

# Axial multi-layer perceptron architecture for automatic segmentation of choroid plexus in multiple sclerosis

Marius Schmidt-Mengin<sup>a,b</sup>, Vito A.G. Ricigliano<sup>a</sup>, Benedetta Bodini<sup>a,c</sup>, Emanuele Morena<sup>a</sup>, Annalisa Colombi<sup>a</sup>, Mariem Hamzaoui<sup>a</sup>, Arya Yazdan Panah<sup>a</sup>, Bruno Stankoff<sup>a,c</sup>, and Olivier Colliot<sup>a,b</sup>

<sup>a</sup>Sorbonne Université, Paris Brain Institute, Inserm, CNRS, AP-HP, Paris, France

<sup>b</sup>Inria, Aramis project-team, Paris, France

<sup>c</sup>AP-HP, Hôpital Saint-Antoine, Department of Neurology, DMU Neurosciences, Paris, France

## ABSTRACT

Choroid plexuses (CP) are structures of the brain ventricles which produce most of the cerebrospinal fluid (CSF). Several postmortem and in vivo studies have pointed towards their role in the inflammatory processes in multiple sclerosis (MS). Automatic segmentation of CP from MRI thus has high value for studying their characteristics in large cohorts of patients. To the best of our knowledge, the only freely available tool for CP segmentation is FreeSurfer but its accuracy for this specific structure is poor. In this paper, we propose to automatically segment CP from non-contrast enhanced T1-weighted MRI. To that end, we introduce a new model called "Axial-MLP" based on an assembly of Axial multi-layer perceptrons (MLPs). This is inspired by recent works which showed that the self-attention layers of Transformers can be replaced with MLPs. This approach is systematically compared with a standard 3D U-Net, nnU-Net, FreeSurfer and FastSurfer. For our experiments, we make use of a dataset of 141 subjects (44 controls and 97 patients with MS). We show that all the tested deep learning (DL) methods outperform FreeSurfer (Dice around 0.7 for DL vs 0.33 for FreeSurfer). Axial-MLP is competitive with U-Nets even though it is slightly less accurate. The conclusions of our paper are two-fold: 1) the studied deep learning methods could be useful tools to study CP in large cohorts of MS patients; 2) Axial-MLP is a potentially viable alternative to convolutional neural networks for such tasks, although it could benefit from further improvements.

**Keywords:** MRI, brain, segmentation, choroid plexus, multiple sclerosis, deep learning, neural networks

## 1. INTRODUCTION

Choroid plexuses (CP) are found in the brain ventricles and produce most of the cerebrospinal fluid (CSF). They are part of the blood–CSF barrier and a gateway for certain substances in and out of the brain. They have been shown to be associated with several diseases including Alzheimer’s disease<sup>1,2</sup> stroke,<sup>3</sup> complex regional pain syndrome<sup>4</sup> and leptomeningeal carcinomatosis.<sup>5</sup> More specifically, there is a growing interest for the study of CP in multiple sclerosis (MS). Several post-mortem studies have shown the involvement of CP in the inflammatory process.<sup>6–8</sup> In vivo, signal changes in the CP have been reported in patients with MS.<sup>9</sup> Very recently, it was demonstrated in vivo that CP are enlarged and inflamed in patients with MS.<sup>10</sup> However, the role of CP in MS pathophysiology, as well as their potential relationship to disease evolution, are not entirely elucidated. There is thus a strong interest in studying CP in large populations of MS patients.

Manual segmentation of CP is time-consuming. As far as we know, the only freely available software for their automatic segmentation is FreeSurfer.<sup>11</sup> Nevertheless, its segmentation of CP is highly inaccurate.<sup>12,13</sup> Hence, the development of more accurate automatic CP segmentation methods would be very useful to allow the study of large cohorts. To the best of our knowledge, only two other approaches have been proposed. The first one<sup>12</sup> clusters voxel intensities in order to separate CP voxels from CSF and ventricular wall inside the ventricles. The segmentation of the ventricles first needs to be obtained, for example by using FreeSurfer. The second one<sup>13</sup>

---

Further author information: (Send correspondence to O.C)

O.C: E-mail: olivier.colliot@sorbonne-universite.fr

uses a 3D U-Net.<sup>14,15</sup> However, it included a very small sample of ten patients and only reported leave-one-out cross-validation results without evaluation on a proper independent testing set. Moreover, neither of these two methods was designed or validated for patients with MS, so it is unclear how they would perform in this case.

Convolutional neural networks have been top performing in many computer vision benchmarks for several years. But recently, Transformer-based architectures<sup>16</sup> have been able to outperform convolutional models on several tasks. Most recent approaches split an image into a grid of patches,<sup>17</sup> which are flattened to vectors and passed to a Transformer as a sequence for classification. SwinTransformers<sup>18</sup> enhanced this architecture for segmentation and object detection. Some works have successfully applied such architectures to medical images.<sup>19–21</sup> Even more recently, it has been shown that the self-attention layers of Transformers can be replaced by simple fully connected layers or multi-layer perceptrons (MLPs).<sup>22–24</sup>

In this paper, we introduce a new model called "Axial-MLP", which is based on an assembly of Axial MLPs, for CP segmentation. We systematically compare this new approach to 3D U-Net, nnU-Net and Freesurfer.

## 2. MATERIALS AND METHODS

### 2.1 Materials

Data originated from three prospective studies performed between May 2009 and September 2017 at the Paris Brain Institute. The studies were approved by the local ethics committees and written informed consent was obtained from all participants. We studied 141 subjects: 44 healthy controls, 61 patients with relapsing remitting MS and 36 patients with progressive MS (13 patients with primary progressive MS and 23 patients with secondary progressive MS).

Images were acquired<sup>10</sup> on two different Siemens 3T MRI scanners (Trio and Prisma) with a 32-channel head coil (92 participants on a Siemens Trio and 49 on a Siemens Prisma). The sequence used for this work was a 3D T1-weighted magnetization-prepared rapid gradient-echo imaging (repetition time = 2300 ms, echo time = 2.98 ms). Other imaging sequences were acquired but were not used in this work. Dataset characteristics are reported in the first line of Table 1.

On each T1-weighted MRI, the choroid plexuses in the two lateral ventricles were segmented by a junior neurologist and corrected by a senior neurologist with 12 years of expertise in MRI processing. More details about image acquisition and manual segmentation procedures can be found in.<sup>10</sup> We divided the population using a stratified split into a training set (91 subjects) and a testing set (50 subjects). The training set was used for cross-validation in order to design and select the models. The testing set was left untouched until the very end of the study.

Table 1. Dataset characteristics: whole dataset, training set (used for cross-validation) and testing set. F: Female, M: Male, T: Siemens Trio MRI scanner, P: Siemens Prisma MRI scanner, HC: healthy control, R: relapsing remitting MS, P1: primary progressive MS, P2: secondary progressive MS.

	Sex	Age	Scanner	Disease Type
All	141 (70F / 71M)	40.7 ± 12.5	92T 49P	44HC 61R 13P1 23P2
Training	91 (45F / 46M)	40.7 ± 12.8	60T 31P	28HC 40R 8P1 15P2
Testing	50 (25F / 25M)	40.7 ± 12.1	32T 18P	16HC 21R 5P1 8P2

### 2.2 Proposed Axial-MLP architecture for CP segmentation

Most Transformer-based vision models<sup>17</sup> partition the input image into a grid of patches and then flatten these patches into embedding vectors. The resulting embeddings are then passed through a Transformer.<sup>16</sup> It has been shown that it is possible to replace the self-attention layers of the Transformer by simple fully-connected layers or MLPs with negligible loss of accuracy.<sup>22–24</sup> We posit that this may be particularly well-suited to our CP segmentation task, as we have registered all images to a common space, intuitively reducing the benefit of

translation invariance. Furthermore, in order to reduce the number of parameters and to increase the inductive bias towards image processing, we apply our MLPs axially, as in Axial Attention Transformers.<sup>25</sup> We call the resulting architecture **Axial-MLP**.

Formally, we start by dividing our input image into non-overlapping patches, resulting in a tensor of shape  $B \times N_d \times N_h \times N_w \times s_d \times s_h \times s_w \times c$  where  $B$  is the batch size,  $N_i, i \in \{d, h, w\}$  are the number of patches along the depth, height and width,  $s_i \in \{d, h, w\}$  are the size of the patches along these dimensions, and  $c$  is the number of channels of the input image. In practice, as the shape of the input image is not necessarily a multiple of the patch size, we start by downsampling it to a multiple of the patch size with trilinear interpolation. Then, we expand this tensor along the channel axis with a fully connected layer. This is comparable to the first convolutional layer of a U-Net. We denote by  $f$  the number of filters of this first layer. The resulting tensor has shape  $B \times N_d \times N_h \times N_w \times s_d \times s_h \times s_w \times f$ . We then pass this tensor through  $L$  layers of our model. A layer consists of 6 fully connected layers. Each fully connected layer is applied along a specific dimension ( $N_d, N_h, N_w, s_d, s_h$ , or  $s_w$ ) plus the channel dimension. The results are passed through a leaky ReLU (negative slope = 0.01) activation function and summed. The resulting tensor is normalized across all but the batch dimension (similarly to Layer Normalization<sup>26</sup>) and we apply a single trainable scalar weight and bias to the output (as opposed to Layer Normalization, where there is one such weight per element). Before each fully connected layer, we apply dropout with a drop rate of 0.02. The dropout is applied along the dimension of the corresponding fully connected layer, meaning that, similarly to Spatial Dropout,<sup>27</sup> the other dimensions are dropped together. Finally, we rearrange the patches into an image and upsample the result so that it has the same shape as the original input image. The whole model is depicted in Figure 1.

We optimized this neural network with Adam<sup>28</sup> ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-8}$ ), with an initial learning rate of  $10^{-2}$  decayed to  $10^{-3}$  after 150 epochs. The training was stopped after 200 epochs. The batch size was 4. We used simple data augmentations using affine transformations provided by TorchIO<sup>29</sup> (scale  $\in [0.9, 1.2]$ , rotation  $\in [-10, 10]$  degrees, translation  $\in [-5, 5]$  voxels).

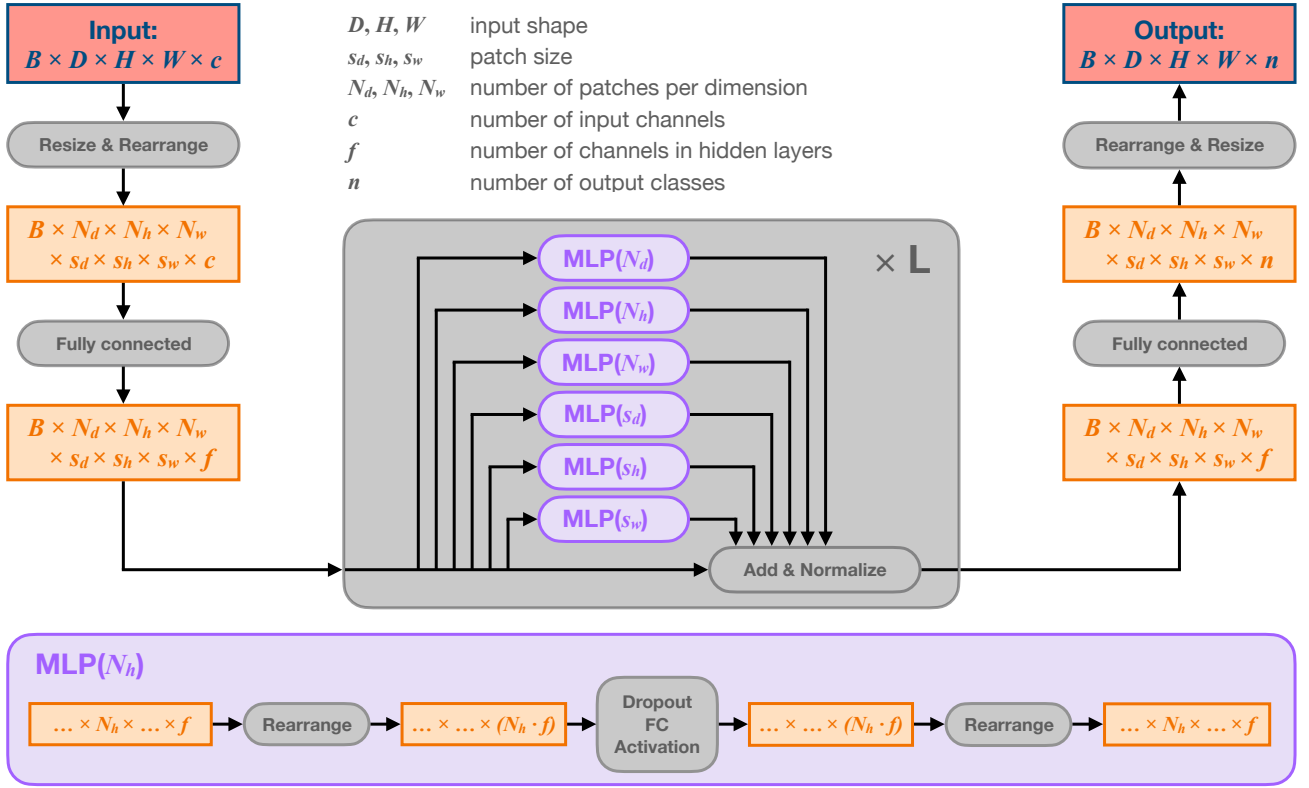


Figure 1. Schematic diagram of the proposed Axial-MLP for segmentation.

## 2.3 Comparison to other approaches

We compared our method to several approaches, ranging from trivial techniques to recent U-Net models.

### 2.3.1 Mean and optimized mask

As the images are registered to a common space, we first established a trivial baseline that does not take into account the MRI images. Given a set of training ground truth masks  $\mathcal{T}$ , we computed

$$\text{Mean mask} = \frac{1}{|\mathcal{T}|} \sum_{y \in \mathcal{T}} y \quad \text{Optimized mask} \approx \operatorname{argmax}_{\bar{y}} \sum_{y \in \mathcal{T}} \text{Dice}(y, \bar{y})$$

The optimized mask is computed approximately by starting from the mean mask and applying 100 gradient descent steps using the Adam optimizer<sup>28</sup> (learning rate 1,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-8}$ ).

### 2.3.2 FreeSurfer and FastSurfer

We performed the segmentation using the **recon-all** pipeline of FreeSurfer.<sup>11</sup> For comparison, we also applied FastSurfer, a deep learning-based approach that aims at reproducing the outputs of FreeSurfer while being considerably faster.<sup>30</sup> Both of these methods were applied to the original, non registered, images.

### 2.3.3 3D-U-Net

We trained a simple 3D U-Net.<sup>14</sup> It’s configuration was identical to the one described in the original paper, except that we varied the initial number of filters and used Instance Normalization<sup>31</sup> and Leaky ReLU. We used the same optimizer and augmentations as for Axial-MLP, except that the learning rate was initially set to  $10^{-3}$  instead of  $10^{-2}$  and decayed to  $10^{-4}$  instead of  $10^{-3}$  after 150 epochs. The training was also stopped after 200 epochs and we used a batch size of 4.

### 2.3.4 nnU-Net

We also applied nnU-Net,<sup>32</sup> a method that aims to automatically determine many hyperparameters of a U-Net in order to optimize its performance on a given dataset. nnU-Net trains up to 3 different configurations: a 2D U-Net that operates on slices, a 3D U-Net that operates on the full resolution images but has to be trained on patches if the images are too large, and a 3D U-Net cascade in which the first U-Net operates on low resolution images, and the second one is applied at full resolution but patch-wise in order to refine the predictions of the first U-Net. In our case, the 3D U-Net cascade was not trained, as the images are small enough to be handled by the full resolution 3D U-Net without patches. For the 2D and 3D U-Net, the following hyperparameters were determined by nnU-Net: no resampling, 4 downsamplings/upsamplings, 32 features in the first stage (doubled in every stage), batch size of 4 for 3D and 308 for 2D. The 3D nnU-Net trainings failed to stop by themselves within 48 hours, so they were killed.

## 2.4 Preprocessing

For FreeSurfer and FastSurfer, we did not preprocess the images. For all other methods, we ran the **t1-linear** pipeline of Clinica.<sup>33,34</sup> It performs bias field correction using the N4ITK method<sup>35</sup> and affine registration using the SyN algorithm<sup>36</sup> from ANTs<sup>37</sup> to align each image to the MNI space with the ICBM 2009c nonlinear symmetric template.<sup>38,39</sup> The resulting images have 1 mm isotropic voxels. We then applied the computed transformations to the segmentation masks using a nearest neighbor interpolation kernel. Finally, we cropped all images using a bounding box of  $102 \times 94 \times 76$  voxels that encompasses all segmentation masks from the training set plus a 10 voxel margin. For all neural network trainings, we first apply a  $z$ -normalization to each image.

## 2.5 Evaluation

All experiments were performed using 32GB Nvidia V100 GPUs.

### 2.5.1 Evaluation metrics

We reported the following metrics:

$$\begin{aligned} \text{Precision } P &= \frac{\sum_{i \in I} x_i y_i}{\sum_{i \in I} x_i} & \text{Recall } R &= \frac{\sum_{i \in I} x_i y_i}{\sum_{i \in I} y_i} & \text{Dice score } D &= \frac{2 \sum_{i \in I} x_i y_i}{\sum_{i \in I} x_i + y_i} = \frac{2}{\frac{1}{P} + \frac{1}{R}} \\ \text{Volume error rate } VER &= \frac{\sum_{i \in I} x_i - y_i}{\sum_{i \in I} y_i} & \text{Absolute volume error rate } AVER &= |VER| \end{aligned}$$

where we denote by  $I$  the set of voxels of a given image,  $x_i$  is the prediction for voxel  $i \in I$  and  $y_i$  the corresponding ground truth label.

We also reported Pearson’s  $r$  between the predicted volumes and the ground truth volumes. All these formulas remain valid and make sense when  $x \in [0, 1]$ . Hence, we chose not to binarize the output of our models, as we believe that it would unnecessarily destroy some information.

The volume error rate, once averaged over the dataset, reflects whether a particular method tends to over or under-estimate the volume, whereas the absolute volume error rate measures the overall accuracy of the estimated volumes. It can be shown that  $D - 1 \leq VER \leq 1/D - 1$ , so that a high enough Dice score ensures a low volume error rate.

### 2.5.2 Validation procedure

We first used a 5-fold cross-validation (CV) on the training set (except for FreeSurfer and FastSurfer, for which there is no need to use CV as they do not require any form of tuning). This means that the training set was split into training and validation sets while the test set was left untouched. We used the same 5 splits for all methods. On each fold, the model with best validation accuracy over the course of the training was selected to perform inference on the fold’s validation set, and the predictions of all 5 folds were aggregated into one large validation set, over which the evaluation metrics were computed. Using the best model of each fold can cause the CV metrics to be over-estimated, but this is not problematic as we have a testing set that was not used until the end of the study and only used for the final evaluation. For inference on this testing set, we averaged the predictions from the best models of each CV fold.

## 3. RESULTS

Cross-validation results are presented in Table 2. We experimented with different configurations of the standard 3D U-Net (4, 8, 16 and 32 initial filters) and the proposed Axial-MLP (4, 8 and 16 hidden channels). Their cross-validation Dice coefficients, as well as resource consumption, are reported in Table 4. We judged that Axial-MLP 8 and 3D U-Net 8 offered the best CV accuracy vs training time compromise, and therefore only these architectures were applied to the testing set (in order not to overfit this hyperparameter). 3D U-Net models were slightly better than Axial-MLP in terms of Dice coefficient, but their number of parameters is also higher.

Results on the testing set are presented in Table 3. As expected, FreeSurfer and FastSurfer resulted in poor segmentation quality (Dice around 0.3-0.4). On the testing set, all deep learning (DL) models outperformed FreeSurfer by a large margin (Dice between 0.71 and 0.72, vs 0.33-0.35 for FreeSurfer and FastSurfer). Dice coefficients of all DL models were comparable. Nevertheless, nnU-Net performed slightly better on almost all metrics. The computation time it used for training was considerably higher (nnU-Net was trained for 48(3D)+12(2D) hours per fold, instead of under 2 hours per fold for all U-Nets and Axial-MLPs). The difference in Dice score between 3D U-Net 8 and Axial-MLP 8 (respectively between nnU-Net 3D and Axial-MLP 8) was small but statistically significant (3D U-Net vs Axial-MLP 8, Dice: 0.72 vs 0.71, paired Student’s t-test,  $T = 4.8$ ,  $p < 10^{-4}$ ; nnU-Net 3D vs Axial-MLP 8, Dice: 0.72 vs 0.71, paired Student’s t-test,  $T = 6.5$ ,  $p < 10^{-7}$ ). Anecdotaly, the constant solution ”optimized mask” had about the same Dice as FreeSurfer. Illustrative examples of automatic and manual CP segmentations are presented on Figure 3 for FreeSurfer and FastSurfer and on Figure 2 for the various deep learning models.

Table 2. 5-fold cross-validation results on the training set. Note that the reported standard deviations are the empirical values computed over the 5 folds and are thus biased estimates of the standard-deviation. Pearson’s  $r$  is not applicable to the mean and optimized mask methods as these are constant across all samples. MVER: Mean volume error rate. MAVER: Mean absolute volume error rate.

	Dice	Precision	Recall	MVER	MAVER	Pearson’s $r$
Mean mask	$0.22 \pm 0.05$	$0.22 \pm 0.07$	$0.23 \pm 0.05$	$0.10 \pm 0.36$	$0.26 \pm 0.26$	N/A
Optimized mask	$0.38 \pm 0.11$	$0.29 \pm 0.10$	$0.57 \pm 0.16$	$1.08 \pm 0.68$	$1.09 \pm 0.68$	N/A
FreeSurfer	$0.33 \pm 0.09$	$0.42 \pm 0.11$	$0.28 \pm 0.09$	$-0.32 \pm 0.22$	$0.35 \pm 0.17$	0.65
FastSurfer	$0.34 \pm 0.09$	$0.41 \pm 0.11$	$0.30 \pm 0.09$	$-0.24 \pm 0.24$	$0.29 \pm 0.17$	0.57
nnU-Net 2D	$0.70 \pm 0.06$	$0.69 \pm 0.09$	$0.72 \pm 0.07$	$0.08 \pm 0.26$	$0.17 \pm 0.22$	0.77
nnU-Net 3D	$0.73 \pm 0.06$	$0.72 \pm 0.09$	$0.75 \pm 0.06$	$0.08 \pm 0.26$	$0.16 \pm 0.23$	0.82
3D U-Net 8	$0.72 \pm 0.06$	$0.69 \pm 0.10$	$0.77 \pm 0.06$	$0.17 \pm 0.34$	$0.20 \pm 0.32$	0.77
Axial-MLP 8 (proposed)	$0.71 \pm 0.06$	$0.67 \pm 0.10$	$0.78 \pm 0.06$	$0.21 \pm 0.35$	$0.24 \pm 0.33$	0.75

Table 3. Results on the independent testing set. Pearson’s  $r$  is not applicable to the mean and optimized mask methods as these are constant across all samples. MVER: Mean volume error rate. MAVER: Mean absolute volume error rate.

	Dice	Precision	Recall	MVER	MAVER	Pearson’s $r$
Mean mask	$0.23 \pm 0.05$	$0.24 \pm 0.07$	$0.24 \pm 0.04$	$0.13 \pm 0.43$	$0.29 \pm 0.34$	N/A
Optimized mask	$0.41 \pm 0.10$	$0.31 \pm 0.10$	$0.62 \pm 0.14$	$1.16 \pm 0.82$	$1.16 \pm 0.82$	N/A
FreeSurfer	$0.33 \pm 0.08$	$0.42 \pm 0.10$	$0.29 \pm 0.08$	$-0.29 \pm 0.18$	$0.32 \pm 0.14$	0.65
FastSurfer	$0.35 \pm 0.09$	$0.40 \pm 0.10$	$0.32 \pm 0.10$	$-0.18 \pm 0.23$	$0.25 \pm 0.15$	0.59
nnU-Net 2D	$0.70 \pm 0.06$	$0.68 \pm 0.10$	$0.73 \pm 0.07$	$0.11 \pm 0.27$	$0.20 \pm 0.21$	0.74
nnU-Net 3D	<b><math>0.72 \pm 0.05</math></b>	<b><math>0.72 \pm 0.10</math></b>	$0.75 \pm 0.06$	<b><math>0.09 \pm 0.26</math></b>	<b><math>0.20 \pm 0.19</math></b>	0.79
3D U-Net 8	$0.72 \pm 0.06$	$0.68 \pm 0.10$	$0.78 \pm 0.06$	$0.19 \pm 0.31$	$0.23 \pm 0.28$	<b>0.80</b>
Axial-MLP 8 (proposed)	$0.71 \pm 0.06$	$0.65 \pm 0.10$	<b><math>0.79 \pm 0.06</math></b>	$0.25 \pm 0.32$	$0.29 \pm 0.29$	0.76

Table 4. Different characteristics of 3D U-Net and Axial-MLP when varying the number of initial filters/hidden channels. The Dice coefficient is from the CV. The reported times represent the duration of a training step with a batch size of 4 in seconds. The GPU memory consumption, given in MB, was obtained with the `nvidia-smi` command and a batch size of 4. We also tried to train a 32 channel Axial-MLP but ran out of GPU memory on a 32GB Nvidia V100.

	Dice	Params	Time	Memory
U-Net 4	$0.711 \pm 0.060$	256,813	0.179	3,930
U-Net 8	$0.719 \pm 0.060$	1,021,505	0.282	6,418
U-Net 16	$0.722 \pm 0.061$	4,082,049	0.596	11,026
U-Net 32	$0.723 \pm 0.058$	16,320,257	1.682	20,668
Axial-MLP 4	$0.704 \pm 0.062$	53,017	0.114	5,026
Axial-MLP 8	$0.711 \pm 0.060$	209,317	0.229	8,614
Axial-MLP 16	$0.713 \pm 0.058$	831,805	0.474	15,654



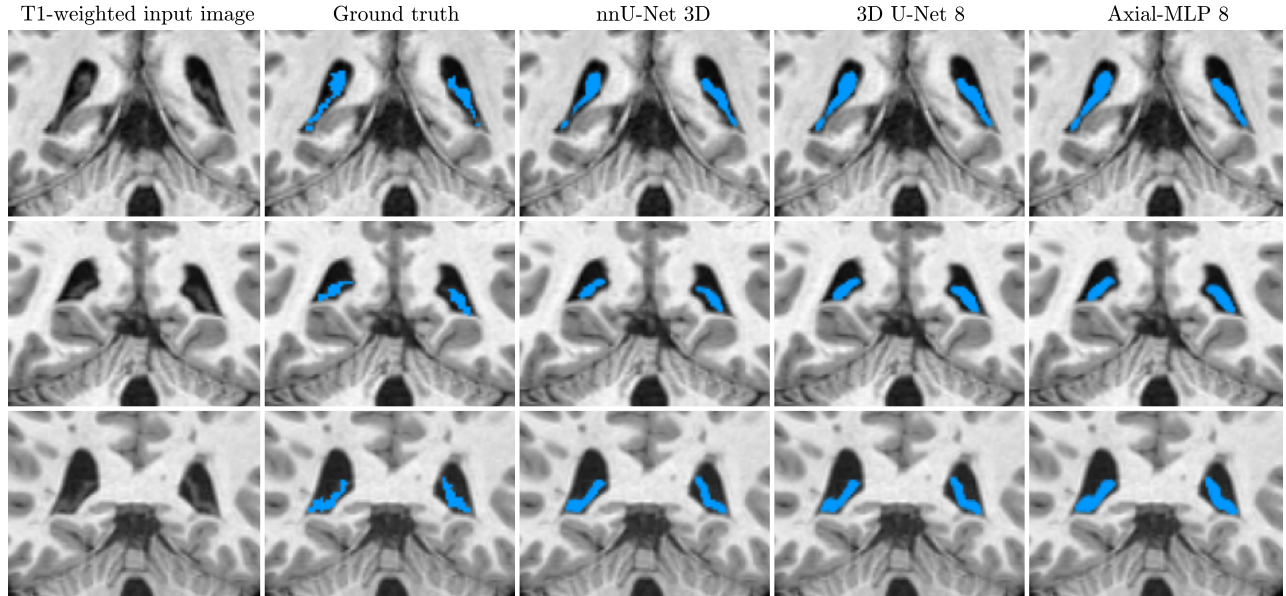


Figure 2. Examples of nnU-Net, 3D U-Net and Axial-MLP segmentation in coronal view for three patients.

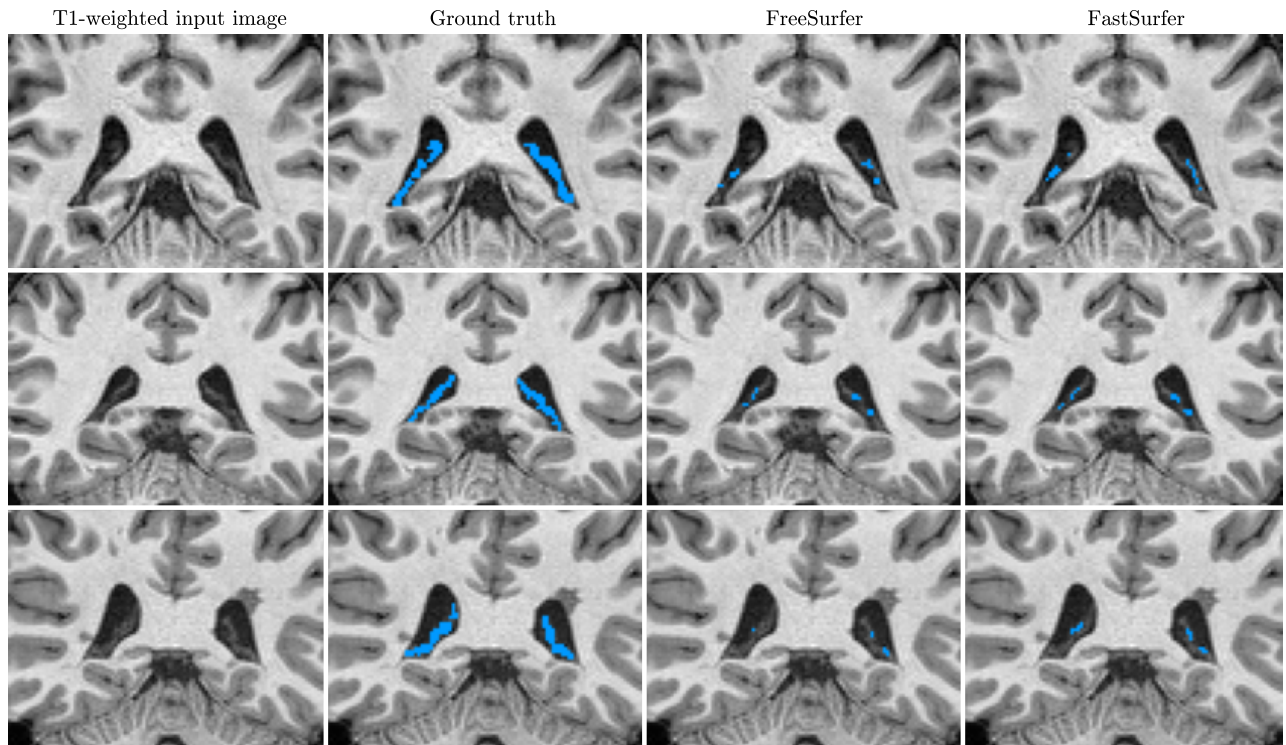


Figure 3. Examples of FreeSurfer and FastSurfer segmentation in coronal view for three patients.

#### 4. DISCUSSION AND CONCLUSION

In this paper, we proposed a new architecture called "Axial-MLP" for segmentation of choroid plexuses on MRI. It was competitive with state-of-the-art segmentation methods and achieved results which were comparable to those of 3D U-Nets and nnU-Net. Its performance, as assessed by the Dice coefficient, was nevertheless slightly lower (0.71 vs 0.72). Overall, our work demonstrates that MLP models can be competitive with convolutional

neural networks for medical image segmentation. We believe that this is useful information for the community and that future improved MLP-based architectures have the potential to outperform convolutional models.

To the best of our knowledge, the only freely available tool for automatic CP segmentation is FreeSurfer. Consistently with previous studies, we showed that it is inaccurate for this specific brain structure.<sup>12,13</sup> As expected, FastSurfer required much less computational time than FreeSurfer. It was also slightly more accurate. All deep learning models vastly outperformed FreeSurfer. Interestingly, in our specific application, nnU-Net did not substantially improve the main performance metrics compared to a standard 3D U-Net even though it required considerably more computational time.

The trained deep learning models resulted in good overlap and volume agreement with the manual reference. They could potentially be useful tools to study choroid plexuses in large cohorts of patients with multiple sclerosis in order to shed further light on their role in disease evolution. Nevertheless, this will require further validation on larger datasets with more variable image acquisition settings.

## ACKNOWLEDGMENTS

The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the “Investissements d’avenir” program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6). This work was granted access to the HPC resources of IDRIS (Jean Zay supercomputer) under the allocation “Accès dynamique” 102000 (number: AD011012863) made by GENCI (Grand Équipement National de Calcul Intensif). The authors are grateful to Mauricio Diaz-Melo for technical assistance to access the Jean Zay supercomputer.

## REFERENCES

- [1] Serot, J.-M., Béné, M.-C., and Faure, G. C., “Choroid plexus, aging of the brain, and alzheimer’s disease,” *Front Biosci* **8**(suppl), s515–s521 (2003).
- [2] Tadayon, E., Pascual-Leone, A., Press, D., Santarnecchi, E., Initiative, A. D. N., et al., “Choroid plexus volume is associated with levels of csf proteins: relevance for alzheimer’s and parkinson’s disease,” *Neurobiology of aging* **89**, 108–117 (2020).
- [3] Egorova, N., Gottlieb, E., Khelif, M. S., Spratt, N. J., and Brodtmann, A., “Choroid plexus volume after stroke,” *International Journal of Stroke* **14**(9), 923–930 (2019).
- [4] Zhou, G., Hotta, J., Lehtinen, M. K., Forss, N., and Hari, R., “Enlargement of choroid plexus in complex regional pain syndrome,” *Scientific reports* **5**(1), 1–5 (2015).
- [5] Mayinger, M., Reibelt, A., Borm, K. J., Ettl, J., Wilkens, J. J., Combs, S. E., Oechsner, M., and Duma, M. N., “MRI based neuroanatomical segmentation in breast cancer patients: leptomeningeal carcinomatosis vs. oligometastatic brain disease vs. multimetastatic brain disease,” *Radiation Oncology* **14**(1), 1–7 (2019).
- [6] Rodríguez-Lorenzo, S., Konings, J., Van Der Pol, S., Kamermans, A., Amor, S., Van Horssen, J., Witte, M. E., Kooij, G., and De Vries, H. E., “Inflammation of the choroid plexus in progressive multiple sclerosis: accumulation of granulocytes and t cells,” *Acta neuropathologica communications* **8**(1), 1–13 (2020).
- [7] Vercellino, M., Votta, B., Condello, C., Piacentino, C., Romagnolo, A., Merola, A., Capello, E., Mancardi, G. L., Mutani, R., Giordana, M. T., et al., “Involvement of the choroid plexus in multiple sclerosis autoimmune inflammation: a neuropathological study,” *Journal of neuroimmunology* **199**(1-2), 133–141 (2008).
- [8] Kooij, G., Kopplin, K., Blasig, R., Stuiver, M., Koning, N., Goverse, G., van der Pol, S. M., van Het Hof, B., Gollasch, M., Drexhage, J. A., et al., “Disturbed function of the blood–cerebrospinal fluid barrier aggravates neuro-inflammation,” *Acta neuropathologica* **128**(2), 267–277 (2014).
- [9] Kim, H., Lim, Y.-M., Kim, G., Lee, E.-J., Lee, J. H., Kim, H. W., and Kim, K.-K., “Choroid plexus changes on magnetic resonance imaging in multiple sclerosis and neuromyelitis optica spectrum disorder,” *Journal of the Neurological Sciences* **415**, 116904 (2020).
- [10] Ricigliano, V. A., Morena, E., Colombi, A., Tonietto, M., Hamzaoui, M., Poirion, E., Bottlaender, M., Gervais, P., Louapre, C., Bodini, B., et al., “Choroid plexus enlargement in inflammatory multiple sclerosis: 3.0-T MRI and translocator protein PET evaluation,” *Radiology*, 204426 (2021).



- [11] Fischl, B., “Freesurfer,” *Neuroimage* **62**(2), 774–781 (2012).
- [12] Tadayon, E., Moret, B., Sprugnoli, G., Monti, L., Pascual-Leone, A., Santarnecchi, E., Initiative, A. D. N., et al., “Improving choroid plexus segmentation in the healthy and diseased brain: Relevance for Tau-PET imaging in dementia,” *Journal of Alzheimer’s Disease* **74**(4), 1057–1068 (2020).
- [13] Zhao, L., Feng, X., Meyer, C. H., and Alsop, D. C., “Choroid plexus segmentation using optimized 3d u-net,” in *[2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)]*, 381–384, IEEE (2020).
- [14] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O., “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *[International conference on medical image computing and computer-assisted intervention]*, 424–432, Springer (2016).
- [15] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in *[International Conference on Medical image computing and computer-assisted intervention]*, 234–241, Springer (2015).
- [16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., “Attention is all you need,” in *[Advances in neural information processing systems]*, 5998–6008 (2017).
- [17] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929* (2020).
- [18] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B., “Swin transformer: Hierarchical vision transformer using shifted windows,” *arXiv preprint arXiv:2103.14030* (2021).
- [19] Hatamizadeh, A., Yang, D., Roth, H., and Xu, D., “Unetr: Transformers for 3d medical image segmentation,” *arXiv preprint arXiv:2103.10504* (2021).
- [20] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y., “Transunet: Transformers make strong encoders for medical image segmentation,” *arXiv preprint arXiv:2102.04306* (2021).
- [21] Xie, Y., Zhang, J., Shen, C., and Xia, Y., “Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation,” *arXiv preprint arXiv:2103.03024* (2021).
- [22] Melas-Kyriazi, L., “Do you even need attention? a stack of feed-forward layers does surprisingly well on imagenet,” *arXiv preprint arXiv:2105.02723* (2021).
- [23] Touvron, H., Bojanowski, P., Caron, M., Cord, M., El-Nouby, A., Grave, E., Joulin, A., Synnaeve, G., Verbeek, J., and Jégou, H., “Resmlp: Feedforward networks for image classification with data-efficient training,” *arXiv preprint arXiv:2105.03404* (2021).
- [24] Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Keysers, D., Uszkoreit, J., Lucic, M., et al., “Mlp-mixer: An all-mlp architecture for vision,” *arXiv preprint arXiv:2105.01601* (2021).
- [25] Ho, J., Kalchbrenner, N., Weissenborn, D., and Salimans, T., “Axial attention in multidimensional transformers,” *arXiv preprint arXiv:1912.12180* (2019).
- [26] Ba, J. L., Kiros, J. R., and Hinton, G. E., “Layer normalization,” *arXiv preprint arXiv:1607.06450* (2016).
- [27] Tompson, J., Goroshin, R., Jain, A., LeCun, Y., and Bregler, C., “Efficient object localization using convolutional networks,” in *[Proceedings of the IEEE conference on computer vision and pattern recognition]*, 648–656 (2015).
- [28] Kingma, D. P. and Ba, J., “Adam: A method for stochastic optimization,” *ICLR* (2015).
- [29] Pérez-García, F., Sparks, R., and Ourselin, S., “Torchio: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning,” *Computer Methods and Programs in Biomedicine*, 106236 (2021).
- [30] Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., and Reuter, M., “Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline,” *NeuroImage* **219**, 117012 (2020).
- [31] Ulyanov, D., Vedaldi, A., and Lempitsky, V., “Instance normalization: The missing ingredient for fast stylization,” *arXiv preprint arXiv:1607.08022* (2016).
- [32] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al., “nnu-net: Self-adapting framework for u-net-based medical image segmentation,” *arXiv preprint arXiv:1809.10486* (2018).

- [33] Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., et al., “Clinica: an open source software platform for reproducible clinical neuroscience studies,” *Frontiers in Neuroinformatics* (2021).
- [34] Wen, J., Thibeu-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., et al., “Convolutional neural networks for classification of alzheimer’s disease: Overview and reproducible evaluation,” *Medical image analysis* **63**, 101694 (2020).
- [35] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C., “N4itk: improved n3 bias correction,” *IEEE transactions on medical imaging* **29**(6), 1310–1320 (2010).
- [36] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C., “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain,” *Medical image analysis* **12**(1), 26–41 (2008).
- [37] Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C., “The insight toolkit image registration framework,” *Frontiers in neuroinformatics* **8**, 44 (2014).
- [38] Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C., and Collins, D., “Unbiased nonlinear average age-appropriate brain templates from birth to adulthood,” *NeuroImage* (47), S102 (2009).
- [39] Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., Group, B. D. C., et al., “Unbiased average age-appropriate atlases for pediatric studies,” *Neuroimage* **54**(1), 313–327 (2011).