



UNIVERSIDAD NACIONAL DE COLOMBIA

Method for the segmentation of brain magnetic resonance images using a neural network architecture based on attention models

Camilo Andres Laiton Bonadiez

Universidad Nacional de Colombia
Minas Faculty, Department of Science, Computation and Decision
Medellín, Colombia
2022

Method for the segmentation of brain magnetic resonance images using a neural network architecture based on attention models

Camilo Andres Laiton Bonadiez

Thesis submitted as partial requirement for the degree of:
Masters in Engineering - Systems Engineering

Mentors:

Ph.D. German Sánchez Torres

Ph.D. John Willian Branch

Investigation pathway:

Artificial Vision and Machine Learning

Research Group:

Grupo de Investigación y Desarrollo en Inteligencia Artificial (GIDIA)

National University of Colombia

Minas Faculty, Department of Sciences, Computation and Decision

Medellín, Colombia

2022

To God for allowing me to study, to my family for their love and patience, to my mentors for their guidance and unconditional support, to my love, Maria, and my friend, Cristian, for being with me in the hardest moments.

Thanks, I love you all.

Resumen

Título en español: Método para la segmentación de imágenes de resonancia magnética cerebrales usando una arquitectura de red neuronal basada en modelos de atención.

En los últimos años, el uso de modelos basados en aprendizaje profundo para el desarrollo de sistemas de salud avanzados ha ido en aumento debido a los excelentes resultados que pueden alcanzar. Sin embargo, la mayoría de los modelos de aprendizaje profundo propuestos utilizan, en gran medida, operaciones convolucionales y de pooling, lo que provoca una pérdida de datos valiosos centrándose principalmente en la información local. En esta tesis, proponemos un enfoque basado en el aprendizaje profundo que utiliza características globales y locales que son importantes en el proceso de segmentación de imágenes médicas. Para entrenar la arquitectura, utilizamos bloques tridimensionales (3D) extraídos de la resolución completa de la imagen de resonancia magnética. Estas se enviaron a través de un conjunto de capas sucesivas de redes neuronales convolucionales (CNN) libres de operaciones de pooling para extraer información local. Luego, enviamos los mapas de características resultantes a capas sucesivas de módulos de autoatención para obtener el contexto global, cuya salida se envió más tarde a la canalización del decodificador compuesta principalmente por capas de upsampling. El modelo fue entrenado usando el conjunto de datos Mindboggle-101. Los resultados experimentales mostraron que los módulos de autoatención permiten la segmentación con un Mean Dice Score $0,90 \pm 0,036$ la cual es mayor en comparación con otros enfoques basados en UNet. El tiempo medio de segmentación fue de aproximadamente 0,032 s por estructura cerebral. El modelo propuesto permite abordar adecuadamente la tarea de segmentación de estructuras cerebrales. Así mismo, permite aprovechar el contexto global que incorporan los módulos de autoatención logrando una segmentación más precisa y rápida. En este trabajo segmentamos 37 estructuras cerebrales y, según nuestro conocimiento, es el mayor número de estructuras bajo un enfoque 3D utilizando mecanismos de atención.

Palabras clave: segmentación de imágenes médicas, aprendizaje profundo, transformers, redes neuronales convolucionales, estructuras cerebrales.

Abstract

Título en inglés: Method for the segmentation of brain magnetic resonance images using a neural network architecture based on attention models.

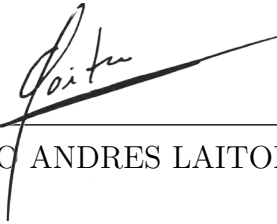
In recent years, the use of deep learning-based models for developing advanced healthcare systems has been growing due to the results they can achieve. However, the majority of the proposed deep learning-models largely use convolutional and pooling operations, causing a loss in valuable data and focusing on local information. In this thesis, we propose a deep learning-based approach that uses global and local features which are of importance in the medical image segmentation process. In order to train the architecture, we used extracted three-dimensional (3D) blocks from the full magnetic resonance image resolution, which were sent through a set of successive convolutional neural network (CNN) layers free of pooling operations to extract local information. Later, we sent the resulting feature maps to successive layers of self-attention modules to obtain the global context, whose output was later dispatched to the decoder pipeline composed mostly of upsampling layers. The model was trained using the Mindboggle-101 dataset. The experimental results showed that the self-attention modules allow segmentation with a higher Mean Dice Score of 0.90 ± 0.036 compared with other UNet-based approaches. The average segmentation time was approximately 0.032 s per brain structure. The proposed model allows tackling the brain structure segmentation task properly. Exploiting the global context that the self-attention modules incorporate allows for more precise and faster segmentation. We segmented 37 brain structures and, to the best of our knowledge, it is the largest number of structures under a 3D approach using attention mechanisms.

Keywords: medical image segmentation, deep learning, transformers, convolutional neural networks, brain structures.

Declaration

I allow myself to affirm that I have carried out this thesis in a autonomously and with the sole help of the permitted means and not different from those mentioned in the thesis itself. All the passages that have been taken verbatim or figuratively from published and unpublished texts, I have acknowledged them in this work. No part of this work has been used in any other type of thesis.

Medellín, 31/07/2022

A handwritten signature in black ink, appearing to read 'Camilo', is written over a horizontal line. The signature is stylized and cursive.

CAMILO ANDRES LAITON BONADIEZ

Content

Figures	ix
Tables	xi
1 Introduction	1
1.1 Research problem	3
1.2 Problem statement	3
1.3 Objectives	4
1.3.1 General objective	4
1.3.2 Specific objectives	4
1.4 Contributions	4
1.5 Document structure	5
2 Background	6
3 Method for the segmentation of anatomical brain structures	11
3.1 Dataset	11
3.2 Data preprocessing	14
3.2.1 Skull Stripping preprocessing step	17
3.3 Deep neural network for brain segmentation	19
3.4 Loss Functions and Class Weights	21
3.4.1 Skull stripping preprocessing step	21
3.4.2 Segmentation of brain structures	22
3.5 Data augmentation techniques	23
3.6 Metrics and training parameters	23
4 Results	26
4.1 Skull stripping preprocessing step segmentation results	26
4.2 Segmentation of brain structures	29
4.2.1 Architecture design determination	33
4.2.2 Patch resolution size determination	38
4.2.3 Comparison with other methods	38

Content	VIII
5 Discussion and future work	43
References	46

Figures

3-1	MRI volume divided into nonoverlapping subvolumes.	15
3-2	Skull stripping algorithm classified in groups. Image taken from [74] . . .	17
3-3	Skull stripping dataset process example taken from [81]	18
3-4	Proposed deep neural network architecture for Skull Stripping	19
3-5	Neural network architecture for 3D brain MRI segmentation.	20
3-6	Skip connections design based on convolutional layers.	20
3-7	Example of elastic deformation in a MRI from the Mindboggle-101 data- set. (a) Original MRI and mask; (b) Elastically deformed images. Source: Author.	24
4-1	Segmentation results of the proposed architecture for skull stripping in the axial, sagittal and coronal planes. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide. Source: Author.	27
4-2	Number of voxels for each of the 37 selected structures from Mindboggle- 101 dataset.	29
4-3	Segmentation results of the proposed architecture in the axial, sagittal and coronal planes where red, green, blue, purple, and yellow colors repre- sent cerebral white matter, cerebellum white matter, cerebellum cortex, thalamus, and putamen structures, respectively. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide. Source: Author.	30
4-4	Segmentation results of the proposed architecture in the axial, sagittal and coronal planes for all 37 brain structures. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide. Source: Author.	31
4-5	Segmentation results of the proposed architecture in 3D. (a) Segmenta- tion of the 37 brain structures; (b) segmentation of the cerebellum cortex (orange), putamen (magenta), and hippocampus structures(yellow); (c) segmentation of the brain stem (gray), insula (yellow), and superior fron- tal structures (green). Source: Author.	33
4-6	Architecture design with Transformer Layers at the beginning of the en- coder path. Source: Author.	34

4-7	Segmentation results of the deep architecture with transformers at the beginning of the encoder path. (a) Ground truth mask; (b) Segmentation result. Source: Author.	34
4-8	Segmentation results of the two-branches network proposal. (a) Ground truth mask; (b) Segmentation result. Source: Author.	36
4-9	Dice score per epoch in the training and validation sets for the proposed architecture. Source: Author.	37
4-10	IoU score per epoch in the training and validation sets for the proposed architecture. Source: Author.	38
4-11	Combined loss function per epoch in the training and validation sets for the proposed architecture. Source: Author.	39
4-12	Patch size influence comparison on structure details segmentation in the axial, sagittal and coronal planes where red, green, and blue colors represent cerebral white matter, cerebellum white matter and cerebellum cortex structures, respectively. (a) Ground truth mask; (b) segmentation with patch resolution size of $16 \times 16 \times 16$; (c) segmentation with patch resolution size of $8 \times 8 \times 8$. Source: Author.	42

Tables

3-1	Label remapping strategy for brain structures having ID 0 set for background.	16
4-1	Segmentation time per brain MRI in the NFBS created validation dataset.	28
4-2	Segmentation results per class in a testing MRI.	28
4-3	Comparison between methods by the Dice Score and segmentation times using the NFBS dataset.	29
4-4	Segmentation results per brain structure in a testing MRI.	32
4-5	Comparison between methods by the Dice Score and p-value for the Wilcoxon signed-rank test comparing proposed-UNet, proposed-DenseUNet samples pairs using the Mindboggle-101 dataset.	39
4-6	Segmentation time per brain structure for a single MRI scan.	40
4-7	Segmentation time per brain MRI in the validation dataset.	41

1 Introduction

The scientific community has developed tools that allow for obtaining brain information so doctors can study the human brain. Among these tools we find brain imaging, which includes methods such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound (US), among others. However, not all methods produce quality images when applied to the brain because high contrast images are required to study the human brain. Therefore, highly sensitive methods such as magnetic resonance imaging or positron emission tomography must be used [1].

Although PET scans are capable of obtaining good quality images, they are not the preferred choice for specialists as they have several disadvantages. PET scans cannot reveal structural information at the microscopic and macroscopic levels in the white and gray matter of the brain; cannot detect changes in brain activation, and pose health risks due to the required radiation [2]. For these reasons, the MRI method has been widely used in the brain for medical studies and scientific research [3, 4]. There are several types of MRI sequences that are capable of improving contrast and brightness in certain types of tissues. T1-weighted, T2-weighted, Fluid Attenuated Inversion Recovery (Flair), and Diffusion Weighted Imaging (DWI) are among the most common MRI sequences [5].

Traditionally, interpretations of medical images have been made by human experts. However, the existence of variations in the criteria among various human specialists is a limitation in relation to the generation of an efficient and precise diagnosis [6]. This weakness has been addressed through the development of computer aided diagnostic systems using computer vision for analysis of medical Images [7]. In this field, traditional algorithms were initially applied for the segmentation of anatomical brain structures, such as thresholding techniques [8], growth of regions [9], machine learning algorithms for classification [10], or grouping [11], among others. Based on the growth in computational capacity and the amount of data available, it is possible to use more robust and complex modern algorithms based on artificial intelligence capable of achieving better results in medical segmentation tasks [12].

In fact, a growing number of researchers are using imaging and machine vision tech-

nology based on artificial intelligence via usage of deep convolutional neural networks. Convolutional networks have shown more accurate results in most application domains, including the medical area [13, 14, 15]. The strength of convolutional networks is that they automatically identify the relevant features without any explicit human supervision [16].

Furthermore, compared to its predecessor, the fully connected neural networks, convolutional networks significantly reduce the number of trainable parameters of the network, facilitating computation and making it possible to build large-scale architectures. Also, they efficiently join the output of the network with the extracted features by jointly training the classification layers with the feature extraction layers [17].

However, despite their strengths, they still have weaknesses. Convolutional networks incorporate layers that make use of pooling operations to reduce the size of the feature map, thus reducing the computation required in subsequent layers. The implications of this operation is that features of the images introduced in the training are lost, requiring a greater amount of data to reach convergence [18]. Similarly, this type of network is not capable of recognizing the pose, texture and orientation of the images, which in brain segmentation is important since in order to have a good segmentation performance we need as much information as possible [19].

Consequently, other types of architectures have been proposed to address the weaknesses of the use of convolutional networks in deep learning architectures. Examples of this are capsule networks [19] or Transformers [20].

Transformers are a type of self-attention-based network that were primarily designed to perform tasks in the field of natural language processing. However, these have gained popularity in the field of images, achieving results comparable to those of architectures based on convolutional networks. Some advantages of using Transformers in the field of images are their ability to use effectively the available computation during training, speed and global context recognition [21].

In the field of medical image analysis, although this type of architecture is being applied in the segmentation of medical images to detect tumors or segment organs in the abdominal area [22], its use in the segmentation of brain anatomical structures has not yet been extensively explored. Due to this, this study was aimed to design a neural network architecture based on attention models directed to the problem of segmentation of anatomical brain structures.

1.1. Research problem

The segmentation of brain magnetic resonance images is of vital importance for the study of the human brain. In fact, the scientific community uses semi-automatic tools to support experts in the field of neuroscience in manual brain segmentation tasks. Tools such as Freesurfer [23], BrainSuite [24] or FSL [25] are some examples of this type of tool. Even so, existing software still requires human intervention for parameterization, is computationally expensive relative to the time spent by a human expert, and expert staff must sometimes make corrections to the resulting segmentation due to system errors.

Due to these limitations, multiple organizations have chosen to use fully automatic software based mainly on artificial intelligence, achieving better results than traditional algorithms in the area of brain segmentation [12]. However, most of the proposed architectures are based on the use of convolutional neural networks as the basic building block. Currently, there is a recent focus on the use of neural attention that has not been widely addressed in this domain. For this reason, it is of interest to explore the benefits of the application of attention models in the problem of automatic segmentation of anatomical brain structures.

1.2. Problem statement

Given an 3D image $X \in \mathbb{R}^{H \times W \times D \times C}$ where H represents height, W width, D depth, and C the number of channels. Each of the images X^j , being j the total number of MRIs of the dataset, have a pixel-wise label map $Y \in \mathbb{Z}^{H \times W \times D}$ where the value of a voxel (h, w, d) in Y is a positive integer that determines the class the voxel in X belongs to in the same coordinates. In this study, the defined classes are multiple brain structures and a background class defined as $Y^j(h, w, d) = \{E_1, E_2, E_3, \dots, E_q\}$.

Be $\Phi(R, \theta, W^k) = \Upsilon$ a neural network that receives an input $\mathbb{R}^{H \times W \times D \times C}$ to which it applies linear and nonlinear transformations defined by hyperparameters θ through k layers with W weights. This neural network should generate as result the output $\Upsilon \in \mathbb{Z}^{H \times W \times D}$ which is a 3D image of predicted classes for each voxel in Z .

Therefore, the problem of segmenting anatomical brain structures given an MRI X is to find the neural network $\Phi(\mathbb{R}, \theta, W^k)$ with the set of hyperparameters θ , the number of layers k , the weights W^i from pairs of images (x^j, y^j) by numerical optimization of the loss function $\ell(Y^j, \Phi)$ that allows to predict the class E_q of each voxel given a new MRI X' .

Although some aspects of the structure of the network and the optimization model have been simplified, this thesis also includes, as a problem, defining the architecture that allows segmenting a set of q brain structures within an MRI.

1.3. Objectives

1.3.1. General objective

To design a method for automatic segmentation of anatomical brain structures from magnetic resonance images using deep learning techniques based on attention models.

1.3.2. Specific objectives

1. Determine the set of brain magnetic resonance images that will be part of the dataset and select the set of anatomical brain structures to segment.
2. Define the preprocessing scheme of the defined magnetic resonance images to model the input of the neural network architecture.
3. Design the neural network architecture that allows the segmentation of the defined anatomical brain structures.
4. Evaluate and compare the implemented neural network model with those existing in the state of the art.

1.4. Contributions

A method capable of segmenting 37 brain structures from magnetic resonance images is designed. It is, to the best of our knowledge, the largest number of structures under a 3D approach using attention mechanisms. From this proposed method it was obtained:

1. A trained deep neural network architecture, based on U-Net [26], capable of segmenting 37 brain structures using attention mechanisms.

2. A trained neural network architecture capable of segmenting brain structures in less time than those established in the state of the art thanks to the use of attention mechanisms.
3. The article *Deep 3D Neural Network for Brain Structures Segmentation Using Self-Attention Modules in MRI Images* which was published in the Q1 level journal *Sensors* that can be found [here](#).

1.5. Document structure

This document is organized as follows: we present an introduction of the addressed problem in Chapter 1, Chapter 2 shows the background of the research problem, Chapter 3 describes the proposed method for the segmentation of anatomical brain structures. Chapter 4 includes the obtained results of the proposed method. Finally, Chapter 5 constitutes the main conclusions as well as future research.

2 Background

In general, multiple methods for the segmentation of brain magnetic resonance images have been proposed. These methods can be grouped as follows: manual methods, spatial dimensionality-based methods (2D and 3D), pixel/voxel intensity-based methods, atlas-based methods, surface-based methods, and methods based on deep learning techniques [27].

The manual segmentation of magnetic resonance images is based on the use of highly trained personnel to obtain the different types of brain tissues. In order to perform segmentation, experts commonly use manual delineation tools that allow them to delineate different regions of the brain. Some examples of these tools are FreeSurfer [23], BrainSuite [24], FSL [25], ITK-SNAP [28], 3D Slicer [29], SPM [10], and Horos [-], among others.

In terms of the spatial dimensionality methods, these are subdivided into 3D and 2D approaches. Three-dimensional-based segmentation approaches seem to be the natural way to approach the problem because it allows to exploit the three-dimensional nature of MRI by considering each voxel and its relationship to neighbors at different acquisition planes (sagittal, coronal, and axial). However, the 3D approach still has limitations, mainly related to the high computational cost in computers with limited memory, the increase in the complexity of the models and the number of parameters, making the learning process slower [30, 31, 32]. Therefore, researchers usually use the 2D representation of a brain MRI to avoid memory restraints and computational limitations of the 3D representation.

Intensity-based methods attempt to find a threshold value that separates the different tissue categories. These methods include techniques of thresholding, growth of regions, classification, and grouping [8]. In [33, 34], authors presented work on pixel intensity using thresholding techniques to segment brain tumors.

Works that use region growth techniques, use the similar characteristics of pixels found together to perform the separation of a common region [35]. Region growth techniques have been applied for the segmentation of brain tumors [9], organs [36], cerebral vessels

[37], and lesions in both the brain and the breast, applying other techniques such as morphological filters [38] or with quantitative values such as the measurement of the roughness of the tumor border [39]. Within this group, we also find classification and grouping techniques that make use of labeled and unlabeled data for their operations. In fact, multiple brain segmentation tools that are widely used in the scientific community, such as FreeSurfer [23], 3D Slicer [29], and SPM [10] make use of Bayesian Classifiers. The k-means algorithm is the most used in clustering because it is simple to implement, relatively fast, and produces good results. Some examples of work for tissue or tumor segmentation are presented in [11] and in [40], where the k-means algorithm is combined with a vector support machine and a Fuzzy C-means algorithm with thresholding techniques, respectively.

Atlas-based methods are those that make use of brain anatomical information from a specific population for image segmentation. Likewise, in [41, 42, 43, 44], it is shown that this method is not limited to the use of a specific atlas; multiple atlases and techniques such as tag fusion can also be used to improve the delimitation of brain regions.

Several works have also been proposed that make use of deformable models. These techniques are part of the group of surface-based methods where the main objective is the delimitation of regions with similar characteristics through the use of elastic curves [45]. Similar to the approaches mentioned above, surface-based methods have been used for the segmentation of brain regions [46] and tumors [47, 48].

Recently, deep learning has become an area of interest in the scientific community due to the important results that have been achieved in multiple disciplines [13, 14, 15, 49, 50]. The strength of convolutional networks is that they automatically identify relevant characteristics without any explicit human supervision [16]. In addition, compared with their predecessors, fully connected neural networks, convolutional networks significantly reduce the number of trainable network parameters, facilitating computation and making it possible to build large-scale architectures. Likewise, they efficiently link the output of the network with the extracted characteristics by jointly training the classification layers and the characteristic extraction layers [17]. Specifically, in the problem of brain magnetic resonance imaging segmentation, deep learning has achieved better results than previously exposed methods [12]. Within the deep learning branch, there are multiple algorithms based on neural networks that have been developed with specific objectives, such as autoencoders [51], Boltzmann machines [52], recurrent neural networks [53], convolutional neural networks [54], and Transformers [20], among others. Convolutional neural networks are precisely the algorithms most widely used by researchers to perform image segmentation and classification tasks, given that they have achieved the best results to date.

Convolutional neural networks are a type of neural network that was created by Le-Cun but was inspired by Fukushima's work on the neocognitron for the recognition of handwritten Japanese characters [55]. In the study of brain magnetic resonance imaging using neural network architectures, it is common to see convolutional neural networks as the basis of the architectures. In fact, in [56], the authors presented a solution for brain tissue segmentation on MRI images taken in infants approximately six to eight months of age using CNNs in the deep learning architecture.

Similarly, the authors of [57] were able to use CNNs for the segmentation of subcortical brain structures using the datasets of Internet Brain Segmentation Repository (IBSR) and LPBA40 [24].

Some of the most important deep learning solutions have been proposed using a 2D representation allowing researchers to segment more structures than a 3D representation allows them to do. In fact, the segmentation of more than 25 brain structures into a 3D representation has been achieved by a few works, while using a 2D representation, deep learning works can segment more than 95 brain structures [58].

The use of this type of neural network is not limited to the segmentation of brain tissues or brain structures. These have also been used in the segmentation of brain lesions, as in [59, 60, 61], the segmentation of brain tumors [62, 63, ?] [60 (missing)], the detection of ischemic strokes [64], and even genomic prediction of tumors [65]. The most important thing to note from these works is that many use branches within their neural network architectures. In general, they use two branches, where one of them is focused on the extraction of globally related characteristics (global context), while the other is in charge of the extraction of local characteristics (local context) to achieve better segmentation.

One of the architectures most commonly used in medical image segmentation tasks is the U-Net architecture [26]. Due to the structure of its architecture, U-Net has advantages over other convolutional neural network architectures of its time. This was built having a path that encodes the characteristics of the image and then continues with its expansion, that is, an encoder-decoder structure. In addition, to avoid the vanishing gradient and explosion problem, the U-Net architecture incorporates skip connections, between the encoder and decoder layers, which improves performance in small datasets compared with other architectures at the time.

Multiple neural network architectures based on U-Net have been proposed for the field of medical image segmentation. The primary goals of these works were to improve the network by using skip connections between the layers of the coding and expansion path [66, 67] and to combine the architecture with others such as SegNet [68]. It is important to note that several of these studies were also applied to the segmentation of white

matter, gray matter, and cerebrospinal fluid from brain magnetic resonance images.

However, convolutional neural networks have serious limitations. One of them is the loss of image characteristics due to pooling operations [18]. This is because the CNNs require these operations to reduce the feature maps resulting from the convolutions and thus reduce the computation required in subsequent layers. Due to this, a large amount of data is necessary in the training process for deep learning networks to be able to generalize and achieve good results [18].

On the other hand, researchers have proposed multiple deep learning architectures based on attention mechanisms such as the Transformer's architecture [32, 33, 69]. This one was initially proposed in the field of natural language processing [70] as being in charge of transforming one sequence into another from multiple attention blocks [20]. The Transformers replaced the recurrent neural network models (RNN) used until then for the translation of texts because it solved its main weakness. This was because the performance of the recurring models fell when very long sequences were introduced due to the long-term dependency learning problem [69], and although this problem was attacked by the Long Short-Term Memory (LSTM) networks, they did not achieve as good results as the Transformers. This became possible since the latter, through self-attention mechanisms, are capable of processing the entire sequence entered, even if it is very long, optimizing processing times due to parallel processes within the network.

Thus, the scientific community has achieved that the Transformer's architecture can obtain results comparable to those established as the state of the art in computer vision methods [71]. In fact, some methods based on Transformers' architectures were proposed for the segmentation of medical images. The TransUNet [22] network, which is based on the U-Net architecture [26], consists of a set of convolutional layers to extract the most important characteristics of the image. The resulting feature maps are then the input to successive attention blocks, which then send this output to the decoder. The decoder is fabricated of convolutional and upsampling layers to achieve the output of a segmented image. It is necessary to mention that the set of convolutional layers is connected with the layers of the decoder through skip connections.

Also, another Transformer-based architecture is the Medical-Transformer network [72], which is based on the use of two branches for its operation. The important thing to highlight in this study is that it has a local and global branch, as has been proposed in various convolutional neural network architectures and the use of convolutions in the feature coding process. Specifically, the local branch has a greater number of encoder and decoder blocks than the global branch. The encoder blocks are fabricated of 1×1 convolutional layers, normalization, Rectified Linear Unit (ReLU) activations, and multiple layers of attention for its operation, while the decoder has closed axial attention

layers.

3 Method for the segmentation of anatomical brain structures

In this study, a 3D architecture of deep neural networks is proposed for the task of segmenting volumes associated with brain structures from MRI. Our proposal uses an encoder/decoder approach, strengthening the connection between them by incorporating self-attention modules and skip connections. The attention modules as well as the convolution operations allow the network to incorporate global and local features, and achieve a more detailed segmentation of the edges of structures.

However, the segmentation of multiple brain structures is a complex problem since computational resources must be used in an optimized way to be able to segment a large number of structures in a 3D representation. Furthermore, it is important to mention that the brain has a different structure that varies in size and shape, which represents a great challenge for the scientific community in terms of developing and training deep neural network architectures since there are not multiple nor large publicly available datasets.

Therefore, this method can be divided into multiple stages: dataset selection, data pre-processing, design of the deep neural network for brain segmentation, loss functions and class weights, and metrics and training parameters.

3.1. Dataset

Public availability of properly sized and labeled datasets for training deep learning models is a common constraint in the medical image field. This problem increases in the subfield of brain segmentation where a large number of correctly labeled brain structures is needed and where data based on MRIs is limited. The selection of the dataset was carried out having in mind two principal characteristics: data volume and the number of manually labeled brain structures.

Therefore, we selected the publicly available Mindboggle-101 dataset [73] which represents the largest dataset of manually labeled brain MRIs. This dataset contains 101 manually labeled human brain images from healthy patients using the Desikan-Killiany-Tourville (DKT) protocol, containing more than 100 brain structures in each volumetric segmentation file `aseg + aparc`. In this dataset, brain structures are labeled based on the brain hemisphere where they are located (e.g., left hippocampus and right hippocampus). Authors scanned five subjects for this dataset and selected others from open-source datasets. The inclusion criteria of external MRIs was:

1. Public MRIs with Non-restrictive license.
2. Healthy participants.
3. High quality MRIs.
4. Multi-modal MRI acquisition (T2- weighted, diffusion-weighted scans, etc.).

Additionally, it is important to mention that authors initially preprocessed and segmented all brain MRI volumes using freesurfer (version 5.1.0) recon-all processing pipeline prior the final segmentation with their proposed Desikan-Killiany-Tourville labeling protocol.

This [processing pipeline](#) can be divided as follows:

1. Step 1 (autorecon1):
 - [Motion correction and conform](#)
 - [Non-Uniform Intensity Normalization](#)
 - [Tairach transform computation](#)
 - [Intensity normalization](#)
 - [Skull strip](#)
2. Step 2 (autorecon2):
 - [Linear volumetric registration](#)
 - [CA intensity normalization](#)
 - [CA non-linear volumetric registration](#)

-
- CA Register Inverse
 - Remove neck
 - EM register, with skull
 - CA label (sseg: Volumetric Labeling) and statistics
 - Intensity normalization 2
 - Mask brain final surface
 - White matter segmentation
 - Fill
 - Tessellation
 - Orig surface smoothing 1
 - Inflation 1
 - QSphere
 - Automatic topology fixer
 - Orig surface smoothing 2
 - Inflation 2
3. Step 3 (autorecon3):
- Spherical inflation
 - Ipsilateral surface registration
 - Jacobian
 - Average curvature
 - Cortical parcellation
 - Surface volume

- [Parcellation statistics](#)
- [Cortical Parcellation](#)
- [Parcellation statistics](#)
- [Cortical sibbon mask](#)
- [Segmentation statistics](#)
- [White matter parcellation](#)
- [Brodmann Area Maps \(BA Maps\) and Hinds V1 Atlas](#)

3.2. Data preprocessing

The Mindboggle-101 dataset is not a dataset that can be used directly for deep learning purposes. We had to apply some preprocessing steps prior training the deep neural network architecture in order to use the data and optimize the usage of our computational resources.

The first step was directed towards the creation of an array representation of the original MRIs and segmentation masks in the dataset making it straightforward for training deep learning models. This is because brain MRIs in this dataset come in a .MGZ file format which is used to store high-resolution structural data and is not transparent for training deep learning models.

Additionally, for the selected structures, we created a label remapping strategy with IDs 1 to 37 having the ID 0 set for background (see Table 3-1). We created the segmentation ground truth files by taking the aseg+aparc files and mapping the existing IDs to our desired IDs. After this, we applied a min-max scaling to the voxel values to be in a range between 0 and 1. This intensity rescaling was performed using the histogram equalization technique over each individual MRI.

Then, we applied a filter in each MRI where empty slides were removed from the brain volumes, leaving on average 192 slides per brain plane. These preprocessed MRI volumes were divided into nonoverlapping subvolumes of size $64 \times 64 \times 64$ voxels that were saved in single files containing stacks of volumes per brain MRI (see Figure 3-1). These resulting arrays were saved in uint8 data type for the preprocessed ground truth files and in float32 for the preprocessed brain volumes. It is important to mention that these data

types were applied to the arrays with the aim of optimizing computational resources. Also, we created the stacks of volumes per MRI because our hard disk could not handle the number of reading petitions per block Tensorflow was asking for.

Finally, we divided the dataset into two sets in the ratio of 8:2. The first set was for training the neural network architecture, and the second one was for validating it. Due to the Mindboggle-101 dataset contains multiple datasets such as OASIS-TRT-20 which contains 20 MRIs, NKI-22 that contains 22 MRIs, among others, we performed the dataset division maintaining the original dataset distribution, making sure that the validation set had at least one MRI from all the datasets.

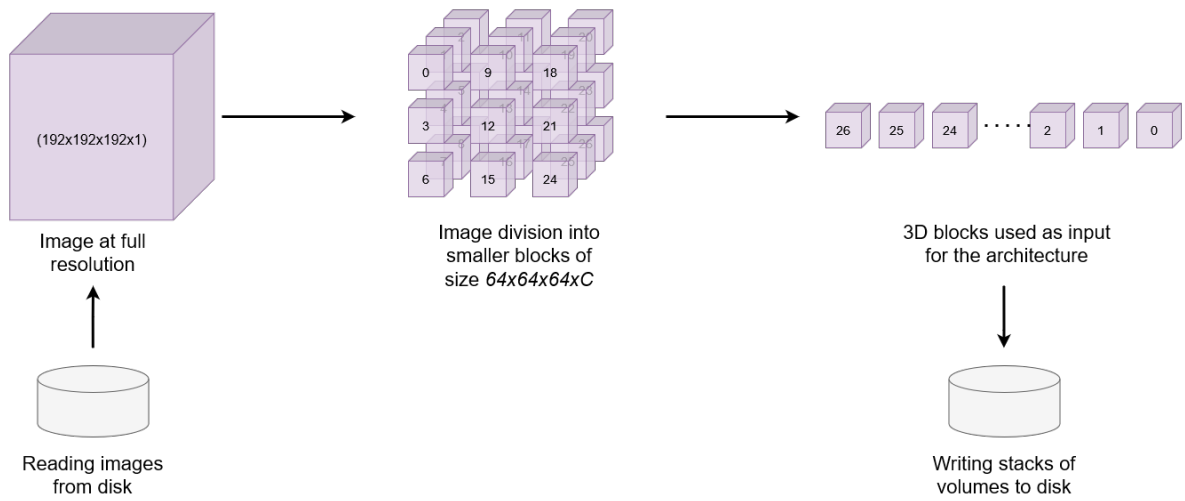


Figure 3-1: MRI volume divided into nonoverlapping subvolumes.

The computational implementations were performed with the open source library for numerical computation Tensorflow and run on a computer with a 5th generation Intel I7 4820k@3.70 GHz processor, 64 GB of RAM memory, and two Nvidia 1080TI video cards with 11 GB of GDDR 5x RAM at 405 MHz.

Table 3-1: Label remapping strategy for brain structures having ID 0 set for background.

Brain Structure	Proposed ID	FreeSurfer ID
Left cerebral white matter	1	2
Right cerebral white matter	2	41
Left cerebellum white matter	3	7
Right cerebellum white matter	4	46
Left cerebellum cortex	5	8
Right cerebellum cortex	6	47
Left lateral ventricle	7	4
Right lateral ventricle	8	43
Left thalamus	9	10
Right thalamus	10	49
Left putamen	11	12
Right putamen	12	51
3rd ventricle	13	14
4th ventricle	14	15
Brain stem	15	16
Left hippocampus	16	17
Right hippocampus	17	53
Left ventral DC	18	28
Right ventral DC	19	60
Ctx left caudal middle frontal	20	1003
Ctx right caudal middle frontal	21	2003
Ctx left cuneus	22	1005
Ctx right cuneus	23	2005
Ctx left fusiform	24	1007
Ctx right fusiform	25	2007
Ctx left inferior parietal	26	1008
Ctx right inferior parietal	27	2008
Ctx left lateral occipital	28	1011
Ctx right lateral occipital	29	2011
Ctx left post central	30	1022
Ctx right post central	31	2022
Ctx right rostral middle frontal	32	1027
Ctx left rostral middle frontal	33	2027
Ctx left superior frontal	34	1028
Ctx right superior frontal	35	2028
Ctx left insula	36	1035
Ctx right insula	37	2035

3.2.1. Skull Stripping preprocessing step

In order to properly perform brain segmentation processes an additional preprocessing step is required. Skull stripping is an important and initial preprocessing step that most MRI studies use to remove non-brain tissue and decrease the amount of propagated error algorithms could have in the process of brain tissue segmentation, volumetric measurement, longitudinal analysis, multiple sclerosis analysis, among others [74].

The inclusion of non-brain tissue in MRI images can lead to misclassifications due to the similar voxel/pixel value intensities have to the desired brain segmentation (classes). In [75], authors showed the importance of using the skull stripping preprocessing step for a more accurate and sensitive analysis of voxel-based morphometry (VBM) in brain morphology. Similarly, in [76], authors demonstrated the positive impact of skull stripping for brain gray-matter segmentation.

It is important to mention that, similar to the brain segmentation field study, there are multiple state-of-the-art brain tools that have developed their own skull stripping algorithms as a preprocessing step such as AFNI [77], ANTS [78] or Brain Surface Extraction (BSE) [79] for further MRI processing. In figure 3-2, a taxonomy on the developed skull stripping algorithms is presented based on [74] where we find manual, classical and more recent skull stripping algorithms, the last one based on deep learning by using mostly convolutional neural networks (CNNs).

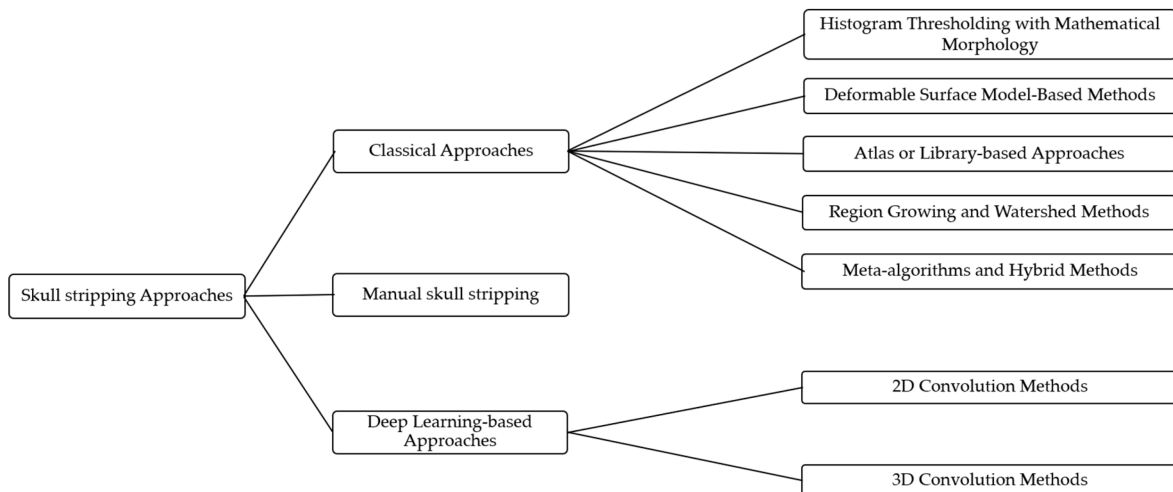


Figure 3-2: Skull stripping algorithm classified in groups. Image taken from [74]

Generally, skull stripping manual and classical methods usually are expensive in human, time and economical resources. For example, the manual delineation of a single MRI volume takes between 15 to 120 minutes requiring trained specialized personnel [80].

Therefore, multiple algorithms have been proposed to tackle this issue where the best results have been achieved in the deep learning field [74].

However, the majority of the proposed deep learning algorithms have been using CNNs as the main block in the architecture design. We aimed to explore the advantages of attention mechanisms for the skull stripping preprocessing step.

Therefore, we proposed an deep neural network architecture for skull stripping using the Neurofeedback Skull-stripped (NFBS) repository [81]. The NFBS repository is a publicly available dataset that has 125 T1-weighted MRI scans that are manually skull-stripped as can be seen in figure 3-3. They have a matrix volume of $256 \times 256 \times 192$ with a voxel size of $1 \times 1 \times 1\text{mm}^3$. Also, the age of subjects ranges from 21 to 45 years old. This dataset was created having in mind machine learning developers. Consequently, the data is ready to use for the creation of machine learning algorithms.

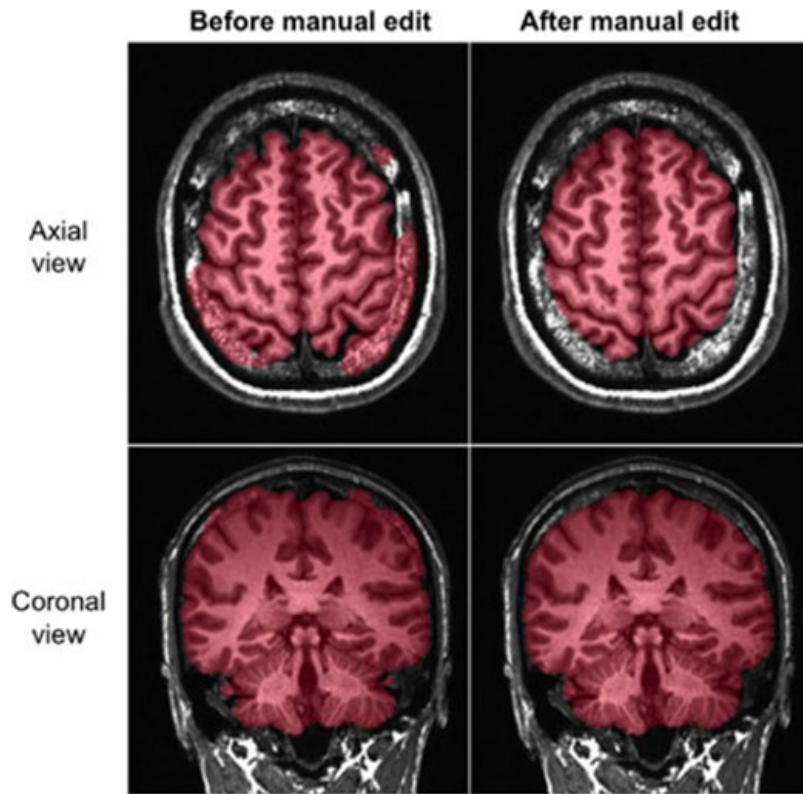


Figure 3-3: Skull stripping dataset process example taken from [81]

In figure 3-4, we show the proposed deep neural network algorithm for the skull stripping preprocessing step using transformer layers as the main block in the architecture. We designed the architecture using a 3D representation in order to have more spatial relationship information.

For the training process, we used the same preprocessing pipeline developed for the Mindboggle-101 dataset (see figure 3-1). The only difference is the label-remapping strategy that, in this case, was used for binary segmentation (label 0 for non-brain tissue, 1 for brain tissue). We explain in detail the self-attention mechanism via transformers in section 3.3 and how it is combined with the rest of the architecture design.

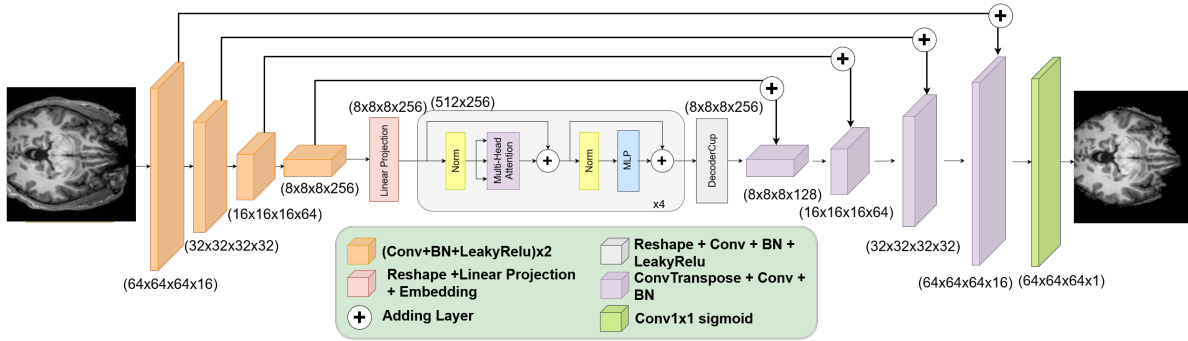


Figure 3-4: Proposed deep neural network architecture for Skull Stripping

3.3. Deep neural network for brain segmentation

The proposed deep neural network architecture is structured as an encoder-decoder architecture. The contracting path follows the typical architecture of a convolutional neural network. However, we applied Transformer layers at the end of this path using the extracted feature maps from the CNN layers. The expansive path was composed of a successive combination of convolutional neural networks and upsampling layers in order to reach the original spatial resolution. This can be seen in figure 3-5. Also, to avoid gradient vanishing and explosion problems, we adopted skip connections between the encoder-decoder paths via the usage of Res paths (see figure 3-6), initially proposed in [68].

We used self-attention mechanisms via Transformers in the encoder path. This consists of successive I layers of Transformers composed of Multi-Head Self-Attention (MHSA) modules and Multi-Layer Perceptron (MLP) blocks, each preceded by a normalization layer. The MLP blocks use the RELU activation function with a regularization dropout layer.

The attention mechanism was computed in parallel inside each of the heads of the MHSA modules in each transformer using a set of vectors named as query, key, and value vectors [20]. The query vector $q \in \mathbb{R}^d$ was matched against all key vectors organized in a matrix $K \in \mathbb{R}^{k \times d}$ using an algebraic operation known as the dot product. The results were then

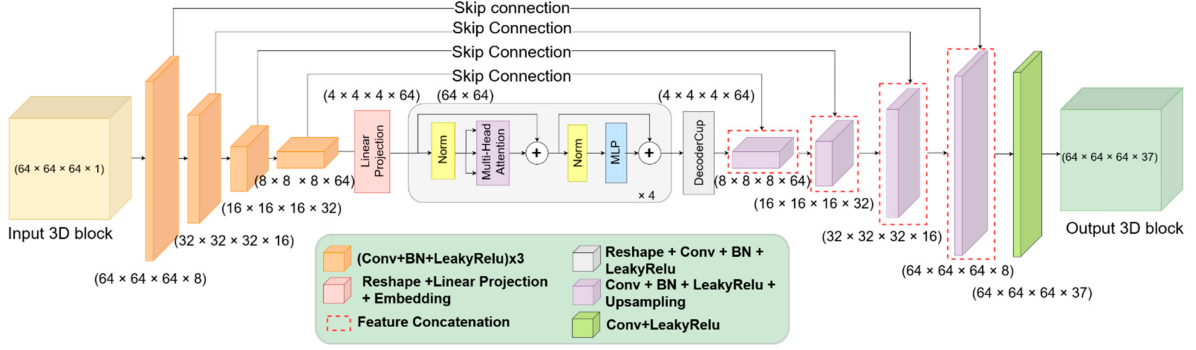


Figure 3-5: Neural network architecture for 3D brain MRI segmentation.

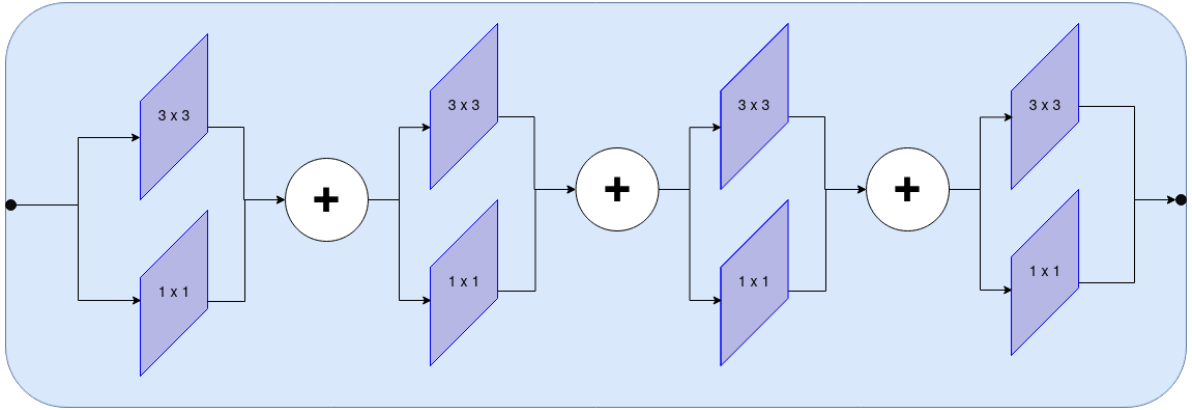


Figure 3-6: Skip connections design based on convolutional layers.

scaled using a scaling factor $\frac{1}{\sqrt{d_k}}$ and normalized using a softmax function to obtain the weights. The attention matrix inside each MHSA head is computed as:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3-1)$$

where Q , K , and V are matrices representing a set of queries, keys, and values, respectively.

Finally, the results of each head were concatenated and linearly projected into a matrix sequence at the end of the MHSA module. This can be described as:

$$MHSA(Q, K, V) = Concat(head_1, head_2, \dots, head_h)W^o \quad (3-2)$$

having $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$ where the projections are parameter

matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$ is the dimension of the Transformer's hidden layers.

In the proposed architecture, we used the extracted feature maps from previous convolutional layers as the input of the first transformer layer, using a trainable linear projection. Indeed, we reshaped the feature maps $X \in \mathbb{R}^{H \times W \times D \times C}$ into a flattened representation as the transformer layers expect a sequence as input. Then, we applied positional embedding over the feature maps to add location information for the segmentation process. This can be described as:

$$f x_i^{q,k,v} = FeatureMapsEmbedding(flatten(x_i)) \quad (3-3)$$

where positional embedding adds location information useful in the segmentation process. This can be seen as:

$$z_0 = [F_1; F_2; F_3; \dots; F_N] + E_{pos}, F \in \mathbb{R}, E_{pos} \in \mathbb{R}^{N \times L} \quad (3-4)$$

where F denotes the feature maps in conjunction with the linear projection and E_{pos} the position embedding and $N = \frac{H \times W \times D \times C}{16}$. After successive layers of Transformers, the output of the last transformer has a shape $z_I \in \mathbb{R}^{d \times N}$. We applied a reshape before the decoder path to recover its 3D dimensionality.

3.4. Loss Functions and Class Weights

3.4.1. Skull stripping preprocessing step

For the segmentation of brain tissue we used a distribution-based loss function called Binary Crossentropy (BCE) [82]. This loss function derives from the Bernoulli distribution and is widely used in the deep learning field [83]. It can be understood as a measure of the difference between two probability distributions for a given random variable or set of events. Formally, it can be defined as:

$$BinaryCrossentropy(x, y) = -(x \log(y) + (1 - x) \log(1 - y)) \quad (3-5)$$

where x represents the ground truth image and y the predicted output given by the

model for the input image.

3.4.2. Segmentation of brain structures

Segmentation of brain structures is a highly imbalanced problem due to the significant differences in size in the structures, presenting greater availability of information in the image for those of greater size. Even the size difference between the structures and the background is usually significant.

Therefore, multiple loss functions and weighting strategies for loss functions were proposed for improving imbalanced brain structure segmentation [84]. In the proposed approach, we used a combination of Dice Loss [85] and Focal Loss [86]. Dice Loss (DL) has its origin in the Dice Similarity Coefficient (DSC), which is widely used as a metric for computer vision segmentation to calculate the similarity between two images.

Later, in [86], it was adapted as a loss function useful in medical image segmentation tasks, improving the imbalance problem between foreground and background. It is formulated as:

$$WeightedDiceLoss = 1 - 2 \frac{\sum_{j=1}^S w_j \sum_{i=1}^N y_{ij} p_{ij}}{\sum_{j=1}^S w_j \sum_{i=1}^N y_{ij} + p_{ij}} \quad (3-6)$$

where w_j is the weight of the j th brain structure and S is the total number of segmentation classes, y_{ij} is the label of voxel i to belong to brain structure j and p_{ij} is the probability of voxel i to belong to brain structure j .

Meanwhile, Focal Loss (FL) is a variation in Binary Cross-Entropy that works better with highly imbalanced datasets. It down-weights the contribution of easy examples and mostly focuses on the hard ones. It can be described as follows:

$$FocalLoss = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3-7)$$

where p_t with $p \in [0, 1]$ is the model's estimated probability for the class, $(1 - p_t)^\gamma$ is the modulating term with γ as the focusing parameter that controls its strength.

The combination of these two loss functions helped us to alleviate the imbalance problem in the segmentation of anatomical brain structures and encourages the correct segmentation of tissue boundaries. Indeed, the use of class weights while training deep

neural network architecture was necessary due to the large number of small structures the brain has compared with the total number of brain voxels. In order to calculate the class weights, we used the median frequency balancing algorithm, which is formulated as follows:

$$\alpha_c = \text{medianFreq}/\text{freq}(c) \quad (3-8)$$

3.5. Data augmentation techniques

There are two main limitations in the process of training a deep neural network architecture in the medical field for classification or detection. The first one is the small data available and the second one is the class imbalance scenario [87]. Therefore, in order to mitigate this problem we used random elastic deformations and random rotations as data augmentation techniques.

Specifically, we applied random rotations by selecting randomly 2 axis and applying the rotation in the remaining axis. On the other side, random elastic deformations in brain imaging is a wide and natural way to apply data augmentation over a dataset. In this technique, the shape, geometry and size of the object are modified imitating the stress field induced by forces over the human living tissue.

An example of applying elastic deformations in a MRI can be seen in figure 3-7. As can be seen in the figure, in order to properly apply elastic deformations in the training dataset, this technique has to be applied in the MRI mask using the same configuration as in the original MRI.

3.6. Metrics and training parameters

In order to evaluate the performance of the proposed segmentation methods, the ground truth and model prediction from the MRIs were compared. The selected metric for this comparison is the DSC which can be seen as a harmonic mean of precision and recall metrics. This metric can be mathematically expressed as:

$$DSC(x, y) = 2 \times \frac{x \cap y}{x + y} \quad (3-9)$$

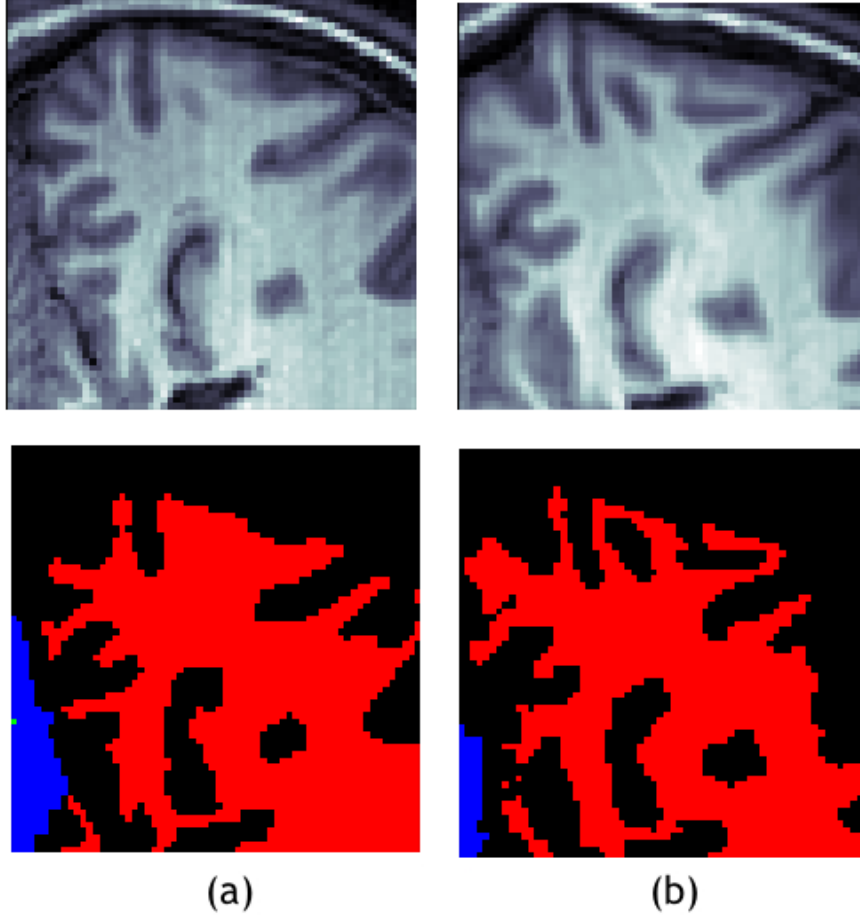


Figure 3-7: Example of elastic deformation in a MRI from the Mindboggle-101 dataset. (a) Original MRI and mask; (b) Elastically deformed images. Source: Author.

where x represents the ground truth image and y represents the predicted output given by the model for the input image. The marked labels for x and y should be binary represented for each class where all voxels included in a given class should have a value of 1 and 0 for all the others. Therefore, the DSC must be calculated individually for each class having output values in a range between 0 (no segmentation) and 1 (perfect segmentation). Consequently, precision is the accuracy of positive predictions while recall is the ratio of positive elements that were predicted correctly by the segmentation model. These metrics can be expressed as:

$$precision = \frac{TP}{TP + FP} \quad (3-10)$$

$$recall = \frac{TP}{TP + FN} \quad (3-11)$$

where TP is the number of true positive predictions, FP the number of false positive predictions and FN the number of false negative predictions for a given class.

Also, the Intersection over Union (IoU) metric is included in this study as an evaluation metric for specific structures. It is useful for comparing similarities between two shapes A and B and determining true positives and false positives from a set of predictions. It can be expressed as:

$$IoU = \frac{A \cap B}{A \cup B} \quad (3-12)$$

The training process used a lineal learning rate schedule, initially set at 0.001 and decreased after the 12th iteration to a power of 0.5, while the batch size is set by default at 8. It used the Adam algorithm as the neural network optimizer. For the transformer architecture based on the Visual Transformer (ViT) architecture [21], we set the successive layers and heads per layer at 4, the hidden size at 64, the MLP size at 192, the dropout rate at 0.1, the normalization rate at 0.0001, and a patch resolution of $8 \times 8 \times 8$. It is important to mention that the hyper-parameters were chosen via experimental design.

4 Results

In this section, the results obtained from the experimentation carried out during the investigation are presented.

4.1. Skull stripping preprocessing step segmentation results

In this subsection, the results of the proposed 3D deep neural network for the segmentation of brain tissue are presented. It is well known in the literature that the process of segmenting brain MRIs using a 3D representation is computationally more expensive than the usual 2D approaches the scientific community has commonly used [74].

However, we consider this approach as the best to be used since it brings contextual information from all the MRI plane perspectives (sagittal, coronal and axial planes) and that the computational resources have continuously increasing allowing the development of these algorithms.

In figure 4-1, we present the segmentation visual results of the proposed architecture. As it can be seen in the image, the segmentation is not perfect. We believe that the voxel value intensity distributions of the skull are, in places close to the brain tissue, very similar to the brain tissue. Then, the model is not able to correctly differentiate between the brain tissue from non-brain tissue. Also, we have to mention that there are multiple artifacts that the process of capturing an MRI from patients leave and that are inherent the process itself.

In table 4-2, the quantitatively evaluation of the binary segmentation model is presented using Precision, Recall and Dice Score for the classes non-brain tissue and brain tissue. In table 4-3, a comparison between state-of-the-art traditional methods, based on CNNs and the proposed method is introduced. Additionally, in table 4-1 we present the segmentation times results (seconds) for the validation dataset in the NFBS repository.

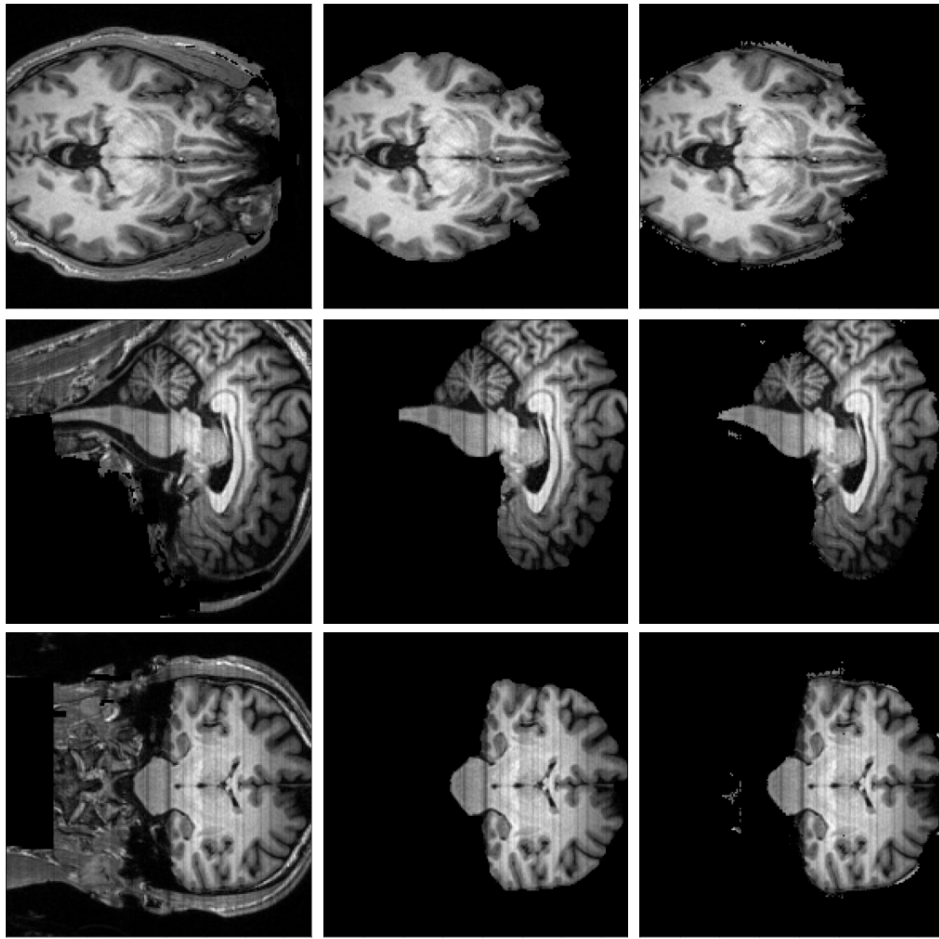


Figure 4-1: Segmentation results of the proposed architecture for skull stripping in the axial, sagittal and coronal planes. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide. Source: Author.

There, we can see that our proposed segmentation method based on self-attention mechanisms is not the best in terms of quality metrics. However, it shows a great performance in segmentation times where transformers take advantage of their parallelized design. It is also worthwhile mentioning that the CNN-based method was published in 2018 where computational resources were more limited than in the moment of the development of this thesis.

Table 4-1: Segmentation time per brain MRI in the NFBS created validation dataset.

MRI name	Segmentation Time (seconds)
A00028185	0,483
A00033747	0,497
A00035072	0,512
A00035840	0,437
A00037848	0,472
A00038998	0,53
A00039431	0,485
A00040193	0,475
A00040573	0,475
A00040628	0,469
A00040944	0,452
A00043520	0,509
A00043704	0,495
A00043722	0,421
A00045590	0,539
A00052560	0,429
A00053851	0,517
A00058999	0,487
A00060632	0,510
A00061204	0,484
A00062210	0,543
A00062266	0,501
A00063008	0,428
A00063589	0,478
A00064081	0,441
Average segmentation time	0,483

Table 4-2: Segmentation results per class in a testing MRI.

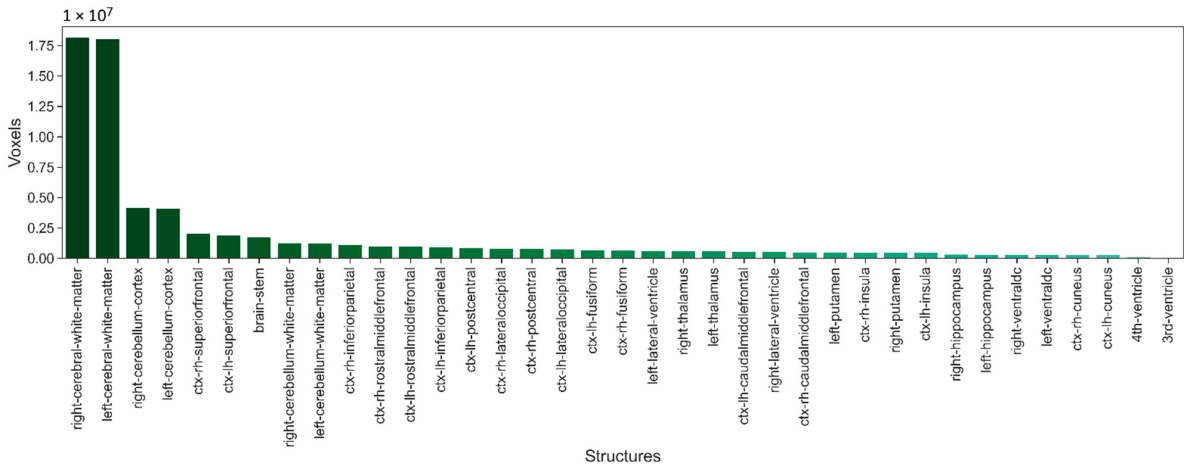
Class	Precision	Recall	Dice Score
Non-brain tissue	0.98	0.99	0.98
Brain tissue	0.94	0.93	0.93
Macro average	0.96	0.96	0.96
Weighted average	0.97	0.97	0.97

Table 4-3: Comparison between methods by the Dice Score and segmentation times using the NFBS dataset.

Skull stripping algorithm	Segmentation Time (s)	Dice Score (%)
AFNI [88]	117,7 \pm 53,3	91,8 \pm 1,0
ANTS [88]	806.9 \pm 87,9	95,2 \pm 0,4
BSE [88]	3,5 \pm 0,7	92,3 \pm 17,2
CNN-based method [88]	4,5 \pm 0,0	96,5 \pm0,4
Proposed Model	0,48 \pm0,06	92,0 \pm 1,4

4.2. Segmentation of brain structures

We quantitatively and visually evaluated the performance of brain structures segmentation. Figure 4-2 shows the number of voxels for each of the 37 selected structures from Mindboggle-101 dataset. It can be seen that there are significant differences between classes. As we mentioned before, to mitigate the effect of class imbalance we use a loss function combining the weighted coefficient Dice and the Focal Loss.

**Figure 4-2:** Number of voxels for each of the 37 selected structures from Mindboggle-101 dataset.

The combination of the aforementioned loss functions and the median frequency balancing algorithm for calculating the class weights allowed us to alleviate the imbalance problem in the segmentation of anatomical brain structures.

A graphic example of the segmentation result of the proposed deep neural network architecture can be seen in figures 4-3 and 4-4, where we show the results in the axial,

sagittal and coronal planes. The local details at the edges of the structures as well as the global features can be noted compared with the reference image.

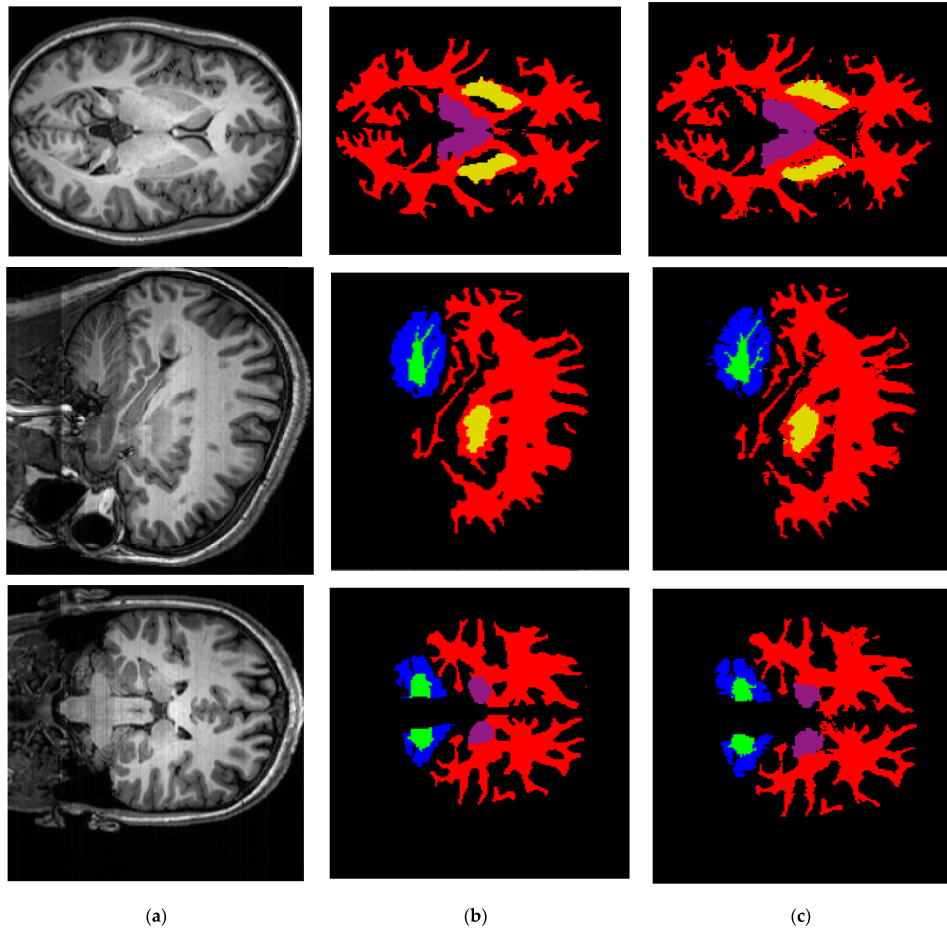


Figure 4-3: Segmentation results of the proposed architecture in the axial, sagittal and coronal planes where red, green, blue, purple, and yellow colors represent cerebral white matter, cerebellum white matter, cerebellum cortex, thalamus, and putamen structures, respectively. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide. Source: Author.

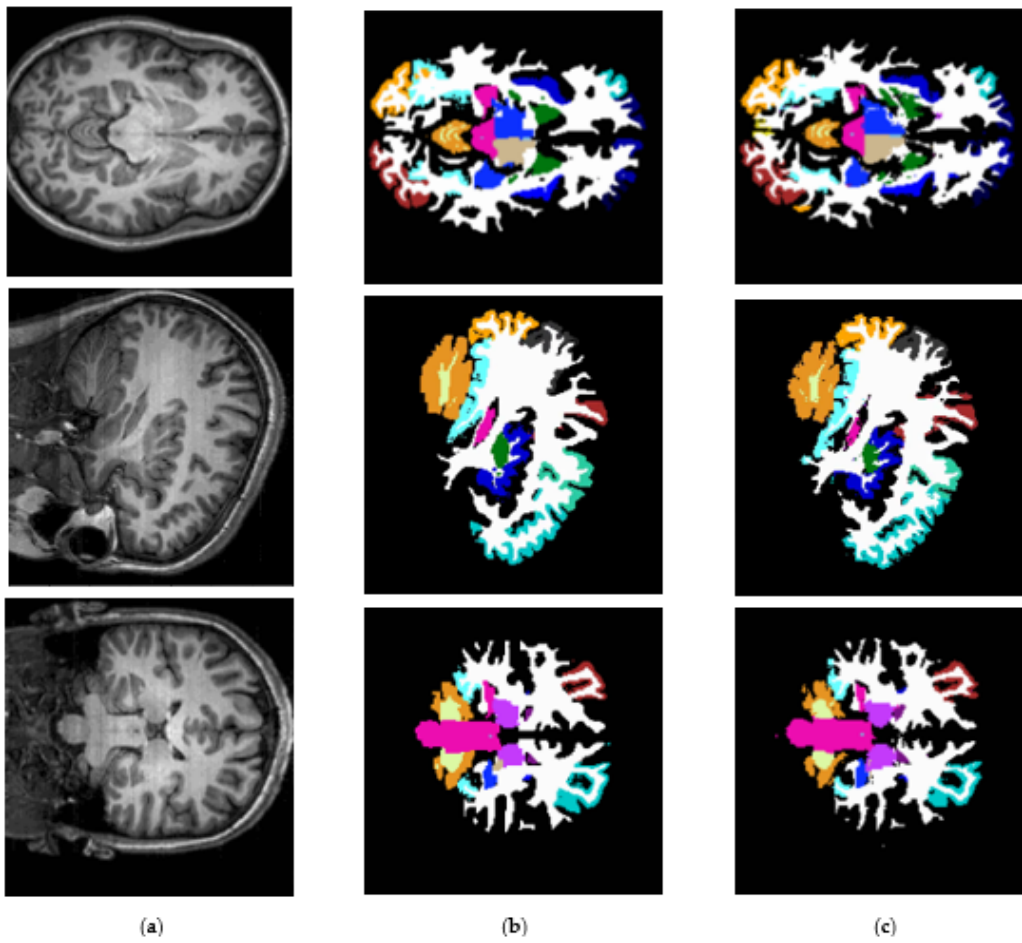


Figure 4-4: Segmentation results of the proposed architecture in the axial, sagittal and coronal planes for all 37 brain structures. (a) Original MRI slide; (b) ground truth mask of the slide; (c) predicted MRI mask of the slide. Source: Author.

Quantitatively, we calculated the Precision, Recall, Dice Score and IoU Score per segmented brain structure. The results in table 4-4 show that there are still problems with the segmentation of some structures, mainly small structures that tend to lower values of quality metrics.

Table 4-4: Segmentation results per brain structure in a testing MRI.

Brain Structure	Precision	Recall	Dice Score	IoU
Left cerebral white matter	0.95	0.91	0.93	0.86
Right cerebral white matter	0.97	0.89	0.93	0.86
Left cerebellum white matter	0.90	0.75	0.82	0.69
Right cerebellum white matter	0.93	0.77	0.85	0.73
Left cerebellum cortex	0.87	0.82	0.84	0.73
Right cerebellum cortex	0.89	0.72	0.80	0.66
Left lateral ventricle	0.64	0.91	0.75	0.60
Right lateral ventricle	0.78	0.91	0.84	0.72
Left thalamus	0.80	0.92	0.86	0.74
Right thalamus	0.90	0.89	0.89	0.80
Left putamen	0.85	0.84	0.85	0.73
Right putamen	0.91	0.81	0.86	0.75
3rd ventricle	0.57	0.96	0.72	0.56
4th ventricle	0.67	0.94	0.78	0.64
Brain stem	0.87	0.93	0.90	0.83
Left hippocampus	0.88	0.67	0.76	0.62
Right hippocampus	0.89	0.80	0.84	0.73
Left ventral DC	0.78	0.83	0.80	0.68
Right ventral DC	0.62	0.87	0.72	0.57
Ctx left caudal middle frontal	0.84	0.43	0.57	0.40
Ctx right caudal middle frontal	0.50	0.24	0.32	0.20
Ctx left cuneus	0.56	0.65	0.60	0.44
Ctx right cuneus	0.54	0.74	0.62	0.46
Ctx left fusiform	0.68	0.61	0.64	0.48
Ctx right fusiform	0.78	0.65	0.71	0.55
Ctx left inferior parietal	0.64	0.54	0.58	0.42
Ctx right inferior parietal	0.60	0.70	0.65	0.49
Ctx left lateral occipital	0.69	0.74	0.71	0.56
Ctx right lateral occipital	0.73	0.69	0.71	0.56
Ctx left post central	0.54	0.82	0.66	0.49
Ctx right post central	0.71	0.70	0.71	0.55
Ctx right rostral middle frontal	0.57	0.81	0.67	0.50
Ctx left rostral middle frontal	0.51	0.82	0.63	0.46
Ctx left superior frontal	0.74	0.81	0.77	0.63
Ctx right superior frontal	0.77	0.79	0.78	0.65
Ctx left insula	0.81	0.84	0.82	0.70
Ctx right insula	0.70	0.87	0.78	0.64
Macro average	0.75	0.78	0.75	0.63
Weighted average	0.97	0.97	0.97	0.95

Although to achieve the training process of the deep neural network architecture the brain is divided into blocks of equal size, the results show that the segmented structures maintain spatial coherence and can recover its representative organic form as can be seen in a 3D visual representation shown in Figure 4-5.

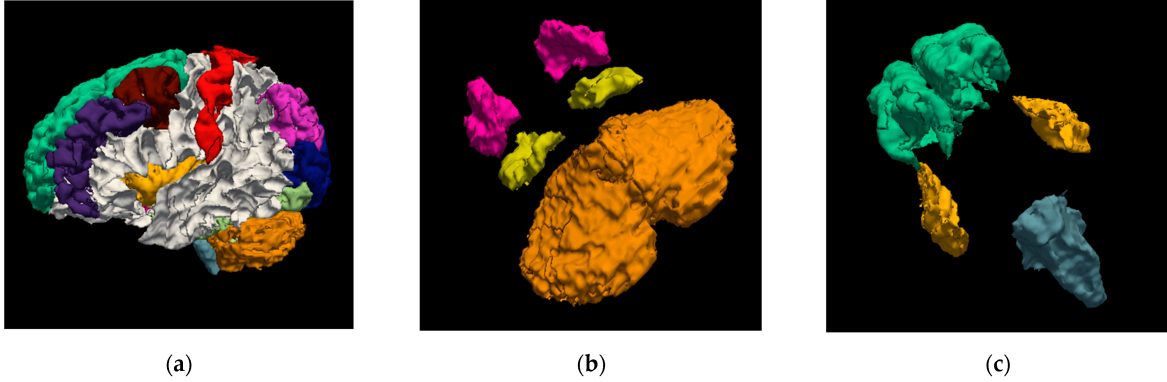


Figure 4-5: Segmentation results of the proposed architecture in 3D. (a) Segmentation of the 37 brain structures; (b) segmentation of the cerebellum cortex (orange), putamen (magenta), and hippocampus structures (yellow); (c) segmentation of the brain stem (gray), insula (yellow), and superior frontal structures (green). Source: Author.

4.2.1. Architecture design determination

The process of designing a deep neural network architecture using transformer layers for obtaining global features of MRI images was developed via experimental design.

First, we tried using the transformer layers at the beginning of the encoder path as a way of getting global features of a full image resolution. In order to do this, we split the MRI at its full resolution into smaller blocks of $64 \times 64 \times 64$ as was explained before but, in this experiment, we gave positional information to these raw split volumes instead of the feature maps. This process and architecture can be seen in figure 4-6.

These experiments showed that by using this design, the deep neural network architecture is not able to recognize local features losing definition in the edges in the final segmentation (see figure 4-7). This is understandable since the nature of the transformer layers comes from the Natural Language Processing (NLP) field and is focused to find relationships between the inputted elements. Nevertheless, it was good at recognizing the area where the brain structure is located in the 3D volumes.

Because of aforementioned issues in the previous architecture design, we tried experi-

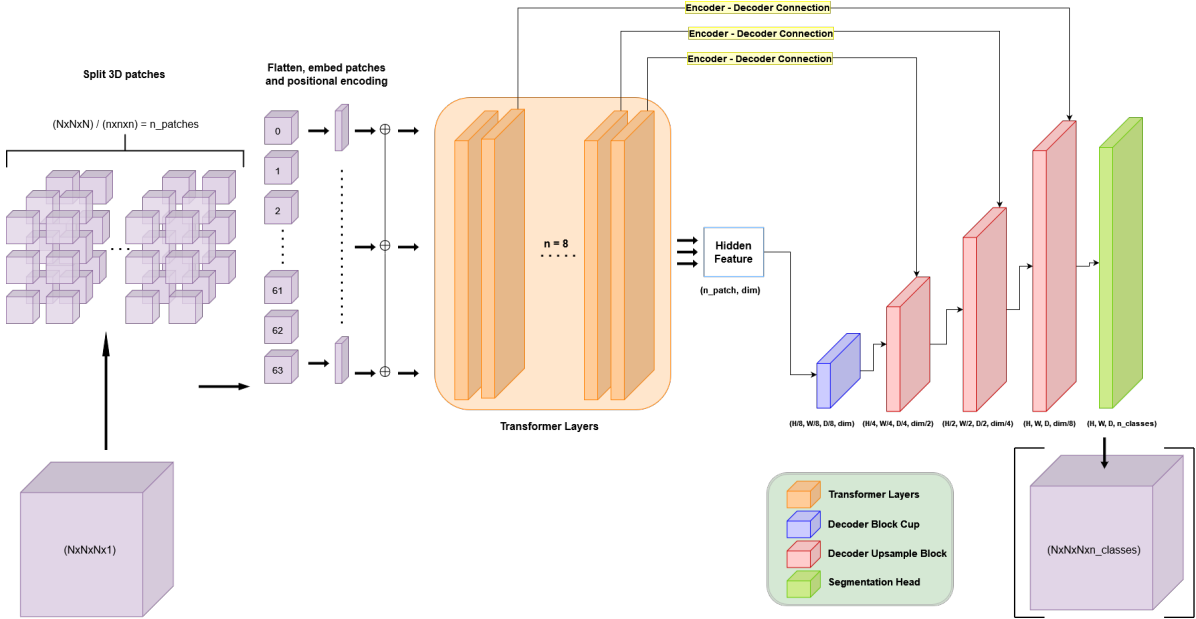


Figure 4-6: Architecture design with Transformer Layers at the beginning of the encoder path. Source: Author.

menting with two branches where one of them was focused on the extraction of globally related features (global context), while the other was in charge of the extraction of local features (local context). This two-branches representation is not new in deep learning architectures and it has been used in multiple medical image proposal such as in [62, 72].

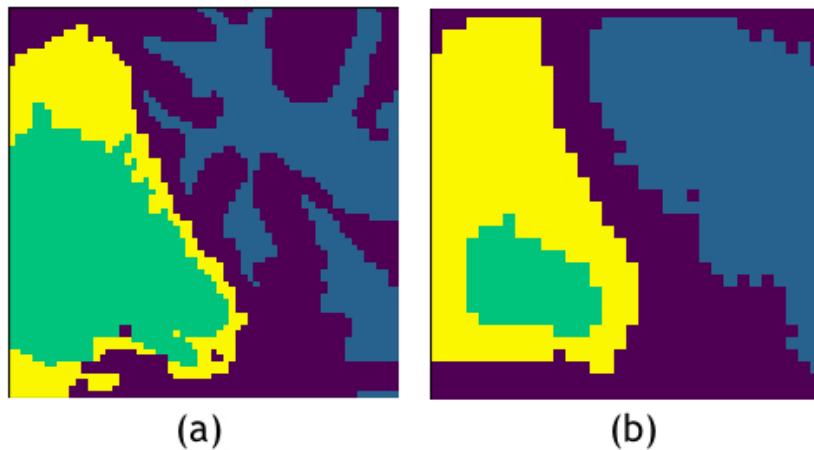


Figure 4-7: Segmentation results of the deep architecture with transformers at the beginning of the encoder path. (a) Ground truth mask; (b) Segmentation result. Source: Author.

Our two-branches proposal used the same volume size of $64 \times 64 \times 64$ where one branch

used the same design proposed in figure 4-6 while the second one used a modified version of MultiResUNet [68] in order to be able to learn from 3D volumes. At the end of the respective branches, we set an adding layer for the combination of the obtained segmentation maps.

However, this architecture did not give us the expected results. The problem of segmenting multiple brain structures is a highly imbalanced problem that requires as much data as available. Then, training a 3D deep neural network architecture that has many millions of parameter is really hard to do in these conditions due to the two branches. In these experiments we used data augmentation techniques such as random rotations and elastic deformations to mitigate this problem.

On the computational side, we had multiple problems training this architecture in our dedicated computer following what is in the literature. The data augmentation techniques, the model and the size of stacks of volumes took the majority of our resources restraining us to do other experiments that big tech companies apply. It is worth remembering that our machine has two Nvidia 1080TI video cards with 11 GB of GDDR 5x RAM.

Therefore, we ran multiple experiments changing the layers to decrease the number of parameters, the loss functions to improve the segmentation definition and also the activation functions to see their effects.

The segmentation results of the two-branch architecture proposal can be seen in figure 4-8. It can be noticed that it is not able to properly segment smaller structures due to the highly imbalanced problem and the size of the architecture. Despite that, these results were more promising in terms of quality metrics and segmentation details such as borders and shape of brain structures thanks to the extraction of local context from the raw volume input through successive layers of convolutional neural networks.

From these experiments, we found that the correct balance between local and global context is crucial for proper segmentation in highly imbalanced datasets and the extraction of local context from the input data is crucial for adding details to the final segmentation.

Therefore, our final proposal shown in figure 3-5 included successive convolutional layers at the beginning of the architecture for local context extraction. Then, we used successive transformer layers for global context extraction from the extracted feature maps of the CNNs.

It is worthwhile mentioning that at this point the encoder path has the most important

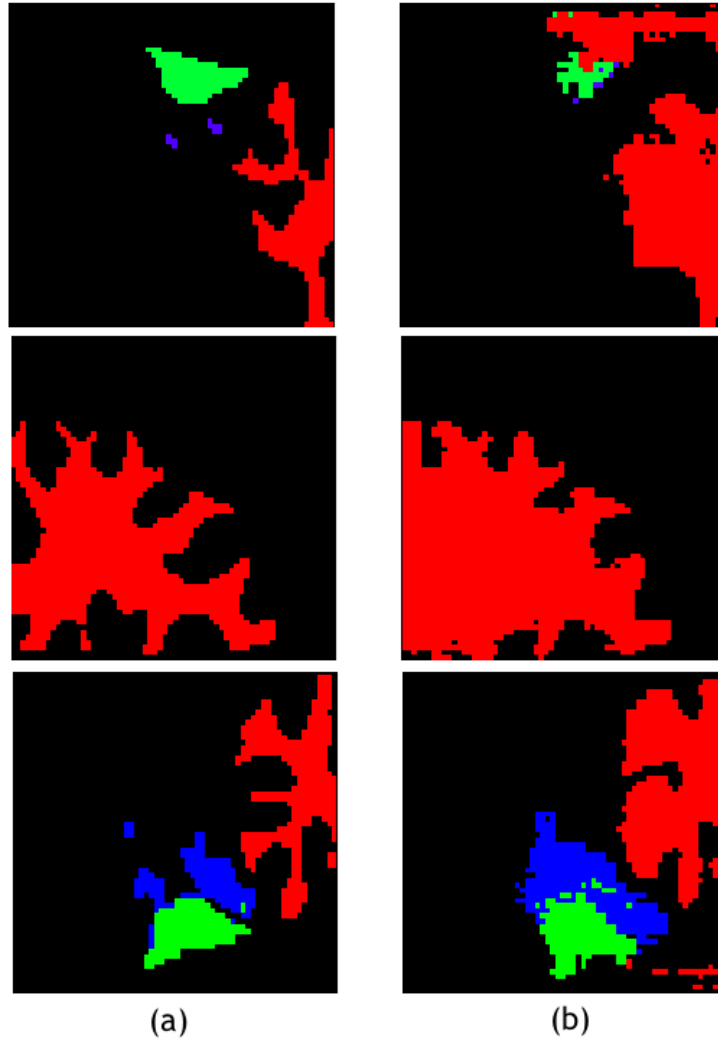


Figure 4-8: Segmentation results of the two-branches network proposal. (a) Ground truth mask; (b) Segmentation result. Source: Author.

features from the 3D image that are useful for global context extraction. Then, we used 3D bilinear upsampling layers combined with CNNs for reaching the original input size resolution for the number of classes. This can be expressed as $64 \times 64 \times 64 \times n_classes$.

Finally, we found that by using skip connections between the encoder and decoder paths based on MultiResUNet [68] paper (see figure 3-6) we improved the segmentation quality metrics in ≈ 2 points. MultiResUNet paper was developed thinking in medical image segmentation, then its use was straightforward for this application. We ran multiple experiments changing the number of internal CNN layers in the skip connection design but it did not improve the segmentation quality results. The usage of only one branch improved our segmentation results for multiple brain structures using an imbalanced

dataset as can be seen in figure 4-4.

In figure 4-9 we can see the Dice Score per epoch in the training and validation sets for the final proposed architecture. There we can see a progressive improvement in the validation and training sets dice score metrics. It is worth noting that, due to the distribution of the dataset explained in section [Data preprocessing](#), we can notice at the beginning of the training that the validation dice score is slightly higher than the training dice score.

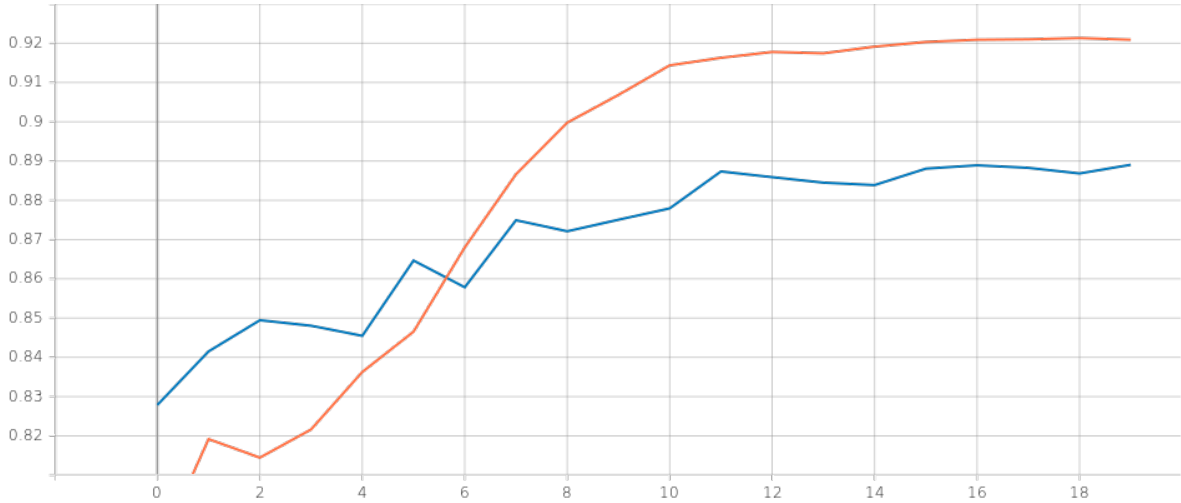


Figure 4-9: Dice score per epoch in the training and validation sets for the proposed architecture. Source: Author.

The same metric behaviour can be seen in figure 4-10 where we show the IoU score for the validation and training sets. The IoU metric has a close mathematical relationship with the Dice Score in the sense that they are both positively correlated and share a similar measurement range. From one to zero indicating a perfect match and completely disjoint, respectively.

Thus, these two metrics are typically considered functionally equivalent. Although, IoU usually penalizes bad classifications more strongly. We include both metrics in this study in order to provide useful information for future research and easy comparison in this area.

Lastly, we include in figure 4-11 the calculated value of the combined loss function (Weighed Dice Loss and Focal Loss) per epoch. As it is mentioned in the literature, the usage of the Leaky ReLU [89] activation function in combination of a He-Normal Initialization [90] technique for the neural network layers was important for a faster convergence of the architecture. The usage of Leaky ReLU attacked the "dying ReLU" problem that could be seen in the training process of the proposed architecture.

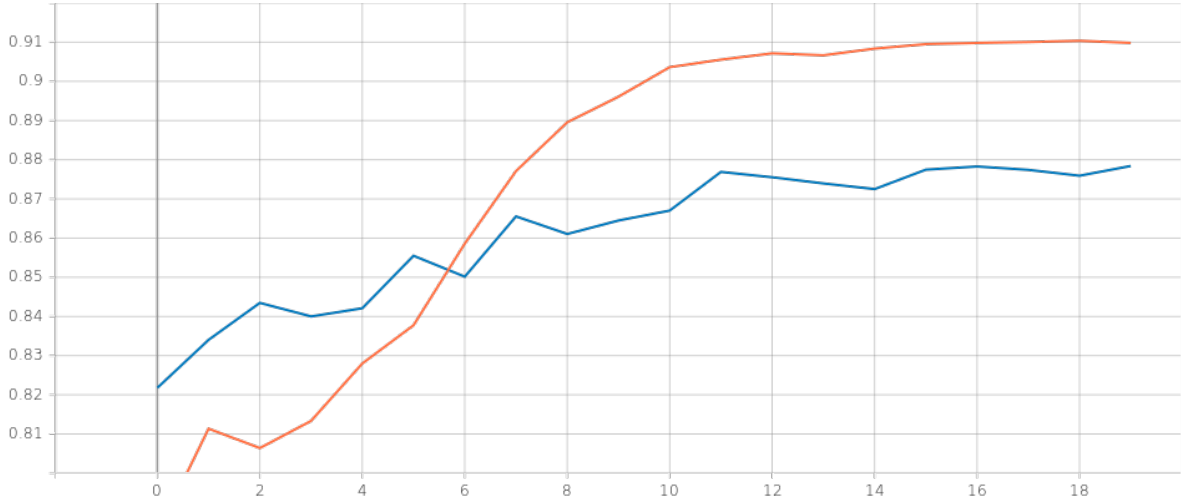


Figure 4-10: IoU score per epoch in the training and validation sets for the proposed architecture. Source: Author.

4.2.2. Patch resolution size determination

We experimented with patch resolution sizes related to the transformer layers. As was initially observed in [22], patch resolution size is important since it dictates the number of complex dependencies that each element will have with others, obtaining finer details in the segmentation process. The ideal case would be to have a patch resolution size of $1 \times 1 \times 1$. Nevertheless, there are not enough computational resources to train a deep neural network architecture based on this patch resolution size.

Consequently, we ran experiments on the segmentation of three structures with patch sizes of $16 \times 16 \times 16$ and $8 \times 8 \times 8$ to see its influence on the segmentation of brain structures (see Figure 4-12). The experiment of a lower patch size was not possible since our computational resources were not enough for this configuration.

For this experiment, we set four successive layers of transformers with patch resolution blocks of $8 \times 8 \times 8$, hidden size at 64, MLP size at 192; dropout rate at 0.1, and normalization rate at 0.0001. Afterwards, we applied a reshape before the decoder path to recover its 3D dimensionality.

4.2.3. Comparison with other methods

At present, the majority of the proposed deep neural network architectures for brain segmentation using Transformers are oriented toward the segmentation of brain tumors.

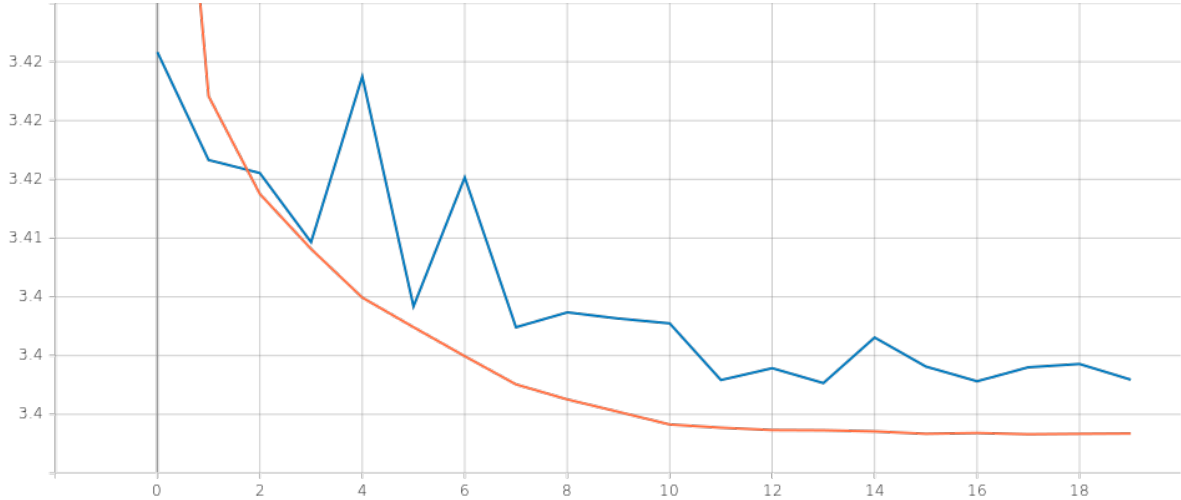


Figure 4-11: Combined loss function per epoch in the training and validation sets for the proposed architecture. Source: Author.

Therefore, it is highly difficult to have a fair comparison since these models were oriented to the segmentation of one imbalanced class, and not for multiple imbalanced classes excluding background.

Because of this, we implemented the 3D U-Net architecture, using it as our baseline, with an identical experimental setup of our proposed architecture. This comparison was carried out by using the Dice score and the Wilcoxon signed-rank test as can be seen in Table 4-5.

Table 4-5: Comparison between methods by the Dice Score and p-value for the Wilcoxon signed-rank test comparing proposed-UNet, proposed-DenseUNet samples pairs using the Mindboggle-101 dataset.

Model	Brain Structures	Mean Dice Score	p-Value
UNet (baseline)	37	0.790 ± 0.0210	0.0012850
DenseUNet (finetuned)	102	0.819 ± 0.0110	0.0211314
Proposed model	37	0.900 ± 0.0360	-

The time needed to perform the segmentation by this architecture and the comparison with other deep learning models is shown in Table 4-6. It is important to mention that transformer layers, thanks to the self-attention mechanism, are capable of processing entire sequences in parallel, optimizing processing times. Unlike CPU processing units, the GPU architecture was specifically designed to process data in parallel, allowing the

proposed model to take full advantage of computational resources and the Transformer’s processing pipeline.

Table 4-6: Segmentation time per brain structure for a single MRI scan.

Model	Brain Structures	Time per Brain Structure	Mean Dice Score
DeepNAT [91]	27	≈ 133 s (on a Multi-GPU Machine)	0.906
QuickNAT [92]	27	$\approx 0,74$ s (on a Multi-GPU Machine)	0.901
DenseUNet	102	0.64 s (± 0.0091 s) (Single GPU Machine)	0.819
FreeSurfer [92]	≈ 190	$\approx 75,8$ s	-
Proposed model	37	0.032 s (± 0.0016 s) (on a Multi-GPU Machine)	0.903

Additionally, in table 4-7 we show the segmentation time taken to segment the MRIs of the validation set. There, we can see that the first MRI to be segmented (HLN-12-12) takes significant more time than the rest of the dataset. This is because the model has to be loaded in our machine before processing the 3D volumes. In this case, this time was considered as part of the segmentation process time.

Table 4-7: Segmentation time per brain MRI in the validation dataset.

MRI name	Segmentation Time (seconds)
HLN-12-12	1,208
HLN-12-6	1,203
MMRR-21-10	1,187
MMRR-21-15	1,190
MMRR-21-1	1,234
MMRR-21-20	1,234
MMRR-21-5	1,203
NKI-RS-22-10	1,234
NKI-RS-22-15	1,218
NKI-RS-22-1	1,218
NKI-RS-22-20	1,234
NKI-RS-22-5	1,219
NKI-TRT-20	1,218
NKI-TRT-20-1	1,203
NKI-TRT-20-20	1,203
OASIS-TRT-20	1,198
OASIS-TRT-20-15	1,203
OASIS-TRT-20-20	1,187
OASIS-TRT-20-5	1,171
Average segmentation time	1,208

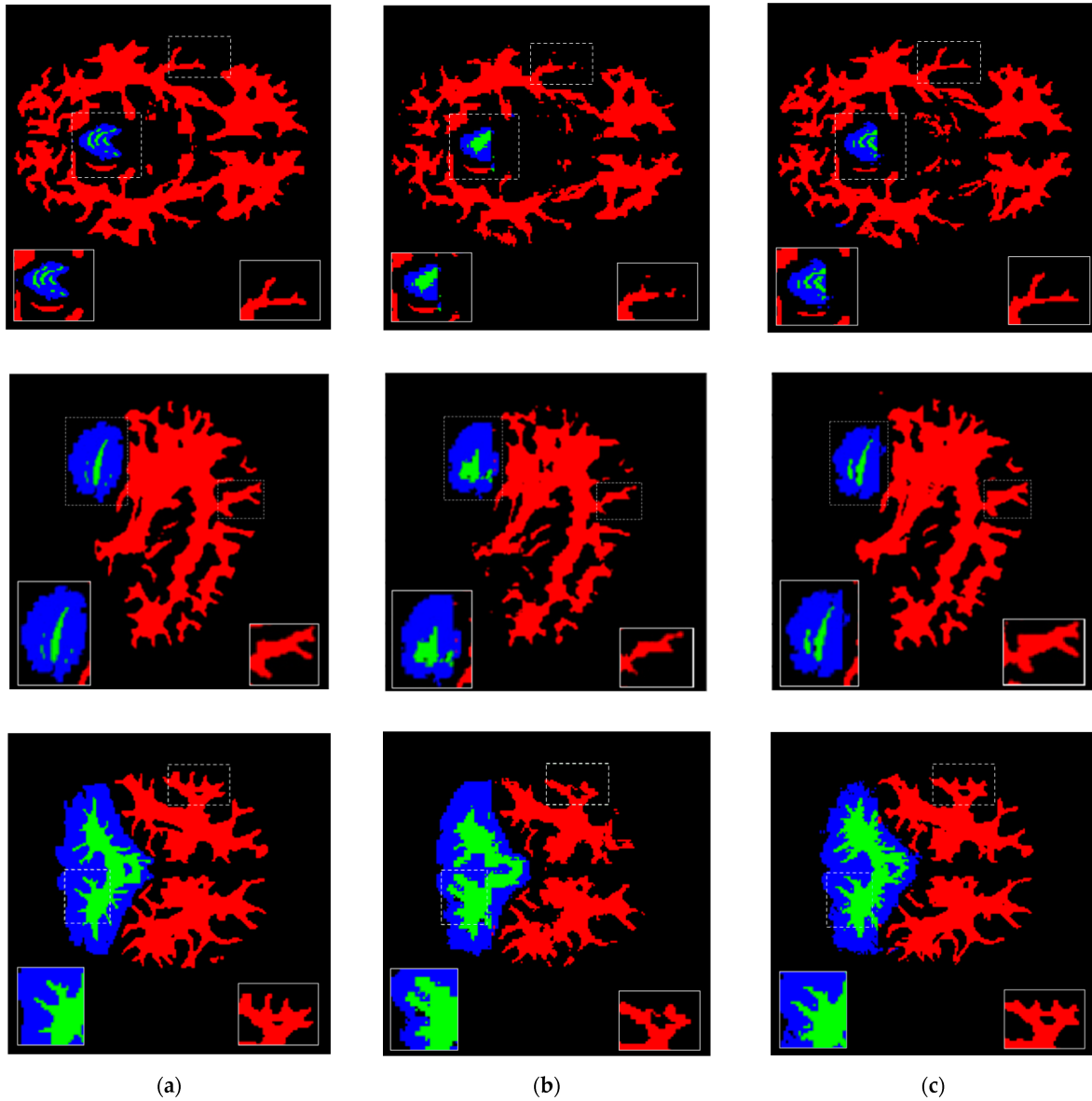


Figure 4-12: Patch size influence comparison on structure details segmentation in the axial, sagittal and coronal planes where red, green, and blue colors represent cerebral white matter, cerebellum white matter and cerebellum cortex structures, respectively. (a) Ground truth mask; (b) segmentation with patch resolution size of $16 \times 16 \times 16$; (c) segmentation with patch resolution size of $8 \times 8 \times 8$. Source: Author.

5 Discussion and future work

This study presents a deep learning-based model for the segmentation of 37 brain structures using transformer models. This network was trained with the manually annotated dataset Mingboggle-101, which contains 101 MRIs with its respective segmentation files processed using the Desikan-Killiany-Tourville (DKT) protocol [73]. In the scientific community, it is common to find multiple approaches to perform the segmentation of MRIs. Therefore, this architecture was indirectly trained to perform segmentation based on the DKT protocol due to the used dataset.

Our architecture includes self-attention modules to strengthen the connection between the encoding and decoding phases based on convolutional neural networks. The capabilities of self-attention modules add to the model the possibility of retaining features across voxels in the input patches of the model. Unlike 2D-based models, the 3D architecture can find voxel relationships in the three different planes, maximizing the use of the spatial nature implicit in MRI.

In addition, the results of the proposed segmentation model show that the quality metrics have a wide range of values. For the Dice Score, for example, the values vary from 0.32 to values of 0.93, showing low-quality segmentations for some structures. We find that the lower values tend to be related to structures with smaller volumes. The geometry at the edges of the structures is a factor that we consider influences the quality of the segmentation. Structures with borders of highly variable geometry tend to have segmentations with more error. Simpler edge structures generally result in more stable quality segmentations.

Our intuition in this regard is that this is due, mainly, to the class imbalance problem and lack of enough data to train the model for those structures specifically. For this reason, it is important in the future to explore other methods that allow addressing this problem, for example, improving the calculation of the weights of the classes used in the loss functions similar to what is performed in [84], or using additional data augmentation techniques to increase the samples of classes with less information.

Another factor that we considered in the analysis is the fact that deep learning methods

based on transformers lack the inductive biases inherent in CNNs requiring large amounts of data to be able to generalize well [21], so their usage in small-size medical datasets remains difficult without any internal modification in their self-attention module. Incorporating these modifications can allow us to improve the segmentation of a large number of highly unbalanced brain structures using a 3D approach.

The patch resolution size is a determining factor to obtain finer details at the edges of segmentations. The experiments show an inverse relationship with size, that is, the smaller the patch size, the more detailed segmentation is obtained at the edges; the larger the patch size, the segmentations tend to be less detailed. It must be considered not all structures have geometrically complex edges, there are structures with simpler geometrics. Therefore, a trade-off between the computational cost of reducing the patch size and the more detailed segmentation requirements must be considered. Given the 3D representation used in this study and the memory requirements, it was not possible to explore values smaller than $8 \times 8 \times 8$.

The results show that our method uses less time for segmentation with a Mean Dice Score similar to those found in the state of the art. Additionally, the segmentation of more than 25 brain structures into a 3D representation is a difficult task that has only been reported by a few groups of authors [58] due to computational and memory limitations. However, it is not competitive in terms of the number of segments where the latest 2D deep learning-based approaches are able to segment more than 100 structures. Consequently, further study should be carried out to optimize the use of memory and computational resources in the proposed architecture to segment more brain structures with a strong focus on the transformer architecture.

On the other hand, due to the lack of a manually skull stripped data for the Mindboggle-101 dataset we had to use the NFBS repository for the skull stripping preprocessing step. It is clear that the usage of two different datasets to solve the problem of segmenting brain structures from MRIs carries some problems. First, we considered that individuals from both datasets do not have the same age distribution. This might be problematic since the brains in both datasets could be different in their structures. However, even if they both have a different age distribution, they are from adult patients. Additionally, these MRIs were taken in different environments with different machine configurations. We do not know if the configuration of the machine that was used to capture the MRIs for both datasets have a final impact in our method. We still have to perform studies and further exploration in order to verify if this have a great impact on the proposed method.

Our method still has deficiencies related to the variation in the segmentation quality for different structures. Class imbalance, as well as the broad geometric nature of the edges

are factors for which our method is still sensitive. The number of segmented structures is also a limitation, it is desirable to be able to segment a greater number of structures, especially compared with 2D-based approaches.

In future work, we will explore the existing computational and memory limitations in our proposed architecture with a high focus on the transformer layers to see whether a different tokenization of the patched feature maps can improve its performance and segment more brain structures.

Bibliografía

- [1] M. M. Miller-Thomas and T. L. Benzinger, “Neurologic applications of pet/mr imaging,” *Magnetic Resonance Imaging Clinics of North America*, vol. 25, no. 2, pp. 297–313, 2017. Hybrid PET/MR Imaging.
- [2] W. Mier and D. Mier, “Advantages in functional imaging of the brain,” *Frontiers in Human Neuroscience*, vol. 9, 2015.
- [3] H. Neeb, K. Zilles, and N. J. Shah, “Fully-automated detection of cerebral water content changes: Study of age- and gender-related h2o patterns with quantitative mri,” *NeuroImage*, vol. 29, no. 3, pp. 910–922, 2006.
- [4] M. E. Shenton, C. C. Dickey, M. Frumin, and R. W. McCarley, “A review of mri findings in schizophrenia,” *Schizophrenia Research*, vol. 49, no. 1, pp. 1–52, 2001.
- [5] G. Widmann, B. Henninger, C. Kremser, and W. Jaschke, “Sequences in head & neck radiology – state of the art mri,” 2017.
- [6] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, “Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives,” *Neurocomputing*, vol. 444, pp. 92–110, 2021.
- [7] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, “A survey on incorporating domain knowledge into deep learning for medical image analysis,” *Medical Image Analysis*, vol. 69, p. 101985, 2021.
- [8] U. Ilhan and A. Ilhan, “Brain tumor segmentation based on a new threshold approach,” *Procedia Computer Science*, vol. 120, pp. 580–587, 2017. 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 22-23 August 2017, Budapest, Hungary.
- [9] W. Deng, W. Xiao, H. Deng, and J. Liu, “Mri brain tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve,” in *2010 3rd International Conference on Biomedical Engineering*

- and Informatics*, vol. 1, pp. 393–396, 2010.
- [10] J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage*, vol. 26, no. 3, pp. 839–851, 2005.
- [11] J. Liu and L. Guo, “A new brain mri image segmentation strategy based on k-means clustering and svm,” in *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 2, pp. 270–273, 2015.
- [12] X. Zhao and X.-M. Zhao, “Deep learning of brain magnetic resonance images: A brief review,” *Methods*, vol. 192, pp. 131–140, 2021. Deep networks and network representation in bioinformatics.
- [13] W. T. Le, F. Maleki, F. P. Romero, R. Forghani, and S. Kadoury, “Overview of machine learning: Part 2: Deep learning for medical image analysis,” *Neuroimaging Clinics of North America*, vol. 30, no. 4, pp. 417–431, 2020. Machine Learning and Other Artificial Intelligence Applications.
- [14] X. Liu, H. Wang, Z. Li, and L. Qin, “Deep learning in eeg diagnosis: A review,” *Knowledge-Based Systems*, vol. 227, p. 107187, 2021.
- [15] A. Nogales, Álvaro J. García-Tejedor, D. Monge, J. S. Vara, and C. Antón, “A survey of deep learning models in medical therapeutic areas,” *Artificial Intelligence in Medicine*, vol. 112, p. 102020, 2021.
- [16] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [17] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, “Review of deep learning: concepts, cnn architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, p. 53, Mar 2021.
- [18] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, and B. Y. Edward, “Capsule networks – a survey,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 1, pp. 1295–1310, 2022.
- [19] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, (Red Hook, NY, USA), p. 3859–3869, Curran Associates Inc., 2017.

- [20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2017.
- [21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2020.
- [22] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, “Transunet: Transformers make strong encoders for medical image segmentation,” 2021.
- [23] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron*, vol. 33, pp. 341–355, Jan. 2002.
- [24] D. W. Shattuck and R. M. Leahy, “Brainsuite: An automated cortical surface identification tool,” *Medical Image Analysis*, vol. 6, no. 2, pp. 129–142, 2002.
- [25] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, “Fsl,” *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012. 20 YEARS OF fMRI.
- [26] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 234–241, Springer International Publishing, 2015.
- [27] I. Despotović, B. Goossens, and W. Philips, “Mri segmentation of the human brain: Challenges, methods, and applications,” *Computational and Mathematical Methods in Medicine*, vol. 2015, p. 450341, Mar 2015.
- [28] P. A. Yushkevich, Y. Gao, and G. Gerig, “ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2016, pp. 3342–3345, Aug. 2016.
- [29] S. Pieper, M. Halle, and R. Kikinis, “3d slicer,” in *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, pp. 632–635 Vol. 1, 2004.
- [30] C. Qin, Y. Wu, W. Liao, J. Zeng, S. Liang, and X. Zhang, “Improved u-net3+ with stage residual for brain tumor segmentation,” *BMC Medical Imaging*, vol. 22, p. 14, Jan 2022.

- [31] J. Sun, Y. Peng, Y. Guo, and D. Li, "Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3d fcn," *Neurocomputing*, vol. 423, pp. 34–45, 2021.
- [32] W. Dai, B. Woo, S. Liu, M. Marques, F. Tang, S. Crozier, C. Engstrom, and S. Chandra, "Can3d: Fast 3d knee mri segmentation via compact context aggregation," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 1505–1508, 2021.
- [33] T. S. Deepthi Murthy and G. Sadashivappa, "Brain tumor segmentation using thresholding, morphological operations and extraction of features of tumor," in *2014 International Conference on Advances in Electronics Computers and Communications*, pp. 1–6, 2014.
- [34] W. Polakowski, D. Cournoyer, S. Rogers, M. DeSimio, D. Ruck, J. Hoffmeister, and R. Raines, "Computer-aided breast cancer detection and diagnosis of masses using difference of gaussians and derivative-based feature saliency," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 811–819, 1997.
- [35] M. Wani and B. Batchelor, "Edge-region-based segmentation of range images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 314–319, 1994.
- [36] J. Wu, F. Ye, J.-L. Ma, X.-P. Sun, J. Xu, and Z.-M. Cui, "The segmentation and visualization of human organs based on adaptive region growing method," in *2008 IEEE 8th International Conference on Computer and Information Technology Workshops*, pp. 439–443, 2008.
- [37] N. Passat, C. Ronse, J. Baruthio, J.-P. Armspach, C. Maillot, and C. Jahn, "Region-growing segmentation of brain vessels: An atlas-based automatic approach," *Journal of Magnetic Resonance Imaging*, vol. 21, no. 6, pp. 715–725, 2005.
- [38] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, "Tumour volume determination from MR images by morphological segmentation," *Physics in Medicine and Biology*, vol. 41, pp. 2437–2446, nov 1996.
- [39] S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms," *Medical Physics*, vol. 23, no. 8, pp. 1337–1345, 1996.
- [40] E. A. A. Maksoud, M. Elmogy, and R. M. Al-Awadi, "Mri brain tumor segmentation system based on hybrid clustering techniques," in *Advanced Machine Learning*

- Technologies and Applications* (A. E. Hassanien, M. F. Tolba, and A. Taher Azar, eds.), (Cham), pp. 401–412, Springer International Publishing, 2014.
- [41] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano, “Combination strategies in multi-atlas image segmentation: Application to brain mr data,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1266–1277, 2009.
- [42] P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, “Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation,” *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.
- [43] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, “Multi-atlas segmentation with joint label fusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, 2013.
- [44] G. Wu, Q. Wang, D. Zhang, F. Nie, H. Huang, and D. Shen, “A generative probability model of joint label fusion for multi-atlas based brain segmentation,” *Medical Image Analysis*, vol. 18, no. 6, pp. 881–890, 2014. Sparse Methods for Signal Reconstruction and Medical Image Analysis.
- [45] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, pp. 321–331, Jan 1988.
- [46] Z. Wu, Y. Guo, S. H. Park, Y. Gao, P. Dong, S.-W. Lee, and D. Shen, “Robust brain roi segmentation by deformation regression and deformable shape model,” *Medical Image Analysis*, vol. 43, pp. 198–213, 2018.
- [47] A. Rajendran and R. Dhanasekaran, “Fuzzy clustering and deformable model for tumor segmentation on mri brain image: A combined approach,” *Procedia Engineering*, vol. 30, pp. 327–333, 2012. International Conference on Communication Technology and System Design 2011.
- [48] H. Khotanlou, J. Atif, O. Colliot, and I. Bloch, “3d brain tumor segmentation using fuzzy classification and deformable models,” in *Fuzzy Logic and Applications* (I. Bloch, A. Petrosino, and A. G. B. Tettamanzi, eds.), (Berlin, Heidelberg), pp. 312–318, Springer Berlin Heidelberg, 2006.
- [49] S. S. Tng, N. Q. K. Le, H.-Y. Yeh, and M. C. H. Chua, “Improved prediction model of protein lysine crotonylation sites using bidirectional recurrent neural networks,” *Journal of Proteome Research*, vol. 21, pp. 265–273, Jan 2022.
- [50] N. Q. K. Le and Q.-T. Ho, “Deep transformers and convolutional neural network in

- identifying dna n6-methyladenine sites in cross-species genomes,” *Methods*, 2021.
- [51] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” 2020.
- [52] G. Montufar, “Restricted boltzmann machines: Introduction and review,” 2018.
- [53] A. Sherstinsky, “Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network,” *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.
- [54] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation Applied to Handwritten Zip Code Recognition,” *Neural Computation*, vol. 1, pp. 541–551, 12 1989.
- [55] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, vol. 36, pp. 193–202, Apr 1980.
- [56] D. Nie, L. Wang, Y. Gao, and D. Shen, “Fully convolutional networks for multi-modality isointense infant brain image segmentation,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1342–1345, 2016.
- [57] S. Bao and A. C. S. Chung, “Multi-scale structured cnn with label consistency for brain mr image segmentation,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 113–117, 2018.
- [58] L. Henschel, S. Conjeti, S. Estrada, K. Diers, B. Fischl, and M. Reuter, “Fastsurfer - a fast and accurate deep learning based neuroimaging pipeline,” *NeuroImage*, vol. 219, p. 117012, 2020.
- [59] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee, and R. Tam, “Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.
- [60] S. Valverde, M. Cabezas, E. Roura, S. González-Vilà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, Àlex Rovira, A. Oliver, and X. Lladó, “Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach,” *NeuroImage*, vol. 155, pp. 159–168, 2017.
- [61] R. E. Gabr, I. Coronado, M. Robinson, S. J. Sujit, S. Datta, X. Sun, W. J. Allen, F. D. Lublin, J. S. Wolinsky, and P. A. Narayana, “Brain and lesion segmentation

- in multiple sclerosis using fully convolutional neural networks: A large-scale study,” *Multiple Sclerosis Journal*, vol. 26, no. 10, pp. 1217–1226, 2020. PMID: 31190607.
- [62] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation with deep neural networks,” *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
- [63] M. Havaei, F. Dutil, C. Pal, H. Larochelle, and P.-M. Jodoin, “A convolutional neural network approach to brain tumor segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (A. Crimi, B. Menze, O. Maier, M. Reyes, and H. Handels, eds.), (Cham), pp. 195–208, Springer International Publishing, 2016.
- [64] L. Chen, P. Bentley, and D. Rueckert, “Fully automatic acute ischemic lesion segmentation in dwi using convolutional neural networks,” *NeuroImage: Clinical*, vol. 15, pp. 633–643, 2017.
- [65] Z. Akkus, I. Ali, J. Sedlar, T. L. Kline, J. P. Agrawal, I. F. Parney, C. Giannini, and B. J. Erickson, “Predicting 1p19q chromosomal deletion of low-grade gliomas from mr images using deep learning,” 2016.
- [66] P. Kumar, P. Nagar, C. Arora, and A. Gupta, “U-segnet: Fully convolutional neural network based automated brain tissue segmentation tool,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3503–3507, 2018.
- [67] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [68] N. Ibtehaz and M. S. Rahman, “Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation,” *Neural Networks*, vol. 121, pp. 74–87, 2020.
- [69] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, “Recent advances in recurrent neural networks,” 2018.
- [70] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2018.
- [71] S. Zheng, J. Lu, H. Zhao, X. Zhu, Z. Luo, Y. Wang, Y. Fu, J. Feng, T. Xiang, P. H. S. Torr, and L. Zhang, “Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers,” 2020.

- [72] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, “Medical transformer: Gated axial-attention for medical image segmentation,” 2021.
- [73] A. Klein and J. Tourville, “101 labeled brain images and a consistent human cortical labeling protocol,” *Frontiers in Neuroscience*, vol. 6, 2012.
- [74] H. Z. U. Rehman, H. Hwang, and S. Lee, “Conventional and deep learning methods for skull stripping in brain mri,” *Applied Sciences*, vol. 10, no. 5, 2020.
- [75] G. Fein, B. Landman, H. Tran, J. Barakos, K. Moon, V. Di Sclafani, and R. Shumway, “Statistical parametric mapping of brain morphology: Sensitivity is dramatically increased by using brain-extracted images as inputs,” *NeuroImage*, vol. 30, no. 4, pp. 1187–1195, 2006.
- [76] J. Acosta-Cabronero, G. B. Williams, J. M. Pereira, G. Pengas, and P. J. Nestor, “The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry,” *NeuroImage*, vol. 39, no. 4, pp. 1654–1665, 2008.
- [77] P. A. Taylor, G. Chen, D. R. Glen, J. K. Rajendra, R. C. Reynolds, and R. W. Cox, “Fmri processing with afni: Some comments and corrections on “exploring the impact of analysis software on task fmri results”,” *bioRxiv*, 2018.
- [78] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, “A reproducible evaluation of ants similarity metric performance in brain image registration,” *NeuroImage*, vol. 54, pp. 2033–2044, Feb 2011. 20851191[pmid].
- [79] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, “Magnetic resonance image tissue classification using a partial volume model,” *Neuroimage*, vol. 13, pp. 856–876, May 2001.
- [80] S. M. Smith, “Fast robust automated brain extraction,” *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [81] B. Puccio, J. P. Pooley, J. S. Pellman, E. C. Taverna, and R. C. Craddock, “The preprocessed connectomes project repository of manually corrected skull-stripped T1-weighted anatomical MRI data,” *GigaScience*, vol. 5, 10 2016. s13742-016-0150-5.
- [82] M. Yi-de, L. Qing, and Q. Zhi-bai, “Automated image segmentation using improved pcnn model based on cross-entropy,” in *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.*, pp. 743–746,

- 2004.
- [83] S. Jadon, “A survey of loss functions for semantic segmentation,” in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, oct 2020.
- [84] T. Sugino, T. Kawase, S. Onogi, T. Kin, N. Saito, and Y. Nakajima, “Loss weightings for improving imbalanced brain structure segmentation using fully convolutional networks,” *Healthcare*, vol. 9, no. 8, 2021.
- [85] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu, eds.), (Cham), pp. 240–248, Springer International Publishing, 2017.
- [86] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” 2017.
- [87] E. Castro, J. S. Cardoso, and J. C. Pereira, “Elastic deformations for data augmentation in breast cancer mass detection,” in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 230–234, 2018.
- [88] G. Valvano, N. Martini, A. Leo, G. Santini, D. Della Latta, E. Ricciardi, and D. Chiappino, “Training of a skull-stripping neural network with efficient data augmentation,” 2018.
- [89] A. L. Maas, “Rectifier nonlinearities improve neural network acoustic models,” 2013.
- [90] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” 2015.
- [91] C. Wachinger, M. Reuter, and T. Klein, “Deepnat: Deep convolutional neural network for segmenting neuroanatomy,” *NeuroImage*, vol. 170, pp. 434–445, 2018. Segmenting the Brain.
- [92] A. Guha Roy, S. Conjeti, N. Navab, and C. Wachinger, “Quicknat: A fully convolutional network for quick and accurate segmentation of neuroanatomy,” *NeuroImage*, vol. 186, pp. 713–727, 2019.