followed by size normalization between inputs and outputs and network concatenation across the encoding and decoding branches. The proposed deep convolutional neural network model was trained and validated using a multimodal stereotactic neuroimaging dataset (BraTS2020) that includes multimodal tumor masks. Evaluation of the pretrained model resulted in the following dice coefficient scores: Whole Tumor (WT) = 0.91, Tumor Core (TC) = 0.85, and Enhanced Tumor (ET) = 0.86. The performance of the proposed 3D-Znet method is comparable to other state-of-the-art methods. Our protocol demonstrates the importance of data augmentation to avoid overfitting and enhance model performance.

Keywords: deep learning; 3D tumor segmentation; encoder–decoder; Znet; multimodal neuroimaging data



Citation: Ottom, M.A.; Abdul Rahman, H.; Alazzam, I.M.; Dinov, I.D. Multimodal Stereotactic Brain Tumor Segmentation Using 3D-Znet. *Bioengineering* **2023**, *10*, 581. <https://doi.org/10.3390/bioengineering10050581>

Academic Editors: Wei Zhang, Feng Gu, Xuan Guo, Sheng Li, Qiushi Fu and Yunfeng Wu

Received: 6 March 2023

Revised: 12 April 2023

Accepted: 19 April 2023

Published: 11 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Contemporary deep learning techniques are widely used in many fields such as agriculture, self-driving cars, fraud detection, and healthcare [1–4]. Adequate attention to deep learning applications in healthcare emerged recently, including brain tumor diagnostics, detection, and segmentation [5–7]. Massive amounts of valuable, multi-source, spatiotemporal, and multiscale data have recently become available in many applied, theoretical, experimental, data science, and healthcare domains [8]. Early detection, accurate prognostication, and precise tracking of diseases contribute heavily to saving lives, finding optimal treatments, and reducing the economic burden for patients and healthcare systems. Similar benefits of machine learning and artificial intelligence are applied to studies of brain tumors and neuro-oncology [9]. In this work, we will present a deep convolutional neural network (CNN) approach, 3D-Znet, to learn the neuroimaging affinities and segment prospective brain tumors using publicly available datasets, BraTS (2020). The 3D-Znet approach is

originally inspired by the variational encoder–decoder framework and the skip connection concept to enable the model to reuse features on multiple levels. The model was evaluated using the BraTS (2020) datasets. Assessment of the proposed approach indicates high overall mean dice coefficient scores for whole tumor (0.91), tumor core (0.85), and enhanced tumor (0.86) masks. The augmentation of the original data sample and appropriate data preprocessing provided a performance boost and enhanced the model predictions.

High-resolution stereotactic medical imaging is critical for patients with brain tumors. Such non-invasive data acquisition protocols rapidly evolve over time, and this drives the development of powerful AI techniques that transform these 3D imaging scans into actionable information and knowledge that can improve the care of patients. Working with medical images such as computed tomography (CT) and magnetic resonance imaging (MRI) is challenging due to data complexity, the large size of the data, and the variability of coordinate representation systems and storage formats [10]. In practice, it is essential to understand the data and medical coordinate systems before applying deep learning techniques, e.g., anatomical and voxel (3D volume-element corresponding to 2D pixel elements) coordinate frameworks. Figure 1A shows the anatomical coordinate system (ACS). ACS comprises three cardinal projection planes illustrating the basic anatomical position of organs in the human body as objects in a solid dense 3D scene. The planes describe the MRI volume orientation. Sagittal cross-sections show 2D image projections of the volume from a side angle (traversing left-to-right direction), the coronal plane from the front view (traversing back-to-front), and the axial plane reveals images as taken from the top down (transverse traversing from top-to-bottom), and Figure 1B shows the brain voxel coordinate system [11–14]. Deep learning systems provide thriving environments for computer vision and semantic segmentation applications. Variational autoencoder–autodecoders are deep learning networks commonly used to translate an input x into an output $r(x)$ via a two-step process—encoder and decoder. Figure 2 shows a demonstration of an autoencoder

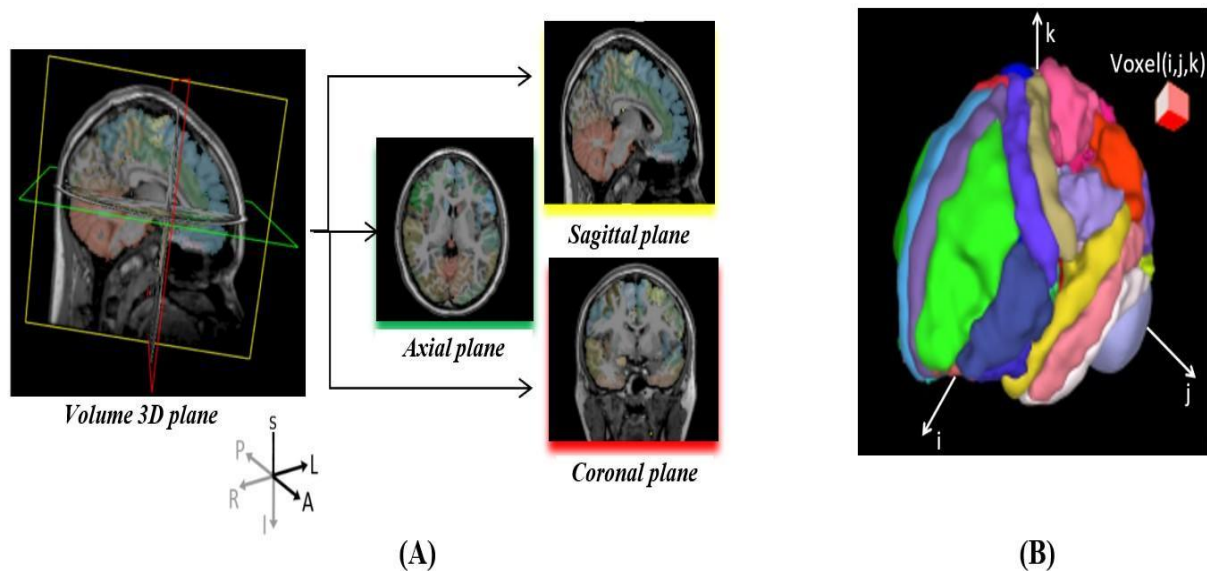


Figure 1. Part (A) is the anatomical coordinate system (ACS); the sagittal plane is vertical to the ground, traversing from right (R) to left (L). A coronal plane is vertical to the ground, perpendicular to the sagittal plane and spanning from anterior (A) to the posterior (P) part of the brain. The transverse axial plane is horizontal, traversing from superior (S) to the inferior (I) part. Part (B) is the voxel coordinate system with i , j , and k coordinates of a point [15,16].

DNN framework. In the encoding phase, the network takes an input x and lowers the representation (size) of x as it moves through the network layers. This process continues until a certain bottleneck is reached, where the output represents a small feature representation. The subsequent decoder phase does the opposite inflationary process by increasing the feature representation to produce an output $r(x)$ with the same dimension as the initial network input x [17,18]. In principle, depending on the application, the output layer may generate objects of size either smaller or larger than the size of the initial input. For example, for producing a higher resolution image reconstruction (super-resolution) from input size 32×32 , an additional layer can be added at the end of the network so that the decoder outputs a larger image of size 64×64 . However, for semantic segmentation, the output size is usually equal to the input size [19].

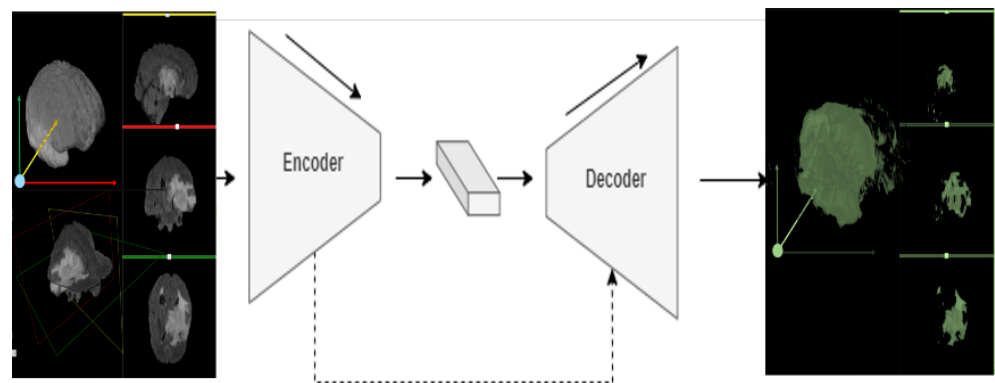


Figure 2. A schematic of an encoder–decoder architecture that maps inputs to outputs preserving the dimensions of the imaging input and the semantic segmentation output. Skip connections are used in some architectures to transfer feature representation from encoder to decoder such as in the Unet [20] and Vnet [21] CNN models.

Deep convolutional neural networks (DCNN) are neural networks based on artificial neurons that are structured into layers. The network layers are connected using virtual edges carrying model weights that are computationally estimated during the DCNN model fitting. The initial layer is called the input layer, and the final layer is the output layer. Intermediate hidden layers, located between input and output layers, recursively transform the feature space of one layer to the next. CNNs contain convolutional layers and non-convolutional layers. Convolutional layers include a kernel (filter) to extract features from the preceding input. Patterns are iteratively learned by sliding the filter over the preceding input and calculating the dot-product of the filter and the prior input; a process called (kernel) convolution, and the result of convolving is called a feature map. In the early stages of the CNN fitting, feature maps contain basic patterns, such as edges and corners (basic building blocks). In contrast, later feature maps deeper into the network layers expose more refined patterns (details) that contribute to forming the final output. Among a series of convolutional layers, the feature maps down-sample the images using pooling operations. CNNs are adaptive and can be used to obtain solutions for various types of data in multiple dimensions. Specifically, 2D convolution is used for 2D imaging data indexed in the height and width dimensions and representing scalar or vector intensities, such as gray-scale or RGB images. Higher-dimensional (hyper) volumes, such as 3D solids, require 3D and appropriate higher-dimensional convolutions that are suitable for data of the given dimension (e.g., height, width, and depth for 3D volumes such as MR images). Figure 3 shows the architectural difference between 2D and 3D convolutions [22–24].

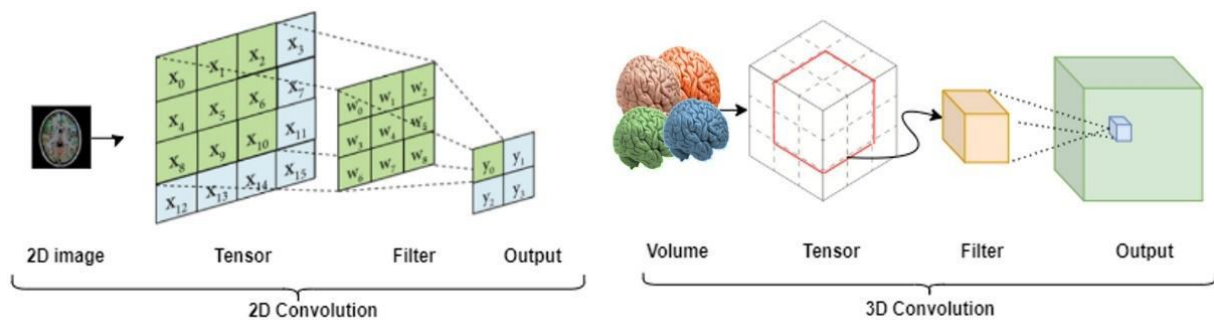


Figure 3. Schematics of 2D and 3D convolution layers; 2D convolution is suitable for classical pixel images, such as CT scans, whereas 3D convolution is used for stereotactic volumetric data, such as CT and MR images [22–24].

The study introduces a new approach for segmenting 3D MR volumes using a deep neural network called the 3D-Znet model. This model is an improvement of the variational autoencoder–autodecoder Znet method and uses fully dense connections to improve its performance. The architecture includes four encoders and four decoders, each containing double convolutional 3D layers, 3D batch normalization, and an activation function. The model was trained and validated using the BraTS2020 dataset, which contains multimodal tumor masks. The evaluation of the model showed that it performed well with dice coefficient scores of 0.91 for Whole Tumor, 0.85 for Tumor Core, and 0.86 for Enhanced Tumor, which are comparable to other state-of-the-art methods. The study also emphasizes the importance of data augmentation in enhancing the model’s performance and avoiding overfitting. The rest of the paper is organized as follows: First, we give an overview of the most related work in the field of medical images segmentation, secondly, we describe the methodology, preprocessing phase, the architecture of the proposed model, the training phase, and the evaluation methods. Thirdly, we discuss the experimental results and a comparison with previous results in the literature, and finally is the conclusion and future work.

2. Related Work

In their study [25], Karayegen and Aksahin proposed a convolutional neural network approach to diagnose and segment 3D brain tumors using a deep neural network pre-trained on the 2020 Brain Tumor Segmentation (BraTS2020) dataset [26]. The authors had normalised the dataset into two classes (background and tumor), however, the dataset has four class categories. During data preprocessing, the authors used histogram equalization to enhance the classifications of edges in each region. To solve memory issues and enhance training performance, they used a random patches mechanism (80 and 90 patches) of sizes $36 \times 36 \times 155$ and $40 \times 40 \times 155$. The evaluation results showed an ability to diagnose and segment tumors with promising results.

Another study by [27] attempts 3D brain tumor segmentation using SegNet algorithm on BraTS dataset. In this study, the investigators trained all the modules separately and integrated them during the post-processing. The four feature maps fused to form one feature map and then a decision tree algorithm was used to classify the output into malignant and benign. Results and evaluation showed a potential for this approach for brain tumor segmentation. A recent report [28] proposed a fusion deep learning called RMU-Net model for 3D semantic segmentation of BraTS datasets. The model is motivated by U-net and MobileNetV2. RMU-Net’s training time is higher than other well known segmentation models, however, the model produces promising results. Another recent study [29] proposed the cascaded V-Nets approach for brain tumor segmentation for multi-modal brain MR imaging. V-Net is considered as a well-performing approach in semantic segmentation using a cascaded structure and ensemble method to enhance segmentation results. The model architecture consists of encoder, decoder, and skip connections. The approach also suggests segmenting the whole tumor first and then splitting the output

into edema, enhancing tumor, and necrosis. The architecture was trained on the BraTS data and validated independently using local hospitals datasets showing considerable improvements in quality of tumor segmentation. The approach of an independent previous study [29] relied on the usage of prior knowledge, training the model jointly on 3D and 2D data, using ensembling methodology, and introducing post-processing to gain better tumor segmentation. The authors utilized three UNets with distinct inputs, then ensemble the equivalent three outputs, and finally applied the post-processing techniques. The first Unet network used 3D patches of multimodal MR images, the second UNet employed brain parcellation as an extra input, and the last network used 2D slices of multimodal MR images. Then brain parcellation and probability maps for each class from the prior network were obtained and tested using BraTS (2018), BraTS (2020), and other local datasets. The final results for this approach were promising, however, compared to other methods, training time is substantially increased due to using multiple Unet DCNNs. A recent study [30] proposed a brain tumor segmentation method based on an ensemble of 3D U-Nets with different hyper-parameters trained on non-uniformly extracted patches. They created a brain tumor segmentation method using an ensemble of 3D U-Nets. Six networks with varying numbers of encoding/decoding blocks, patch sizes, and loss weights were trained and ensemble by averaging the final prediction probabilities. The ensemble model outperformed any of the single models in terms of results. However, the ensemble method requires extensive computational power and is time-consuming. Moreover, [31] proposed a novel transformer-based method for 3D medical image segmentation. The method is effective at extracting local and global characteristics; in addition, the authors designed a combination of transformer structure and CNN, as well as an ETrans (Enhanced Transformer) model, to enhance detail feature extraction. This model was used to extract local detailed features, allowing the model to perform well in segmenting categories that occupy a small portion of the image. However, due to the extensive use of the transformer structure, the performance when segmenting the edges was insufficient. Another study [32] used magnetic resonance images (MRI) to classify images of Alzheimer's disease (AD) using deep convolutional neural networks (CNN) involving CNN and transfer learning (Visual Geometry Group (VGG)16 and VGG19). Images of Alzheimer's disease were divided into four categories by neurologists, and the results were assessed using a range of metrics, where VGG-19 was the best in three categories. The research in [33] proposed a machine-learning diagnostic system for COVID-19. For a quicker and more accurate detection of possible COVID-19 instances, four machine learning algorithms—Random Forest (RF), XGBoost, and Light Gradient Boosting Machine (LGBM)—were applied. The dataset utilized the pertinent symptoms for the identification of a suspicious person from COVID-19 symptoms. The results showed that real-time data capturing can efficiently diagnose COVID-19 patients. A recent study [34] suggested a three-stage approach to address brain tumor segmentation. First, a morphological operation pre-processing is applied to remove the skull bone from the image. Then, the particle swarm optimization (PSO) algorithm with a two-way fixed-effects analysis of variance (ANOVA)-based fitness function is utilized to find the optimal block containing the brain lesion. Finally, the K-means clustering algorithm is used to distinguish the detected block as tumor or non-tumor. The study used the BraTS 2015 database and their private dataset from Kouba imaging center-Algiers (KICA), which showed the model's capability to segment brain tumors.

Despite the availability of various methods and algorithms for brain tumor segmentation, achieving high accuracy in detecting the tumor area and distinguishing it from healthy brain tissue remains challenging and requires further investigation, and the development of methods that are more efficient and effective in detecting and segmenting brain tumors will contribute to the field, especially in cases where the tumor size is small or the tumor is located in a complex area of the brain.

3. Methods

3.1. Dataset and Pre-Processing

In our study, we used the BraTS 2020 dataset to evaluate the performance of the 3D-Znet model [34,35]. The dataset consists of multi-modal MR images for 369 patients that were captured using two methods of segmentation consisting of Gross total resection ($n = 359$) and Subtotal resection ($n = 10$). Two types of brain tumor neuroimaging data were available, which consisted of high-grade gliomas ($n = 237$) and low-grade gliomas ($n = 132$). The meta-data provides two characteristics of the patients from whom the scans were extracted—the survival (days) and age (years). Using linear regression analysis [2], it was observed that there was a significant downward relationship between length of survival and age. A 10-year increase in age decreased the length of survival by 118 days ($b = -118$; 95% CI: $[-160, -70]$; $p < 0.001$). These results reflect the importance of this study to develop a better detection tool for early detection to improve the survival rate of patients with brain tumors.

The data (multimodal 3D Brats 2020 scans) were compiled from 19 contributing sites [35] and are available in compressed neuroimaging NIFTI file format (.nii.gz) [36]. For each patient, there are five volumes, and each volume dimension is $240 \times 240 \times 155$ representing height, width, and depth, respectively: fluid-attenuated inversion recovery (Flair), T1-weighted (T1), contrast T1-weighted (T1ce), T2-weighted (T2), and ground truth segmentation (seg). The ground truth volume (seg) has four distinct pixel values: no tumor (the value of 0), non-enhancing tumor core NET (the value of 1), peritumoral edema ED (the value of 2), enhancing tumor ET (the value of 4), and the value of 3 represents missing label. The ground truth was segmented manually by one-to-four expert neuro-radiologists [26,36,37]. Figure 4 shows the sample data volume from the BraTs dataset using Nilearn (Statistics for NeuroImaging in python) [38]. The raw or compressed NIFTI files can also be easily displayed using the SOCR BrainViewer webapp (<https://socr.umich.edu/HTML5/BrainViewer/>) (accessed on 6 March 2023).

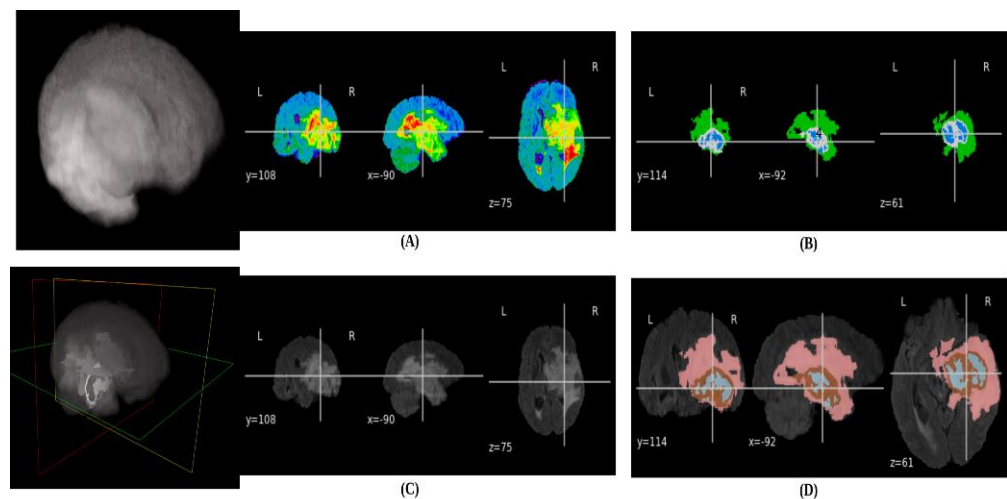


Figure 4. (A) Cardinal projection cross sections of the Echo Planar Imaging (EPI) data: Coronal, Axial, and Sagittal planes of a sample Flair volume. (B) EPI plots: Coronal, Axial, and Sagittal for a ground truth volume (mask). (C) Anatomical plots of Coronal, Axial, and Sagittal for a sample Flair volume. (D) Region of Interest overlap between ground truth and Flair volumes: Coronal, Axial, and Sagittal cross sectional planes.

The brain MR images were collected from 19 institutions. This site heterogeneity requires pre-processing to establish corresponding homologies between datasets from different locations and enhance model performance. The 3D volumes were co-registered to a standard anatomical template (an atlas) of an exact resolution of $1 \times 1 \times 1 \text{ mm}^3$, and then skull-stripped [34] to remove extra-cerebral tissue. Min-max normalization was used to temper intensity variation and to scale the intensity to a uniform scale between 0 and 1. We also reshaped the original volumes dimension from $240 \times 240 \times 155$ to $128 \times 128 \times 128$, which represent the stereotactic height, width, and depth (slices) dimensions, respectively. Resizing can affect model accuracy; however, it was performed to meet the available hardware resources and to minimize training time.

Data augmentation is used to produce more samples of data from the available dataset using techniques to modify the existing data, such as flipping images, zooming in and out, rotating by a certain angle, or using more complex synthetic algorithms. We used rotation by angles -5° to 15° degrees to expand the training data from 369 to 1845 volumes, where 60% were used for training and the rest for testing. Figure 5 shows a sample volume before and after data augmentation. The segmented volume (ground truth) was used to generate three volumes: Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC). WT volume is a copy of ground truth volume, where pixel values are a combination of nonenhancing tumor core NET (1), peritumoral edema ED (2), enhancing tumor ET (4), and the values of the remaining pixels are non-tumor (value 0). ET is another copy of the segmented volume, where pixel values correspond to enhancing tumor ET (4) and the rest of the pixel values correspond to non-tumor (value 0). The third copy of the segmented volume (TC) has the values for non-enhancing tumor core NET (1), enhancing tumor ET (4) and, the rest are 0 (non-tumor). Finally, the three generated volumes are stacked together to form one multi-modal tensor of dimension $(3 \times 128 \times 128 \times 128)$. On the other hand, actual image volumes fluid-attenuated inversion recovery (Flair), T1-weighted (T1), contrast T1-weighted (T1ce), and T2-weighted (T2) are stacked together to form one multi-modal volume with the dimension of $(4 \times 128 \times 128 \times 128)$. Figure 6 shows the entire workflow pipeline demonstrating the data preprocessing, model fitting, Znet assessment, and result reporting protocol.

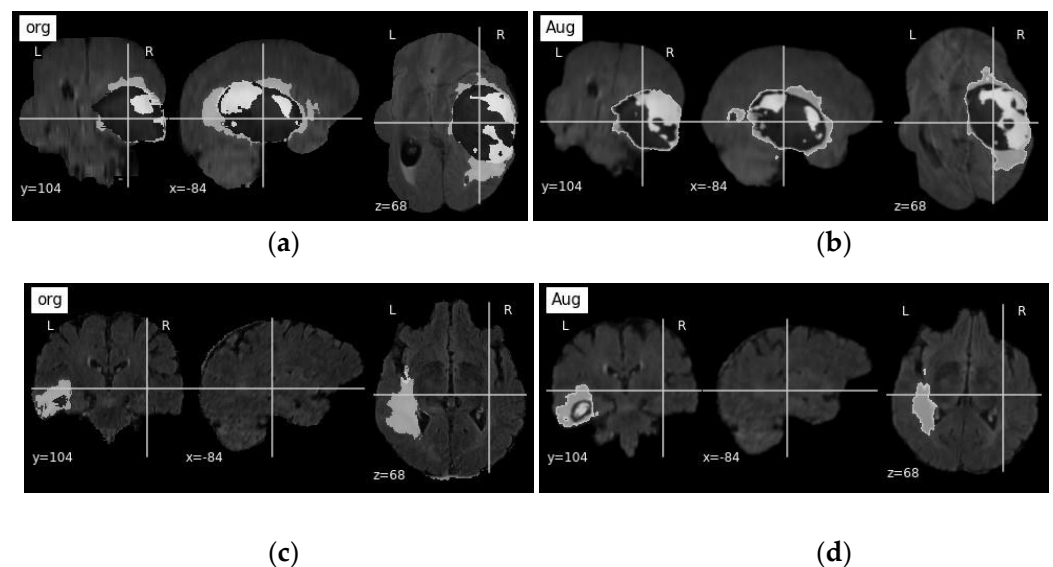


Figure 5. Sample of data augmentation using rotation technique of 10 and -5 degrees: (a) original volume; (b) augmented volume rotated by 10 degrees; (c) original volume; and (d) augmented volume rotated by -5 degrees.

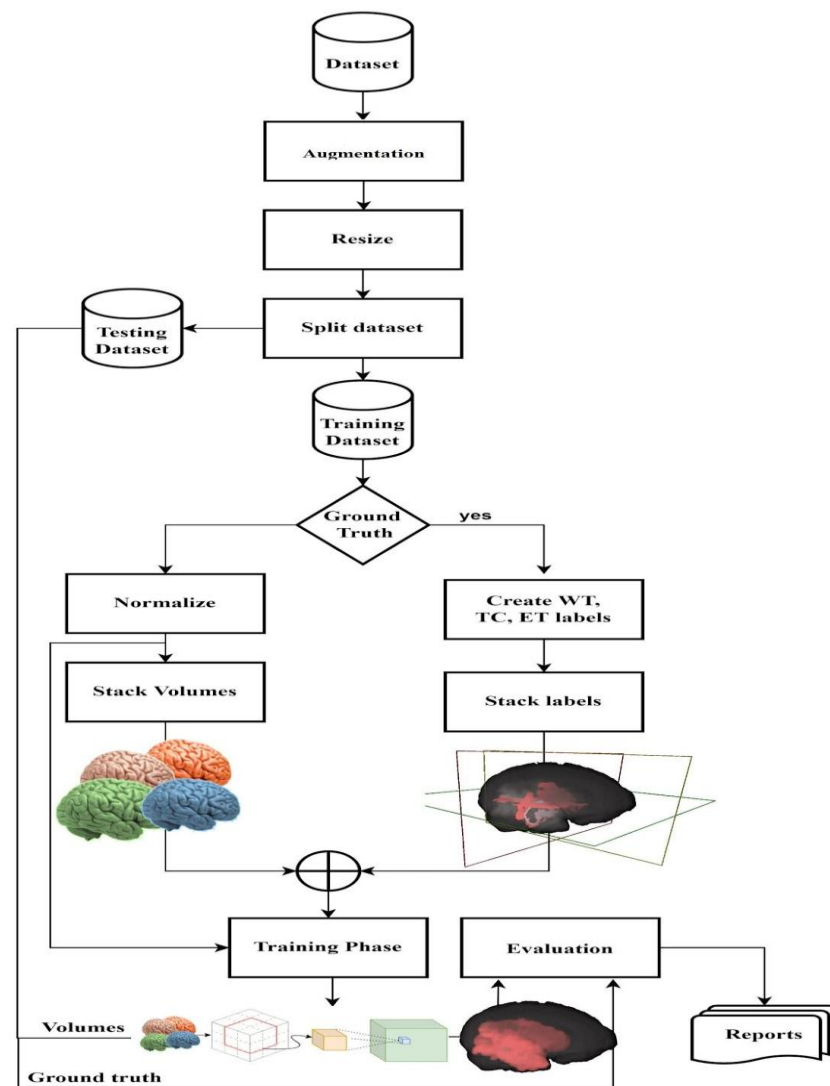


Figure 6. Flowchart of the proposed end-to-end Znet pipeline workflow protocol.

3.2. 3D-Znet Architecture

The prior 2D-Znet model encoder–decoder framework [39] inspired the new 3D-Znet architecture, which is used for stereotactic (3D) neuroimaging volumes, such as multi-modal MR images. The 3D-Znet architecture relies on fully connected connections (dense connections), which is very powerful in biomedical applications, segmentation, and prediction [40].

The objective of dense connections is to enable the model to reuse features on multiple levels to improve model performance [41], where every block of input layers is densely connected to the subsequent block of nodes in the next layer. 3D-Znet incorporates four encoders and four decoders paired with input and output blocks. Each block of the encoder–decoder consists of double convolutional 3D layers, 3D batch normalization, and an activation function (ReLU). These are followed by size normalization between inputs and outputs to facilitate network concatenation and to produce the inputs for the subsequent encoder–decoder blocks. Encoder blocks differ from decoder blocks by using the 3D-maxpooling to downsample the input along its width, height, and depth. In contrast, the decoder block utilizes upsampling to generate super-resolution inputs and retain the original volumetric dimensions at the last decoder block. Conv3d uses a 3D convolution over an input tensor. In the simplest case, the output value of the conv3d with input size (N, C_{in}, D, H, W) is a tensor sized $(N, C_{out}, D_{out}, H_{out}, W_{out})$, where N is the batch size, C is the number of channels, D is the depth, H , and W are highest and width, respectively.

The conv3d parameters are kernel_size = 3, stride = 1, and padding = 1. The role of batch normalization is to make the training quicker and more stable by re-centering and re-scaling the input tensor. The overall 3D-Znet architecture is illustrated in Figure 7.

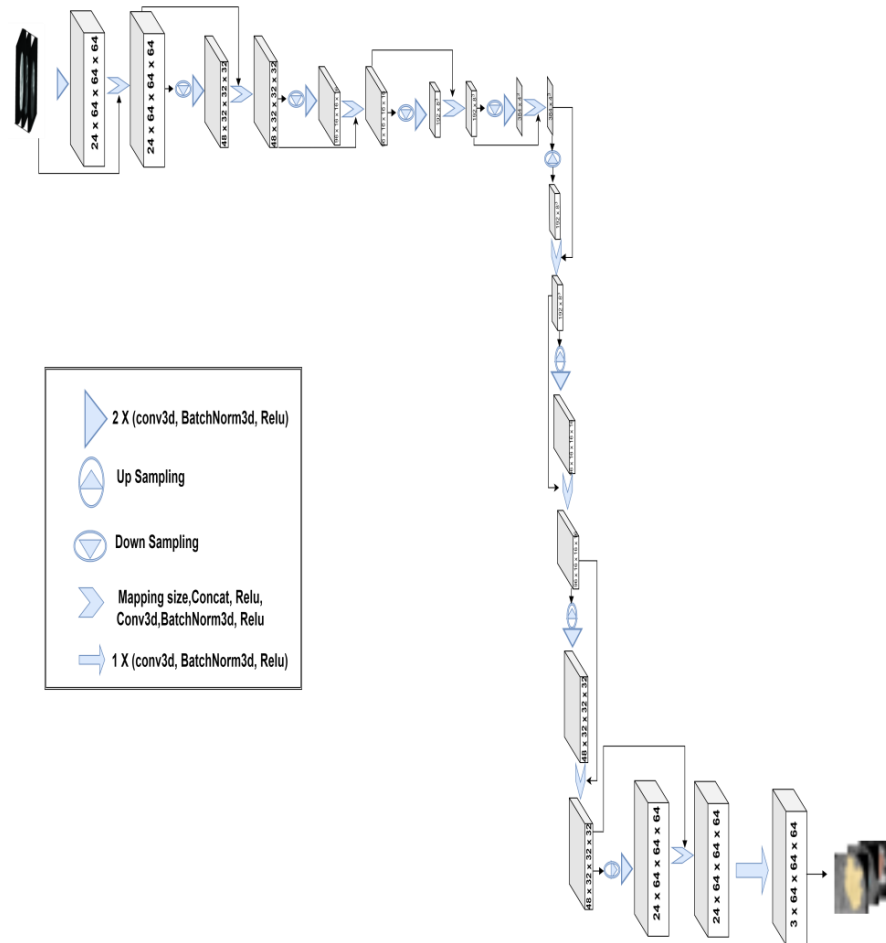


Figure 7. The proposed 3D-Znet architecture for 3D MRI brain tumor segmentation, composed of encoder–decoder blocks and fully connected connections (dense connections) for a sample spatial dimensions of (3,64,64,64).

3.3. Evaluation Metrics

A key part of DNN model evaluation requires reliable similarity measures to quantify the similarity (or discrepancy) between the ground truth output and the DNN-generated output (Znet segmented masks). Assessing image segmentation is non-trivial, since there is no unique and perfect evaluation framework [42,43]. However, metrics such as the dice similarity coefficient are useful for evaluating and tracking the similarity between segmented image outputs and the corresponding target tumor masks [28]. The set-theoretic *dice* coefficient is a measure comparing a pair of sets, *MS* (machine segmentation) and *GT* (ground truth), by calculating their intersection sizes divided by their union. The analytical form for the *dice* coefficient is shown in following equation:

$$dice = \frac{|MS \cap GT|}{|MS \cup GT|}$$

3.4. Model Training

We trained the model for 50 epochs using an adaptive moment estimation (ADAM) optimizer [44], images volumetric of size $128 \times 128 \times 128$ pixels, a batch size of 1 due to memory limitation, and a binary cross-entropy loss function [45]. Hardware and

software specifications include 2×16 -core Intel Xeon CPUs, $1 \times$ NVidia Titan 12 GB GPU, 128 GB RAM, 6 TB HDD storage, Ubuntu 18.04.5 LTS, Nvidia GPU driver v460.91, CUDA 11.2 + CuDNN 8.1, Torch v1.10.0, torchvision v0.11.1, Spyder v4.2.5, and other supporting python libraries, with a running time of about 45 min for each epochs. See the project GitHub repository for details, code, and complete end-to-end protocol, https://github.com/SOCR/DL_ZNet_3D_BrainSeg (accessed on 6 March 2023).

4. Experimental Results

In this section, we summarize the 3D-Znet model performance using a series of experiments aiming to identify an optimal multimodal tumor mask segmentation using the BraTS (2020) public dataset containing the stereotactic MR volumes. Initially, the data augmentation process was necessary to expand the data samples from the original sample of 369 raw volumes to a larger 1,845 sample of training volumetric data using 3D affine transformations. The resulting augmented dataset was divided into training (80%) and testing (20%) sets. Then, all stereotactic data were resized to $128 \times 128 \times 128$ tensors to reduce training complexity and fit the joint 3D-Znet model fitting on all training data within the available RAM limits. Each data sample contains the annotated mask (ground truth), which was processed to obtain three masks called ET (Enhanced Tumor), TC (Tumor Core), and WT (Whole Tumor). The results of training the previously discussed approach 3D-Znet (Figure 7) showed a mean dice correlation of 0.91 for segmenting the whole tumor, 0.85 for tumor core, and 0.86 for segmenting the most difficult enhanced tumor. This corresponds to an overall average dice segmentation coefficient of 0.87. These results provide strong evidence of the ability of the proposed 3D-Znet DCNN method to reliably segment different types of brain tumors where the DCNN-generated tumor masks are in very good agreement with human expert delineations and other state-of-the-art models (see Table 1).

Table 1. Comparison of the 3D-Znet model to previous segmentation models based on the mean dice coefficient using three types of outputs; tumor core (TC), enhanced tumor (ET), and whole tumor (WT).

Model Information	Dice Coefficient			Avg.	Dataset	Ref.
	WT	TC	ET			
Robust Deep Learning and Ranger for brain tumor segmentation 3D U-Net	88.9%	81.4%	84.1%	85.0%	Brats2020	[46]
Modality-Pairing learning method using 3D U-Net	89.1%	81.6%	84.2%	84.9%	BraTS2020	[47]
Hybrid High-resolution and Non-local Feature Network	91.3%	78.8%	85.5%	85.2%	BraTS2020	[48]
MobileNetV2 with residual blocks as encoder and upsampling part of U-Net as decoder	91.4%	83.3%	88.1%	87.6%	BraTS2020	[28]
Asymmetric U-Net embedding network for 3D brain tumor segmentation	80.7%	69.7%	75.2%	75.2%	BraTS2020	[49]
Deep Convolutional Neural Networks with spherical space transformed input data	86.9%	79.0%	80.7%	82.2%	BraTS2020	[50]
Context Aware 3D UNet for Brain Tumor Segmentation	89.1%	79.1%	84.7%	84.3%	BraTS2020	[51]
Cascade of three Deep Layer Aggregation neural networks	88.6%	79.0%	83.0%	83.5%	BraTS2020	[52]
Multi-encoder Network for brain tumor segmentation	70.2%	73.9%	88.3%	77.5%	BraTS2020	[53]
3D-Znet encoder-decoder Network for 3D brain tumor segmentation	90.6%	84.5%	85.9%	87.0%	BraTS2020	Current

Some visualization examples of the Znet output on testing data are shown in Figure 8. To visually inspect the raw brain images, the corresponding anatomical expert-drawn

manually delineated tumor masks, and contrast these against the 3D-Znet predictions, the figure shows axial 2D cross-sections of the 3D volumes. The left panels show the ground truth tumor masks superimposed on the observed MRI volumetric data (image sections) and displayed in green color. The middle panels depict the 3D-Znet prediction masks overlayed on the MRI sections, and the right panels illustrate the overlap between ground truth and DNN-derived tumor masks. There is good agreement between the actual (human expert) and the 3D-Znet (CNN) predicted tumor boundaries. The latter appear a little more regular, smoother, and less complex compared to the native masks, which tend to have highly curved boundaries. Subsequent studies may need to investigate the issue of DCNN regularization of predicted imaging results, understand the underlying causes, and potentially correct for or adjust the DCNN parameters to allow for more irregular boundary shapes.

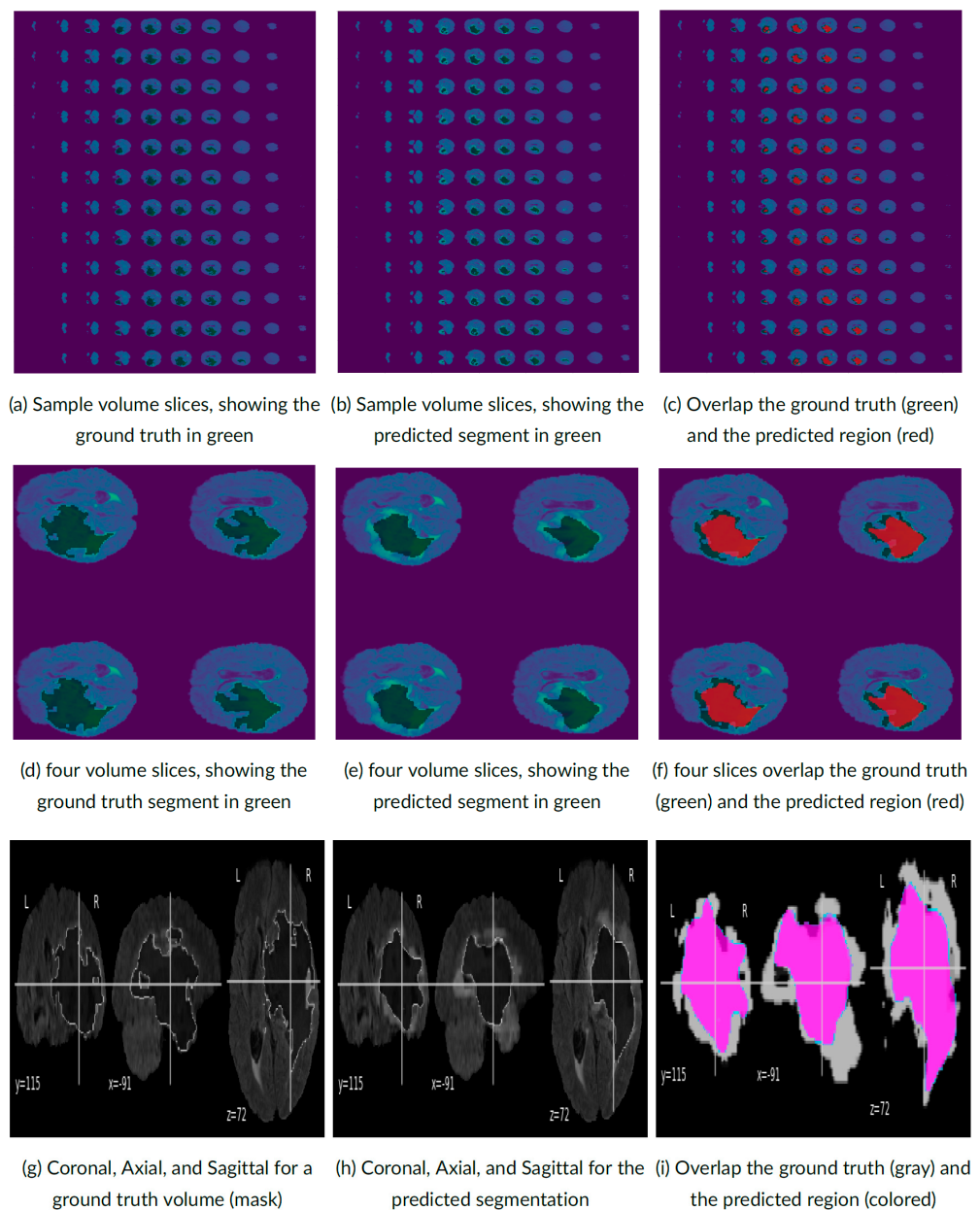


Figure 8. An example of applying the 3D-Znet model to segment the expected brain tumor mask using one random validation-set test-case.

5. Conclusions

In this manuscript, we proposed an efficient deep convolutional neural network (CNN) approach, 3D-Znet, to learn the stereotactic neuroimaging affinities and segment prospective brain tumors using publicly available datasets, such as BraTS (2020). The same 3D-Znet model can be retrained, or refined using transfer learning, on other supervised learning problems. The proposed approach was originally inspired by the variational encoder–decoder framework and the skip connection concept to enable the model to reuse features on multiple levels. The 3D-Znet model includes four encoders and four decoders along with input and output blocks. The model was evaluated using the BraTS (2020) datasets. Assessment of the proposed approach indicates high overall mean dice coefficient scores for whole tumor (0.91), tumor core (0.85), and enhanced tumor (0.86) masks. The augmentation of the original data sample and appropriate data preprocessing provided a performance boost and enhanced the Znet model predictions. In addition, we found that data augmentation plays an important role in avoiding model overfitting. On the other hand, data augmentation requires significantly more computational resources, longer training time, and significant computational infrastructure during the learning process. These upfront costs of training the Znet model on augmented data do not present a computational burden during the subsequent Znet tumor prediction, model validation, translations, and clinical assessment. We found that predicting the enhanced and tumor concentrated masks represented the most challenging tumor segmentation problem in the Brats (2020) archive. This may be explained by the low number of pixels in these types of tumors and potentially highly-subjective diagnoses by trained neuro-radiologists for complex tumor types. Prospective work to expand, improve, and generalize the proposed Znet model may involve alternative strategies to overcome limited data samples, swapping the 3D convolutional layers in the DCNN by 3D wavelet or 3D fractal encoding-decoding transformations, and utilizing different learning techniques. The problem of generating efficient, reliable, and realistic algorithms for segmenting high-dimensional and multimodal neuroimaging data with supervised ground truth labels is difficult. Solutions to this problem may have direct implications to advancing clinical care as well as provide novel mechanisms to synthetically generate unlimited (simulated) realistic neuroimaging data that can be used to train the next-generation AI/ML algorithms that are more sensitive, expeditious, and pragmatic. In addition, transfer learning approaches based on the proposed 3D-Znet may also reduce the training time and provide more accurate predictions. Deep learning models are complex models that often have a large number of parameters. This complexity can lead to inherent uncertainty, which refers to the fact that the model may not fully capture the underlying data distribution. Moreover, there are many hyperparameters that can be tuned in deep learning models, such as learning rate, batch size, and dropout rate. The choice of hyperparameters can affect the model's performance and uncertainty. Addressing these uncertainties is an ongoing area of research in deep learning.

All software, pretrained 3D-Znet models, and end-to-end electronic python notebook used in this study are available in the project GitHub repository (https://github.com/SOCR/DL_ZNet_3D_BrainSeg, accessed on 6 March 2023).

Author Contributions: M.A.O. contributed to the conception, design, and data analysis of the paper. All authors contributed to data interpretation and drafting/editing the manuscript. M.A.O., H.A.R., I.M.A. and I.D.D. were involved in revising the manuscript, providing critical comments. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported in part by NSF grants 1916425, 1734853, 1636840, NIH grants UL1 TR002240, R01 CA233487, R01 MH121079, R01 MH126137, and T32 GM141746. The funding agencies played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Many colleagues at the University of Michigan Statistics Online Computational Resource (SOCR) contributed ideas, infrastructure, and support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All software, pretrained 3D-Znet models, and end-to-end electronic python note-book used in this study are available in the project GitHub repository (https://github.com/SOCR/DL_ZNet_3D_BrainSeg, accessed on 6 March 2023). The raw or compressed NIFTI files can also be easily displayed using the SOCR BrainViewer webapp (<https://socr.umich.edu/HTML5/BrainViewer/>, accessed on 6 March 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Shinde, P.P.; Shah, S. A review of machine learning and deep learning applications. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA), Pune, India, 16–18 August 2018; IEEE: New York, NY, USA, 2018; pp. 1–6.
- Dinov, I.D. *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*; Computer Science; Springer: Berlin/Heidelberg, Germany, 2018.
- Bolhasani, H.; Mohseni, M.; Rahmani, A.M. Deep learning applications for IoT in health care: A systematic review. *Inform. Med. Unlocked* **2021**, *23*, 100550. [[CrossRef](#)]
- Ottom, M.A. Convolutional neural network for diagnosing skin cancer. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 333–338. [[CrossRef](#)]
- Chen, B.; Zhang, L.; Chen, H.; Liang, K.; Chen, X. A novel extended kalman filter with support vector machine based method for the automatic diagnosis and segmentation of brain tumors. *Comput. Methods Programs Biomed.* **2020**, *200*, 105797. [[CrossRef](#)] [[PubMed](#)]
- Preethi, S.; Aishwarya, P. An efficient wavelet-based image fusion for brain tumor detection and segmentation over PET and MRI image. *Multimed. Tools Appl.* **2021**, *80*, 14789–14806. [[CrossRef](#)]
- Hu, A.; Razmjoo, N. Brain tumor diagnosis based on metaheuristics and deep learning. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 657–669. [[CrossRef](#)]
- Dinov, I.D.; Velez, M.V. *Data Science: Time Complexity, Inferential Uncertainty, and Space-time Analytics*; Walter de Gruyter GmbH & Co KG: Berlin/Heidelberg, Germany, 2021.
- Ahmad, M.A.B. *Mining Health Data for Breast Cancer Diagnosis Using Machine Learning*; University of Canberra: Canberra, Australia, 2013.
- Olender, G.; Hurschler, C.; Fleischer, B.; Friese, K.I.; Sukau, A.; Gutberlet, M.; Becher, C. Validation of an anatomical coordinate system for clinical evaluation of the knee joint in upright and closed MRI. *Ann. Biomed. Eng.* **2014**, *42*, 1133–1142. [[CrossRef](#)]
- Rohlfing, T.; Zahr, N.M.; Sullivan, E.V.; Pfefferbaum, A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum. Brain Mapp.* **2009**, *31*, 798–819. [[CrossRef](#)] [[PubMed](#)]
- Sharkey, J.M.; Quarrington, R.D.; Magarey, C.C.; Jones, C.F. Center of mass and anatomical coordinate system definition for sheep head kinematics, with application to ovine models of traumatic brain injury. *J. Neurosci. Res.* **2022**, *100*, 1413–1421. [[CrossRef](#)]
- Ratti, C.; Wang, Y.; Piper, B.; Ishii, H.; Biderman, A. PHOXEL-SPACE: An interface for exploring volumetric data with physical voxels. In Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, Cambridge, MA, USA, 1–4 August 2004; pp. 289–296.
- Cassinelli, A.; Ishikawa, M. Volume Slicing Display. In Proceedings of the SA09: SIGGRAPH ASIA 2009, Yokohama, Japan, 16–19 December 2009; Association for Computing Machinery: New York, NY, USA.
- SOCR University of Michigan, 3D Brain Viewer Using XTK—Boston Children Hospital. Available online: <https://socr.umich.edu/HTML5/BrainViewer/> (accessed on 20 January 2022).
- Multiple Sclerosis Org, Basic Plane Mathematics of MRI. Available online: https://my-ms.org/mri_planes.htm (accessed on 12 February 2022).
- Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [[CrossRef](#)] [[PubMed](#)]
- Du, X.; Yin, S.; Tang, R.; Zhang, Y.; Li, S. Cardiac-DeepIED: Automatic pixel-level deep segmentation for cardiac bi-ventricle using improved end-to-end encoder-decoder network. *IEEE J. Transl. Eng. Health Med.* **2019**, *7*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Chollet, F. *Deep Learning with Python*; Simon and Schuster: New York, NY, USA, 2021.
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
- Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; IEEE: New York, NY, USA, 2016; pp. 565–571.
- Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [[CrossRef](#)]
- Despotović, I.; Goossens, B.; Philips, W. MRI segmentation of the human brain: Challenges, methods, and applications. *Comput. Math. Methods Med.* **2015**, *2015*, 450341. [[CrossRef](#)]

24. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [CrossRef]
25. Karayegen, G.; Aksahin, M.F. Brain tumor prediction on MR images with semantic segmentation by using deep learning network and 3D imaging of tumor region. *Biomed. Signal Process. Control* **2021**, *66*, 102458. [CrossRef]
26. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef]
27. Alqazzaz, S.; Sun, X.; Yang, X.; Nokes, L. Automated brain tumor segmentation on multi-modal MR image using SegNet. *Comput. Vis. Media* **2019**, *5*, 209–219. [CrossRef]
28. Saeed, M.U.; Ali, G.; Bin, W.; Almotiri, S.H.; AlGhamdi, M.A.; Nagra, A.A.; Masood, K.; Amin, R.U. RMU-Net: A Novel Residual Mobile U-Net Model for Brain Tumor Segmentation from MR Images. *Electronics* **2021**, *10*, 1962. [CrossRef]
29. Hua, R.; Huo, Q.; Gao, Y.; Sui, H.; Zhang, B.; Sun, Y.; Mo, Z.; Shi, F. Segmenting Brain Tumor Using Cascaded V-Nets in Multimodal MR Images. *Front. Comput. Neurosci.* **2020**, *14*, 9. [CrossRef] [PubMed]
30. Feng, X.; Tustison, N.J.; Patel, S.H.; Meyer, C.H. Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features. *Front. Comput. Neurosci.* **2020**, *14*, 25. [CrossRef] [PubMed]
31. Jiang, Y.; Zhang, Y.; Lin, X.; Dong, J.; Cheng, T.; Liang, J. SwinBTS: A Method for 3D Multimodal Brain Tumor Segmentation Using Swin Transformer. *Brain Sci.* **2022**, *12*, 797. [CrossRef] [PubMed]
32. Ajagbe, S.A.; Amuda, K.A.; Oladipupo, M.A.; Oluwaseyi, F.A.; Okesola, K.I. Multi-classification of Alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (CNN) approaches. *Int. J. Adv. Comput. Res.* **2021**, *11*, 51. [CrossRef]
33. Awotunde, J.B.; Ajagbe, S.A.; Oladipupo, M.A.; Awokola, J.A.; Afolabi, O.S.; Mathew, T.O.; Oguns, Y.J. An Improved Machine Learnings Diagnosis Technique for COVID-19 Pandemic Using Chest X-ray Images. In *International Conference on Applied Informatics*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 319–330. [CrossRef]
34. Atia, N.; Benzaoui, A.; Jacques, S.; Hamiane, M.; El Kourd, K.; Bouakaz, A.; Ouahabi, A. Particle swarm optimization and two-way fixed-effects analysis of variance for efficient brain tumor segmentation. *Cancers* **2022**, *14*, 4399. [CrossRef]
35. CBICA University of Pennsylvania, Multimodal Brain Tumor Segmentation Challenge 2020: Data. Available online: <https://www.med.upenn.edu/cbica/brats2020/data.html> (accessed on 15 December 2021).
36. Whitcher, B.; Schmid, V.J.; Thornton, A. Working with the DICOM and NIFTI Data Standards in R. *J. Stat. Softw.* **2011**, *44*, 1–29. [CrossRef]
37. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 170117. [CrossRef]
38. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv* **2018**, arXiv:181102629.
39. Huntenburg, J.; Abraham, A.; Loula, J.; Liem, F.; Dadi, K.; Varoquaux, G. Loading and plotting of cortical surface representations in Nilearn. *Res. Ideas Outcomes* **2017**, *3*, e12342. [CrossRef]
40. Ottom, M.A.; Rahman, H.A.; Dinov, I.D. Znet: Deep Learning Approach for 2D MRI Brain Tumor Segmentation. *IEEE J. Transl. Eng. Health Med.* **2022**, *10*, 1–8. [CrossRef]
41. Chen, L.; Bentley, P.; Mori, K.; Misawa, K.; Fujiwara, M.; Rueckert, D. DRINet for medical image segmentation. *IEEE Trans. Med. Imaging* **2018**, *37*, 2453–2462. [CrossRef]
42. Xiao, H.; Feng, J.; Wei, Y.; Zhang, M.; Yan, S. Deep Salient Object Detection with Dense Connections and Distraction Diagnosis. *IEEE Trans. Multimed.* **2018**, *20*, 3239–3251. [CrossRef]
43. Yeghiazaryan, V.; Voiculescu, I. *An Overview of Current Evaluation Methods Used in Medical Image Segmentation*; Department of Computer Science, University of Oxford: Oxford, UK, 2015.
44. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
45. Torch Contributors, Binary Cross Entropy. Available online: <https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html> (accessed on 16 December 2021).
46. Fidon, L.; Ourselin, S.; Vercauteren, T. Generalized Wasserstein dice score, distributionally robust deep learning, and ranger for brain tumor segmentation: BraTS 2020 challenge. In *Proceedings of the International MICCAI Brainlesion Workshop*, Lima, Peru, 4 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 200–214.
47. Wang, Y.; Zhang, Y.; Hou, F.; Liu, Y.; Tian, J.; Zhong, C.; Zhang, Y.; He, Z. Modality-pairing learning for brain tumor segmentation. In *Proceedings of the International MICCAI Brainlesion Workshop*, Lima, Peru, 4 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 230–240.
48. Jia, H.; Cai, W.; Huang, H.; Xia, Y. H² NF-Net for Brain Tumor Segmentation Using Multimodal MR Imaging: 2nd Place Solution to BraTS Challenge 2020 Segmentation Task. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 58–68.

49. Messaoudi, H.; Belaid, A.; Allaoui, M.L.; Zetout, A.; Allili, M.S.; Tliba, S.; Salem, D.B.; Conze, P.H. Efficient embedding network for 3D brain tumor segmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Lima, Peru, 4 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 252–262.
50. Russo, C.; Liu, S.; Di Ieva, A. Impact of Spherical Coordinates Transformation Pre-processing in Deep Convolution Neural Networks for Brain Tumor Segmentation and Survival Prediction. In Proceedings of the International MICCAI Brainlesion Workshop; Springer: Berlin/Heidelberg, Germany, 2021; pp. 295–306. [\[CrossRef\]](#)
51. Ahmad, P.; Qamar, S.; Shen, L.; Saeed, A. Context aware 3D UNet for brain tumor segmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Lima, Peru, 4 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 207–218.
52. Silva, C.A.; Pinto, A.; Pereira, S.; Lopes, A. Multi-stage Deep Layer Aggregation for Brain Tumor Segmentation. In Proceedings of the International MICCAI Brainlesion Workshop, Lima, Peru, 4 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 179–188. [\[CrossRef\]](#)
53. Zhang, W.; Yang, G.; Huang, H.; Yang, W.; Xu, X.; Liu, Y.; Lai, X. ME-Net: Multi-encoder net framework for brain tumor segmentation. *Int. J. Imaging Syst. Technol.* **2021**, *31*, 1834–1848. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.