Computers in Biology and Medicine 182 (2024) 109164

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

ASCHOPLEX: A generalizable approach for the automatic segmentation of choroid plexus

Valentina Visani [a], Mattia Veronese [a,b], Francesca B. Pizzini [c], Annalisa Colombi [d], Valerio Natale [e], Corina Marjin [f], Agnese Tamanti [f], Julia J. Schubert [b], Noha Althubaity [b,g], Inés Bedmar-Gómez [b], Neil A. Harrison [h], Edward T. Bullmore [i,j,k], Federico E. Turkheimer [b], NIMA Consortium [1], Massimiliano Calabrese [f], Marco Castellaro [a,*]

[a] Department of Information Engineering, University of Padova, Padova, Italy
[b] Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[c] Department of Engineering for Innovation Medicine, University of Verona, Verona, Italy
[d] Unit of Neurology, Fondazione Poliambulanza, Brescia, Italy
[e] Department of Diagnostic and Public Health, University of Verona, Verona, Italy
[f] Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy
[g] Department of Radiological Sciences, College of Applied Medical Science, King Saud bin Abdulaziz University for Health Sciences, King Abdullah International Medical Research Center (KAIMRC), Riyadh, Saudi Arabia
[h] Cardiff University Brain Research Imaging Centre (CUBRIC), Cardiff University, Cardiff, UK
[i] Department of Psychiatry, School of Clinical Medicine, University of Cambridge, Cambridge, UK
[j] Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK
[k] Immuno-Psychiatry, Immuno-Inflammation Therapeutic Area Unit, GlaxoSmithKline R&D, Stevenage, UK

ARTICLE INFO

Keywords:
Choroid plexus
Deep neural networks
Ensemble
Semantic segmentation
Magnetic resonance imaging

ABSTRACT

Background: The Choroid Plexus (ChP) plays a vital role in brain homeostasis, serving as part of the Blood-Cerebrospinal Fluid Barrier, contributing to brain clearance pathways and being the main source of cerebrospinal fluid. Since the involvement of ChP in neurological and psychiatric disorders is not entirely established and currently under investigation, accurate and reproducible segmentation of this brain structure on large cohorts remains challenging. This paper presents ASCHOPLEX, a deep-learning tool for the automated segmentation of human ChP from structural MRI data that integrates existing software architectures like 3D UNet, UNETR, and DynUNet to deliver accurate ChP volume estimates.

Methods: Here we trained ASCHOPLEX on 128 T1-w MRI images comprising both controls and patients with Multiple Sclerosis. ASCHOPLEX's performances were evaluated using traditional segmentation metrics; manual segmentation by experts served as ground truth. To overcome the generalizability problem that affects data-driven approaches, an additional fine-tuning procedure ($ASCHOPLEX_{tune}$) was implemented on 77 T1-w PET/MRI images of both controls and depressed patients.

Results: ASCHOPLEX showed superior performance compared to commonly used methods like FreeSurfer and Gaussian Mixture Model both in terms of Dice Coefficient (ASCHOPLEX 0.80, $ASCHOPLEX_{tune}$ 0.78) and estimated ChP volume error (ASCHOPLEX 9.22%, $ASCHOPLEX_{tune}$ 9.23%).

Conclusion: These results highlight the high accuracy, reliability, and reproducibility of ASCHOPLEX ChP segmentations.

* Correspondence to: Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131, Padova, Italy.
E-mail addresses: valentina.visani@phd.unipd.it (V. Visani), mattia.veronese@unipd.it (M. Veronese), francescabenedetta.pizzini@univr.it (F.B. Pizzini), annalisa.colombi@poliambulanza.it (A. Colombi), valerio.natale01@gmail.com (V. Natale), corinamarjin@gmail.com (C. Marjin), agnese.tamanti@univr.it (A. Tamanti), julia.schubert@kcl.ac.uk (J.J. Schubert), thubaityn@ksau-hs.edu.sa (N. Althubaity), ines.bg.med@gmail.com (I. Bedmar-Gómez), HarrisonN4@cardiff.ac.uk (N.A. Harrison), etb23@medschl.cam.ac.uk (E.T. Bullmore), federico.turkheimer@kcl.ac.uk (F.E. Turkheimer), massimiliano.calabrese@univr.it (M. Calabrese), marco.castellaro@unipd.it (M. Castellaro).
[1] See Appendix.

https://doi.org/10.1016/j.compbiomed.2024.109164
Received 9 April 2024; Received in revised form 23 August 2024; Accepted 16 September 2024
Available online 25 September 2024

1. Introduction

1.1. The choroid plexus

The Choroid Plexus (ChP) is a brain anatomical structure situated within the cerebral ventricular system across all four ventricles, which constitutes a significant component of the Blood - Cerebrospinal Fluid - Barrier (BCSFB). ChP primary function involves the production of the CSF [1], but it also serves as a mediator for brain clearance pathways, thereby contributing to the maintenance of brain homeostasis [2,3] and glymphatic function [4]. The functional role of the ChP has been primarily characterized in animals [5] but recent literature in clinical cohorts has evidenced its role in the mediation of peripheral and central inflammation [6]; importantly, increased ChP volume has been validated as a robust marker of neuroinflammation across diagnostic neurological and psychiatric cohorts [7].

1.2. How to visualize the choroid plexus

Due to its clinical relevance, several quantitative neuroimaging modalities, including Diffusion Weighted Imaging [8], perfusion imaging [9], and Positron Emission Tomography [10,11], have been used to investigate both its function and morphology. ChP investigations require precise segmentation from structural data and are typically performed on T1-weighted (T1-w) MRI. Note that the gold standard MRI sequence to image the ChP is the T1-w MRI sequence enhanced with contrast injection [12,13]. However, due to its invasive nature [14] and limited routine acquisition, T1-w without contrast agent injection are instead used [15] given that estimates of ChP volume (ChPV) using non contrast-enhanced T1-w images compared to contrast-enhanced ones are fairly correlated [12,16].

1.3. The choroid plexus role in clinical studies

In Alzheimer's disease an increase in ChPV is related to worse cognitive impairment, as well as reduced clearance of CSF proteins like amyloid-beta that leads to protein accumulation in tissue [2,12, 17,18]. Concerning Multiple Sclerosis (MS) an alteration of ChPV is related to neuroinflammation processes [5,7,19–21]. Particularly, a recent study [22] has also demonstrated a potential correlation between the ChP enlargement and both increased CSF albumin concentration and the clinical disability grade of MS patients [23]. In psychiatric cohorts, an enlarged ChP is associated with lower cognition in psychosis and reduction of BCSFB permeability in depression [10,24,25]. Therefore, the robustness, generalizability and reliability of ChP segmentation are important to guarantee the validity and reproducibility of ChP biomarkers in these cohorts [26].

1.4. The need for a choroid plexus automatic segmentation

Segmentation of ChP on T1-weighted MRI is routinely done manually but the process is time-consuming and prone to inter- and intra-operator variability [12,27]. These variabilities are further exacerbated by the high heterogeneity of ChP in terms of size and tortuosity, as well as by the limited contrast between the ChP and the ventricles [14,27]. Whereas manual segmentation is still the gold standard and considered the ground truth (GT MSeg), these weaknesses make the study of ChP on large amounts of data problematic. To enable studies on large patient cohorts, few automatic tools for ChP segmentation have been proposed in the literature, although with sub-optimal results. FreeSurfer (FS) [28], an open-source software routinely used for the automatic segmentation of brain images, was the first to be tested, but it poorly correlates with the GT MSeg [12,29]. The Gaussian Mixture Model (GMM), an add-on method applied by Tadayon et al. [12], has limited applicability to diseased subjects with morphological abnormalities [29]. Recently, the single 3D U-Net method [30] has proven to return better spatial overlapping than FS, but its validation was limited, since the study was carried out only on ten subjects. A recent study exploited the use of 2-step 3D U-Net emphasizing the potential of this simple architecture [14]. Choroid Plexus Segmentation toolbox (ChPSeg) [6] is also based on 3D U-Net, but input undergo preprocessing steps, including registration to standard space and cropping. However, the benefits of preprocessing in medical imaging deep learning-based segmentation are still under investigation [31] and require supplementary execution time rather than prediction with a trained network on unprocessed data. The GMM method has been recently improved to incorporate other conventional MRI sequences like FLAIR, but the improvement lacks robustness [29]. Recently, FastSurfer (FastS) [32], a faster deep learning-based version of FS, can segment the ChP, and showed satisfactory results in an Alzheimer's cohorts [33], but it has been proven to fail on MS patients [27]. As of today, a reliable and generalizable automatic segmentation of the ChP for population studies involving large cohorts of both patients and controls is still unavailable, though highly desiderable.

1.5. The aim of this work

The aim of this study is to propose ASCHOPLEX (Automatic Segmentation of CHOroid PLEXus), an automatic Deep Learning tool to obtain a reproducible and reliable ChP segmentation on T1-w MRI data minimizing the error in ChPV estimation. Differently from other applications of deep learning for ChP segmentation, ASCHOPLEX increases the robustness of the ChP segmentation through the ensemble of the predictions made by the most performant trained models and overcomes the generalizability problem that affects data-driven approaches by including a fine-tuning step to learn dataset-specific features. ASCHOPLEX implementation relies on the MONAI Auto3dSeg pipeline [34] to include the architecture template of 3D U-Net [35], UNETR (UNEt TRansformers) [36] and DynUNet, a variant of the nnU-Net [37], with different path-size and loss functions. 3D U-Net is widely recognized as the workhorse in medical imaging segmentation tasks; however, we also included UNETR and nnU-Net which outperformed the simpler U-Net architecture in recent Medical Imaging Segmentation challenges [36,37]. ASCHOPLEX was tested on three independent datasets composed of controls and patients with different brain disorders (MS and Depression) acquired from three different scanners. All data were manually segmented by experts. Two datasets were used to train the core networks, while the third dataset was used to test the generalizability of the tool implementing a fine-tuning procedure. The proposed approach has been compared to the commonly used literature approaches, FS, FastS and GMM, and the recently published tool ChPSeg.

2. Materials and methods

2.1. Datasets

Data were obtained from three studies for a total of 205 subjects.[2] Datasets 1 and 2 were provided by the Multiple Sclerosis Centre of the University Hospital of Verona and were acquired prospectively between March 2019 and October 2021. All subjects gave their written informed consent prior to participating in the study. All procedures were performed in accordance with the Declaration of Helsinki (2008) and the study protocols were approved by the local Ethical Committee. Below is a detailed description of the two datasets.

[2] In the *Supplementary Materials - Supplementary Figures (Figure TI)* is shown a visual comparison of the three datasets.

1. Dataset 1: It comprises 67 subjects (M/F 21/46): 24 healthy controls (HC) (age: 37.2 ± 9.5 years) and 43 Relapsing-Remitting Multiple Sclerosis (RRMS) (age: 40.9 ± 9.9 years) patients. T1-w MRI images were acquired on a Philips Achieva TX with 8-channels head coil (Software version R3.2.3.2). Parameters of 3D T1-w MPRAGE sequences were: resolution 1 × 1 × 1 mm; SENSE acceleration factor: 2.5; TE/TR/TI: 3.7/8.4/1037 ms; FA: 9°; total acquisition time: 4 min 50 s.
2. Dataset 2: It comprises 61 RRMS patients (M/F 13/48, age: 36.7 ± 10.1 years). T1-w MRI images were acquired on a Philips Elition S with 32-channels head coil (Software version R5.7.2.1). Parameters of 3D T1-w MPRAGE sequences were: resolution: 1 × 1 × 1 mm; compressed SENSE acceleration factor: 4; TE/TR/TI: 3.7/8.4/1016 ms; FA: 8°; total acquisition time: 3 min 20 s.

Dataset 3 was collected in the United Kingdom as part of the Biomarkers in Depression Study (BIODEP, NIMA consortium, https://www.neuroimmunology.org.uk/). All procedures were approved by the local ethics committee and performed in accordance with the Declaration of Helsinki. Participants gave their written informed consent prior to participating. The BIODEP study was approved by the NRES Committee East of England Cambridge Central (REC reference:15/EE/0092). In brief, Dataset 3 comprises 77 subjects (M/F 26/51): 51 depressed participants (age: 36.2 ± 7.3 years) and 26 matched HC (age: 37.3 ± 7.8 years) (Full details are reported in the study protocol [10,38]). Each participant underwent a structural MRI simultaneously to PET imaging with a GE SIGNA PET/MR scanner. Parameters of 3D T1-w Fast SPGR sequences were: resolution: 1 × 1 × 1 mm; TE/TR/TI: 2.99/6.96/450 ms; FA: 12°; total acquisition time: 6 min.

2.2. Pre-processing

Images of Datasets 1 and 2 did not undergo any preprocessing before being passed as input to ASCHOPLEX. Images of Dataset 3 underwent a rigid-body transformation to the MNI Talairach ICBM 152 2006c template provided via MIAKAT v4.2.6 software (https://www.miakat.org/MIAKAT2/index.htm). To obtain the FS segmentation, the T1-w sequence of Datasets 1 and 2 have undergone a lesion filling procedure based on Lesion Prediction Algorithm available in the Lesion Segmentation Toolbox for SPM12 (www.statistical-modelling.de/lst.html) [39].

2.3. Choroid plexus manual segmentation

The GT MSeg of each subject was obtained in T1-w native space. ChP segmentation was performed only in the two lateral ventricles because the ChP is often challenging to reliably distinguish in the third and fourth ventricles using conventional 3T MRI imaging [15]. For Datasets 1 and 2 the ChP was segmented by consensus of a junior and a senior neuroradiologist for each dataset using ITK-snap [40]. The same was done for Dataset 3 using Analyze software v.12 (https://analyzedirect.com). The GT was used as reference to compare the proposed method with FS, GMM, FastS, and ChPSeg. Moreover, it was also employed during the training of the deep neural network (DNN) models and as reference to test their performances in the validation/testing phase.

2.4. ASCHOPLEX implementation

ASCHOPLEX (Automatic Segmentation of CHOroid PLEXus) has been implemented (Fig. 1) in MONAI version 1.0.1 [34]. MONAI (Medical Open Network for Artificial Intelligence) is a collection of open-source, freely available collaborative frameworks. Particularly, MONAI has a streaming loading modality to improve the training efficiency, many DNNs already implemented, as domain specific metrics and loss functions. Specifically, ASCHOPLEX is a modified version of Auto3DSeg (https://github.com/Project-MONAI/tutorials/tree/main/auto3dseg), a MONAI application for 3D medical data that exploits cutting-edge DNN architectures with an ensemble strategy to propose a flexible tool to perform image segmentation in the medical imaging domain. Auto3DSeg is agnostic to the image domain of application and exploits structured format that generates the algorithm folders for each selected DNN based on the MONAI bundle. A bundle is a self-descriptive network that combines its architecture definition with the metadata, like the global information on the data, and the scripts to train and infer the model network. The main advantages of Auto3DSeg are easier distribution of the networks, user friendly approach (the user only needs to provide the data folder and its datalist), and reproducibility of the results. A further feature of Auto3DSeg is the ensemble module that ranks trained algorithms based on checkpoints validation accuracy, selects the top N models, and generates predictions through ensemble methods, implementing both average and major voting. The output segmentations derived from various DNN models might exhibit significant variability. Therefore, an effective solution to this challenge is the implementation of an ensemble procedure, which consolidates and integrates all individual predictions.

Table 1
Overview of study datasets division into training/validation/testing sets.

Dataset	Dataset	Subjects Division
ASCHOPLEX Core Networks: Training and Ensemble		
Dataset 1	24 HC 43 RRMS	Training: 29 RRMS, 17 HC Testing: 14 RRMS, 7 HC
Dataset 2	61 RRMS	Training: 46 RRMS Testing: 15 RRMS
Total	**128 subj**	Training: 92 subj Testing: 36 subj
Study of the generalizability of ASCHOPLEX		
Dataset 3	**77 subj** blinded division of depressed and HC	Fine-tuning range: **from 0 to 10 subj** Validation: 5 subj Testing: 62 subj

Subj=subject; HC=healthy controls; RRMS=Relapsing-Remitting Multiple Sclerosis.

2.4.1. ASCHOPLEX core networks: training and ensemble

Datasets 1 and 2 were used in the initial stage to train the core networks. We will refer to the merge of Datasets 1 and 2 as a single dataset renamed "core networks training" dataset. In the training/validation/testing splitting process we balanced the phenotype (RRMS/HC) and the scanner (Table 1). The training procedure was implemented following a five-fold cross-validation. The testing set was left untouched until the end of the study to compare the performance of ASCHOPLEX with the commonly used literature approaches. We started comparing DNNs of varying architectures and hyperparameters. We selected three DNNs architectures already implemented in MONAI: 3D U-Net [35]; DynUnet, MONAI implementation of nnU-Net [37], and UNETR [36]. All the configurations were implemented with 3D patch inputs. The 3D U-Net was configured with a single input channel. It consisted of 5 layers, with a number of filters equal to 16 for the first layer and doubled for the following ones. Data were downsampled/upsampled in the encoder/decoder part using 2-strided convolutions residual units with a down/up-sampling kernel size of 3 × 3 × 3 voxels, followed by instance normalization and PReLU activation blocks. The DynUnet is part of a new generation of self-configuring DNNs. It accepted 3D images as input with a single input channel. It was configured with four blocks: input, downsample, bottleneck, upsample. Data were down/up-sampled using [1, 2, 2, 1] strides convolutions with a kernel size of 3 × 3 × 3 voxels, followed by instance normalization and LeakyReLU activation. UNETR is an adaptation of vision transformers. It is composed by a vision transformer in the encoder part connected to the decoder part via skip

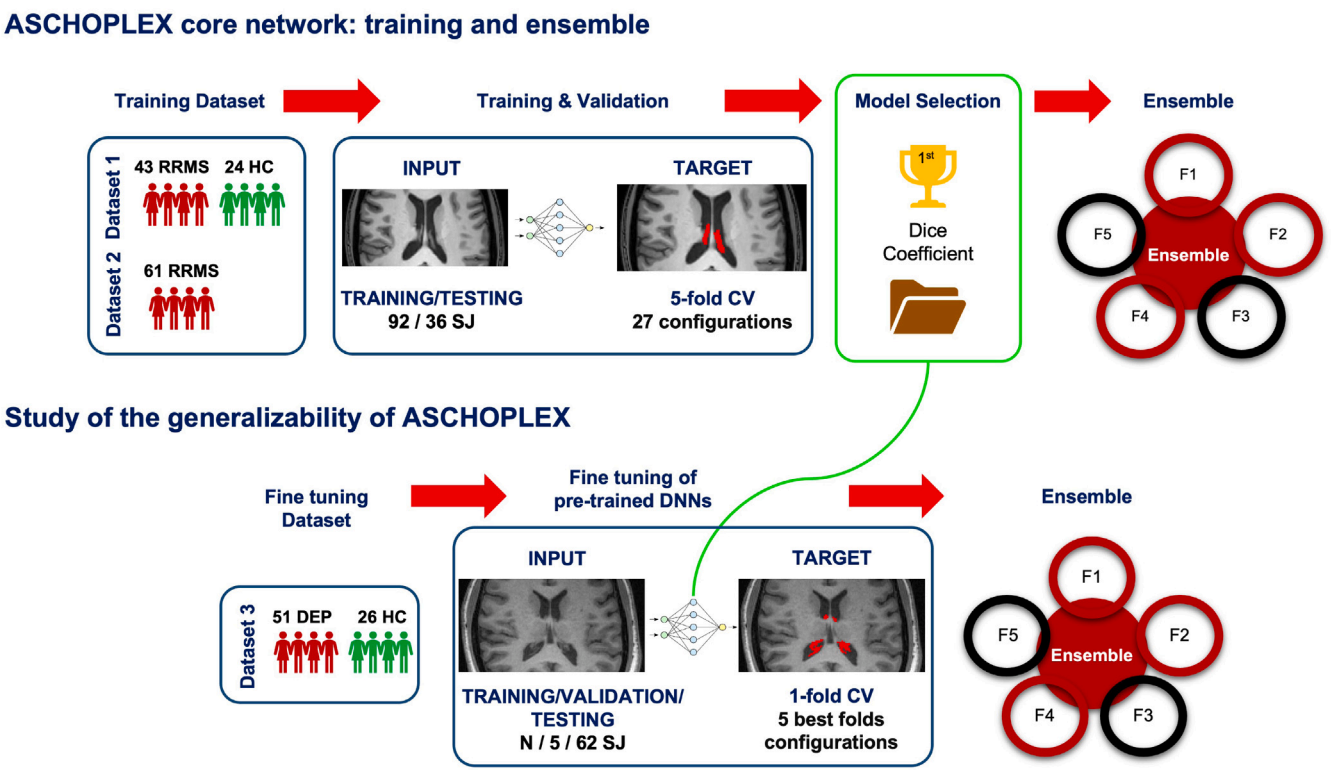


Fig. 1. Workflow of the implementation of ASCHOPLEX. ASCHOPLEX core networks: training and ensemble (1) T1-w MRI data derived from two different MRI scanners were manually segmented and used as input (RRMS: Relapsing-Remitting Multiple Sclerosis; HC: Healthy Controls) (2) Training and validation of twenty-seven DNNs configurations using five-fold cross-validation (3) Selection of the best model for each fold (Fx) based on Dice Coefficient (4) Ensemble by major voting of the five selected model predictions to obtain the final outcome of ASCHOPLEX. Study of the generalizability of ASCHOPLEX: (1) Data from a hybrid PET/MRI scanner were manually segmented and used as input (DEP: depressed subjects) (2) Fine-tuning of the pre-trained five models on N subjects (3) Ensemble by major voting of the five fine-tuned models predictions to obtain the final segmentation.

connections to merge the major benefit of both transformers and U-Net, which inspired this architecture: respectively, great capability in learning global information and properly capturing localized features. The configuration accepted single 3D inputs. The dimension of the network feature size was set to 16, the hidden layer size to 768, and the feedforward layer size to 3072. The transformer in the encoder branch used 12 attention heads. To preserve spatial information of the patches, a perceptron positional embedding layer was used. At every level of resolution, the reshaped tensors from the embedding space were projected into the input space through a series of successive $3 \times 3 \times 3$ convolutional layers, which were subsequently followed by normalization layers. As hyperparameters we considered: three patch size ($64 \times 64 \times 64$, $96 \times 96 \times 96$, $128 \times 128 \times 128$) and three loss functions (Generalized Dice (Dice), CrossEntropy (CE), combination of Dice and CE (DiceCE)), for a total of twenty-seven DNNs. MONAI Auto3DSeg bundle was extended to include each DNN among the twenty-seven selected configurations. We trained each configuration separately in a five-fold cross-validation fashion using data augmentation transforms (rotations and flip among all three axes, intensity shifts with probability=15%) to extend the dataset features to ensure the model will be less prone to overfitting. Each DNN was trained using the Adam-Weighted optimizer [41] with parameters: learning rate (1e-04), weight decay (1e-05), maximum number of iterations (2e04), batch size (1). The training of the configurations was executed using a 16 GB NVIDIA Tesla V100 GPU. The overall time required for this initial training step was 20 days (483 h). Following the cross-validation scheme, we use the Dice Coefficient to select the best configuration among the twenty-seven configurations trained for each fold, obtaining the five best models to perform the ensemble by majority voting to increase the robustness of the predicted segmentation on the testing set.

2.4.2. Study of the generalizability of ASCHOPLEX

The principal limitation of data-driven methodologies lies in the limited capacity for model generalization when applied to datasets distinct from those employed during the training phase with potentially different patterns of acquisition due to either the scanner (e.g. MR scanner hardware, MR sequence parameters or MR software version) or to the brain morphology of the subjects involved. Changing the MRI dataset, even if the MR sequence contrast is the same, means showing the DNNs images with untrained features. This is a generalizability problem that especially affects deep learning when it deals with MR images [42]. This issue can be in some cases overcome with a fine-tuning procedure (Fig. 1). Dataset 3, denoted as the "fine-tuning" dataset, was employed to assess the tool's generalizability. The dataset was partitioned into distinct subsets for training, validation, and testing, as outlined in Table 1. The validation set, composed by five subjects, served the purpose of model validation throughout the training process, ensuring its robustness. The testing set remained unaltered until the completion of the study, allowing for the evaluation of ASCHOPLEX's performance on a novel dataset. This evaluation also facilitated a comparative analysis with conventional approaches found in the existing literature. Firstly, to test the generalizability of the tool, ASCHOPLEX was inferred on the 62 testing subjects of the fine-tuning dataset with a no fine-tuned network ensemble. This approach was named ASCHOPLEX_0. Secondly, a fine-tuning step is carried out. This approach, called $\text{ASCHOPLEX}_{\text{tune}}$, aimed to gauge the tool's performance across a spectrum of training scenarios, providing insights into its robustness and versatility. To better understand the minimum number of subjects that might be needed when adapting ASCHOPLEX to a new unseen dataset, we started from the five selected core networks as pretrained models and we varied the number of subjects in the fine-tuning training procedure from one to ten. The

Table 2
Results of the segmentation metrics on 36 testing subjects of the core network training dataset. Comparison of the ensemble after the training of ASCHOPLEX, ChPSeg, FreeSurfer (FS), FastSurfer (FastS) and Gaussian mixture Model (GMM) methods with the reference manual segmentation (GT MSeg).

Metrics	ASCHOPLEX	ChPSeg	FS	FastS	GMM	GT MSeg
Volume	**2432 ± 668**	2062 ± 567	1319 ± 469	1464 ± 430	1928 ± 580	2480 ± 837
Dice	**0.80 ± 0.07**	0.60 ± 0.09	0.32 ± 0.09	0.34 ± 0.09	0.54 ± 0.09	–
95% HD	**2.50 ± 2.89**	13.40 ± 9.87	9.11 ± 2.79	9.06 ± 2.22	10.52 ± 5.99	–
$\Delta Vol\%$	**0.70 ± 11.31**	−13.35 ± 29.90	−43.05 ± 22.12	−38.60 ± 22.09	−15.08 ± 34.93	–
$\lvert\Delta Vol\%\rvert$	**9.22 ± 6.59**	26.69 ± 18.98	44.82 ± 18.26	41.11 ± 16.98	29.80 ± 23.66	–
Recall	**0.80 ± 0.08**	0.56 ± 0.15	0.25 ± 0.1	0.28 ± 0.09	0.50 ± 0.12	–
Precision	**0.80 ± 0.09**	0.68 ± 0.11	0.46 ± 0.13	0.49 ± 0.13	0.63 ± 0.14	–
r	**0.95**	0.36	0.36	0.40	n.s.	–

All values (except for Absolute Volume and Pearson's Volume Correlation Coefficient) are presented in the form mean ± standard deviation. Absolute Volume is expressed in [mm^3], while |ΔVol%| in [%]. Legend: n.s.- non significant.

number of necessary manually labeled subjects must be reasonable: if the number is too low, the model struggles to learn the characteristics of the dataset; using more data, the manual segmentation becomes time-consuming. As before, the MONAI bundle was modified to consider the new algorithm templates with the addition of the fine-tuning step. The fine-tuning procedure consists of the complete training (with parameters described as follows) of the five selected models without freezing any model or layers to match the dataset characteristics better. The training parameters were the same as in the core networks training except for the maximum number of iterations (1e04). During this phase, the five-fold cross-validation methodology was not employed; a single validation step was introduced, eliminating the need of dividing the training set into folds. Similarly to previous iterations, the ultimate prediction for each testing subject was generated by conducting an ensemble process, employing a majority voting mechanism among the segmentations derived from the five fine-tuned models. The total time taken for this step, fine-tuning and segmentation saving, was 20 h.

2.5. Performance evaluation

We conducted a comparative analysis between the segmentations generated by ASCHOPLEX and those generated by all the freely available software up to date (FS, FastS, ChPSeg and GMM). For all the tested software we used as input only the T1-w image. This comparison was performed separately on the testing sets of both the core networks and the fine-tuning dataset. We generated segmentations using the FS recon-all pipeline (version 7.1.1 for core networks training dataset, version 6.0.0 for fine-tuning dataset) [28], and applied the GMM method, following the GMM pipeline available on GitHub [12]. To obtain the FastS segmentations, we used version 2.2.0 of FastS provided in a docker image [32]. We employed the ChPSeg tool version 1.0.0, that is available in its docker version on GitHub [6]. The performance of the compared approaches was evaluated using the following metrics: Dice Coefficient $Dice = \frac{2\Sigma min(x_i,y_i)}{\Sigma x_i+\Sigma y_i}$; 95% Hausdorff Distance (*95% HD*); Pearson's Volume Correlation (r); Percentage Volume Difference $\Delta\text{Vol\%} = \frac{\Sigma_i(x_i-y_i)}{\Sigma_i y_i}\%$; Absolute Percentage Volume Difference ($|\Delta\text{Vol\%}|$); Recall $Recall = \frac{\Sigma_i x_i y_i}{\Sigma_i y_i}$; Precision $Precision = \frac{\Sigma_i x_i y_i}{\Sigma_i x_i}$, where $x_i, y_i \epsilon X, Y$ are the predicted segmentation and the GT MSeg, respectively. All metrics have been evaluated considering the entire 3D images and not considering the background voxels which, since the plexus is very small compared to the brain, could overestimate good performance. Dice Coefficient gives an estimate of the similarity of the compared segmentations, just like Recall and Precision. On the contrary, *95% HD* focused on how far the compared segmentations are: to reach good performance, the segmentation approach must maximize the Dice and minimize the *95% HD*. In addition, $|\Delta\text{Vol\%}|$, $\Delta\text{Vol\%}$, and r focus on the ChPV estimate.

Concerning ASCHOPLEX training, one-way ANOVA and post-hoc t-test ($\alpha = 0.05$) were performed for Absolute Volume between all available segmentations for the core network training dataset (ASCHOPLEX, FS, FastS, GMM, ChPSeg, and GT MSeg). Concerning ASCHOPLEX_0 and ranging the number of subjects of $\text{ASCHOPLEX}_{\text{tune}}$ from one to ten, one-way ANOVA and post-hoc t-test ($\alpha = 0.05$) were performed on the fine-tuning dataset for Dice Coefficient, *95% HD*, and Absolute Volume to identify the minimum number of training subjects for the fine-tuning step. Once the optimal number of subjects for $\text{ASCHOPLEX}_{\text{tune}}$ was identified, one-way ANOVA and post-hoc t-test ($\alpha = 0.05$) were performed for Absolute Volume calculated on the fine-tuning dataset for the tested methods: $\text{ASCHOPLEX}_{\text{tune}}$, FS, FastS, GMM, ChPSeg, and GT MSeg. A Bonferroni correction for multiple comparisons was used.

2.6. Software availability

The tool proposed in this paper is freely available, published as a user-friendly package on GitLab (https://gitlab.dei.unipd.it/fair/aschoplex). The end user can directly infer the segmentation with pre-trained models based on the core network training dataset or perform the manual segmentation for a subsample of a new dataset running the fine-tuning step, adapting the model to the new dataset features to improve the reliability of the predicted segmentations. Moreover, a sample case from Dataset 1 is available in the GitLab repository.

3. Results

3.1. ASCHOPLEX core networks: training and ensemble

The five selected configurations were characterized by UNETR and DynUnet architectures. All models were 128 patch-based with Dice or DiceCE loss function. Table 2 reports the results for the segmentation metrics on the 36 testing subjects of the core network training dataset.[3] The compared approaches were ASCHOPLEX, ChPSeg, FS, FastS, GMM and GT MSeg. Dice Coefficient is higher for the ensemble approach (0.80 ± 0.07) than other methods proposed by the literature. Concerning *95% HD*, ASCHOPLEX prediction location (2.50 ± 2.89) is the nearest to GT MSeg. Focusing on the volume analysis, Pearson's correlation analysis between each segmentation reports a positive correlation between the volume of the ASCHOPLEX and GT MSeg (0.95) that is higher than that between FS, FastS, and ChPSeg, and GT MSeg (respectively, 0.36, 0.36, 0.40), while GMM does not show any statistically significant correlation. One-way ANOVA reveals a statistically significant main effect on the absolute volume among the tested methods ($p_{value} < 0.0001$). FS, FastS, and GMM have statistically lower absolute volume when compared to GT MSeg (GMM corrected $p_{value} = 0.003$, FS and FastS corrected $p_{value} < 0.0001$), while there are no statistically significant differences between both ASCHOPLEX (corrected $p_{value} = 1.0$) and ChPSeg (corrected $p_{value} = 0.07$) and GT MSeg volumes (Fig. 2). ASCHOPLEX provides the lower $|\Delta\text{Vol\%}|$ (9.22 ± 6.59 %). In Fig. 2(a) the distributions of ASCHOPLEX and GT MSeg

[3] The correspondent boxplots are shown in the *Supplementary Materials - Supplementary Figures (Figure A)*, while a representative figure for a particular case subject is reported in *Figure C*.

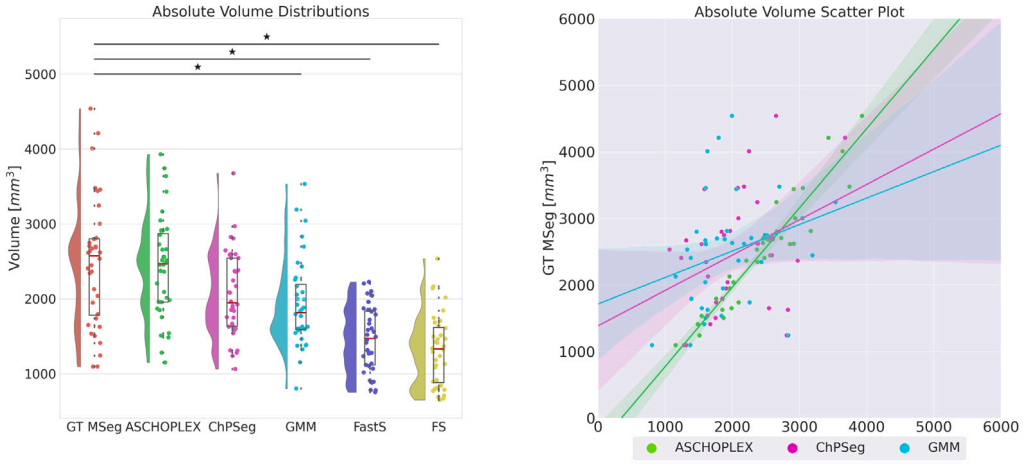


Fig. 2. (a) Rainplot of the absolute volume of the predicted segmentation obtained on the 36 testing subjects of the core network training dataset with the compared methods ASCHOPLEX, ChPSeg, FreeSurfer (FS), FastSurfer (FastS), Gaussian Mixture Models (GMM), and manual segmentation (GT MSeg), used as reference. Asterisks indicate the statistical differences between the absolute volumes of the compared methods and GT MSeg. (b) Scatter plot and linear regression model fit with 95% confidence interval bounds of the absolute volume of the predicted segmentation obtained on the 36 testing subjects of the core network training dataset with the compared methods ASCHOPLEX, ChPSeg, Gaussian Mixture Models (GMM), and manual segmentation (GT MSeg), used as reference.

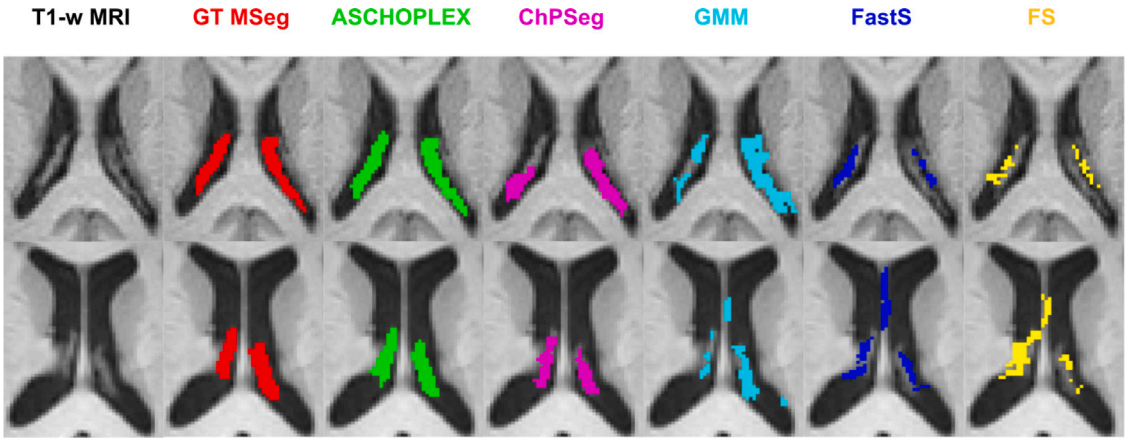


Fig. 3. Comparison of Choroid Plexus predicted segmentation. From left to right, the figure shows a zoom-in of the 3D T1-w MRI with superimposed manual segmentation (GT MSeg) in red, proposed method ASCHOPLEX in green, FastSurfer (FastS) in blue, ChPSeg in pink, Gaussian Mixture Model (GMM) in light-blue, and FreeSurfer (FS) in yellow. Percentage Volume Difference metric (reference GT MSeg) for the top subject (best case patient): ASCHOPLEX +0.32%, ChPSeg +5.92%, FS −48.79%, FastS −40.79%, GMM +19.52%; for the bottom subject (worst case patient): ASCHOPLEX −15.37%, ChPSeg −53.94%, FS −68.47%, FastS −62.80%, GMM −40.22%.

volumes are comparable. ChPSeg gives a moderate underestimation bias and FS, FastS, and GMM underestimate ChPV more than both ASCHOPLEX and ChPSeg. In Fig. 2(b) the linear regression between the ChPV predicted by ASCHOPLEX and that calculated on the GT MSeg provide a more robust result, shown by narrowed confidence bounds, while GMM and ChPSeg provide a wider confidence interval of the linear regression coefficient. Regression models of FastS and FS are not reported due to lower performance compared to other methods and they are both not optimize to segment the ChP only. Regarding the other metrics, Recall (0.80 ± 0.08) and Precision (0.80 ± 0.09) of ASCHOPLEX are higher than those of the other methods. Fig. 3 shows the automatic segmentation results overlapped to the T1-w MRI image for two representative patients with GT Mseg reported as reference.

3.2. Study of the generalizability of ASCHOPLEX

Table 3 reports the results for the segmentation metrics on the testing set of the fine-tuning dataset. The compared approaches were: ASCHOPLEX_0, $\text{ASCHOPLEX}_{\text{tune}}$, FS, FastS, GMM, ChPSeg, and GT MSeg.

Firstly, we have compared ASCHOPLEX_0 and $\text{ASCHOPLEX}_{\text{tune}}$. Fig. 4 reports the boxplot of the performance metrics for this analysis.[4] Focusing on Dice Coefficient, ANOVA revealed an effect of group between all compared segmentations in terms of Dice Coefficient ($p_{value} < 0.0001$). In fact, the fine-tuning step improves the agreement between the predicted segmentation and the GT MSeg compared to ASCHOPLEX_0. Particularly, the Dice Coefficient is not statistically different in $\text{ASCHOPLEX}_{\text{tune}}$ from four to ten subjects, while ASCHOPLEX_0 is statistically significantly different from all $\text{ASCHOPLEX}_{\text{tune}}$ training strategies considering this metric (corrected $p_{value} < 0.0001$). Concerning *95% HD*, segmentations obtained with $\text{ASCHOPLEX}_{\text{tune}}$ are closer to GT MSeg than ASCHOPLEX_0. Particularly, $\text{ASCHOPLEX}_{\text{tune}}$ approaches using a number of training subjects between three and ten are not statistically significantly different for this metric (corrected $p_{value} = 1.0$). Focusing on the volume analysis, Pearson's correlation

[4] In *Supplementary Materials - Supplementary Tables (Table A, B, C)* are reported the multiple comparison results for, Dice Coefficient, *95% HD*, and Absolute Volume, respectively.

Table 3
Results of the segmentation metrics on the 62 testing subjects of the fine-tuning dataset. Comparison of $ASCHOPLEX_0$ (inference without fine-tuning), $ASCHOPLEX_{tune}$ (fine-tuning varying the number of subjects from one to ten), FreeSurfer (FS), FastSurfer (FastS), Gaussian Mixture Model (GMM), and ChPSeg with the reference manual segmentation (GT MSeg).

	Volume [mm^3]	Dice	95% HD	$\Delta Vol\%$	$\lvert\Delta Vol\%\rvert$	Recall	Precision	r
GT MSeg	2032 ± 526	–	–	–	–	–	–	–
$ASCHOPLEX_0$	1267 ± 479	0.59 ± 0.16	4.75 ± 6.42	−33.80 ± 26.59	35.60 ± 24.12	0.51 ± 0.19	0.78 ± 0.07	n.s.
$ASCHOPLEX_{tune}$ 1 SJ	2328 ± 599	0.69 ± 0.06	4.11 ± 1.69	16.96 ± 24.40	25.13 ± 15.81	0.75 ± 0.09	0.65 ± 0.09	0.63
$ASCHOPLEX_{tune}$ 2 SJ	2509 ± 670	0.71 ± 0.07	3.88 ± 1.99	25.17 ± 24.65	28.85 ± 20.22	0.79 ± 0.06	0.65 ± 0.11	0.70
$ASCHOPLEX_{tune}$ 3 SJ	2464 ± 601	0.75 ± 0.05	2.94 ± 1.50	22.68 ± 17.44	24.30 ± 15.09	0.83 ± 0.06	0.68 ± 0.08	0.83
$ASCHOPLEX_{tune}$ 4 SJ	2375 ± 600	0.75 ± 0.05	2.54 ± 1.29	17.98 ± 16.44	20.44 ± 13.25	0.82 ± 0.06	0.71 ± 0.08	0.84
$ASCHOPLEX_{tune}$ 5 SJ	2191 ± 506	0.78 ± 0.03	1.87 ± 1.03	9.23 ± 12.64	12.46 ± 9.48	0.82 ± 0.05	0.75 ± 0.06	0.87
$ASCHOPLEX_{tune}$ 6 SJ	2287 ± 499	0.78 ± 0.03	2.02 ± 1.23	14.37 ± 13.93	16.25 ± 11.68	0.83 ± 0.05	0.73 ± 0.06	0.87
$ASCHOPLEX_{tune}$ 7 SJ	2257 ± 506	0.78 ± 0.03	2.15 ± 1.83	12.63 ± 12.69	14.72 ± 10.20	0.83 ± 0.05	0.75 ± 0.06	0.88
$ASCHOPLEX_{tune}$ 8 SJ	2327 ± 506	0.78 ± 0.03	1.86 ± 0.87	16.26 ± 12.89	17.47 ± 11.19	0.84 ± 0.05	0.73 ± 0.05	0.89
$ASCHOPLEX_{tune}$ 9 SJ	2344 ± 514	0.78 ± 0.03	1.95 ± 1.48	17.07 ± 13.09	18.30 ± 11.29	0.80 ± 0.05	0.73 ± 0.05	0.89
$ASCHOPLEX_{tune}$ 10 SJ	2176 ± 513	0.79 ± 0.03	1.93 ± 1.85	8.19 ± 11.13	10.83 ± 8.59	0.82 ± 0.05	0.76 ± 0.05	0.92
FS	1037 ± 290	0.31 ± 0.08	11.94 ± 5.29	- 47.95 ± 11.58	47.95 ± 11.58	0.24 ± 0.07	0.46 ± 0.10	0.64
FastS	1029 ± 320	0.31 ± 0.07	5.88 ± 2.30	- 48.9 ± 11.09	48.9 ± 11.09	0.24 ± 0.07	0.46 ± 0.09	0.71
GMM	1581 ± 430	0.61 ± 0.05	8.08 ± 5.04	- 20.92 ± 15.32	22.55 ± 12.80	0.55 ± 0.07	0.70 ± 0.09	0.75
ChPSeg	2231 ± 612	0.65 ± 0.05	3.08 ± 1.33	10.86 ± 19.17	17.18 ± 13.81	0.68 ± 0.08	0.63 ± 0.07	0.77

All values (except for Absolute Volume and Pearson's Volume Correlation Coefficient) are presented in the form mean ± standard deviation. Legend: X SJ - number of subjects used during training of $ASCHOPLEX_{tune}$; n.s.- non significant.

analysis reported a positive and significant correlation between the Absolute Volume of all $ASCHOPLEX_{tune}$ approaches and the GT MSeg, particularly with more than five subjects ($r > 0.87$), while $ASCHOPLEX_0$ coefficient is non-significant. ANOVA test revealed an effect of group across all compared methods in Absolute Volume ($p_{value} < 0.0001$). In detail, GT MSeg is not statistically significantly different from $ASCHOPLEX_{tune}$ with more than five subjects (corrected $p_{value} = 1.0$) in the training set. On the contrary, $ASCHOPLEX_0$ is statistically significantly different from both all $ASCHOPLEX_{tune}$ configurations and GT MSeg (corrected $p_{value} < 0.0001$). $|\Delta Vol\%|$ and $\Delta Vol\%$ metrics confirm previous results. $ASCHOPLEX_{tune}$ with five subjects has been selected for the comparison with literature approaches. Fig. 5(a) shows the Absolute Volume distributions for $ASCHOPLEX_{tune}$ using five subjects in the training set, FS, FastS, GMM, ChPSeg, and GT MSeg.[5] In Fig. 5(b), the linear regression analysis shows a good ChPV prediction obtained by $ASCHOPLEX_{tune}$ corroborated by narrow confidence bounds. GMM underestimates ChPV with a systematic bias, while ChPSeg shows greater variability and shows a proportional bias in estimating the ChPV. For clarity, results obtained with FastS and FS are not reported due to lower performance compared to other methods. $ASCHOPLEX_{tune}$ with five subjects in the training set reports a higher Dice (0.78 ± 0.03) and lower *95% HD* (1.87 ± 1.03) compared to other methods (Table 3). One-way ANOVA on Absolute Volume highlighted an effect group ($p_{value} < 0.0001$). Post-hoc t-test reported that $ASCHOPLEX_{tune}$ is not statistically different from GT MSeg (corrected $p_{value} = 0.86$), as confirmed by other volume metrics ($\Delta Vol\% = 9.23 \pm 12.64$ %, $r = 0.86$), like ChPSeg (corrected $p_{value} = 0.27$), while FS, FastS, and GMM are significantly different (corrected $p_{value} < 0.0001$). Precision and recall confirm previous results. Fig. 6 shows the relative automatic segmentation results overlapped on the T1-w MRI image for two representative patients. The GT MSeg is reported as reference.

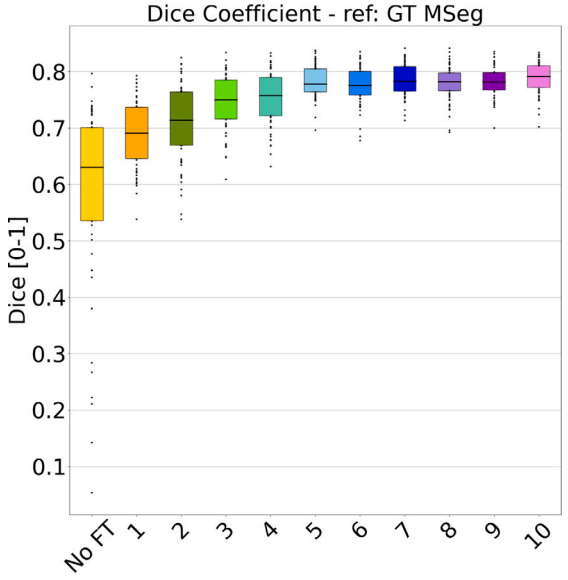


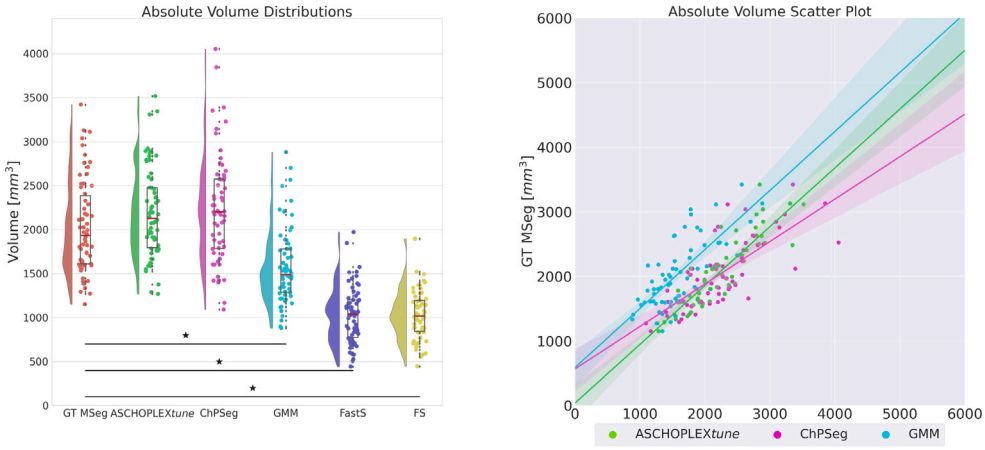
Fig. 4. Boxplot of Dice Coefficient segmentation metrics calculated over the 62 testing subjects of the fine-tuning dataset. Comparison of $ASCHOPLEX_0$ (No FT, ASCHOPLEX without fine-tuning procedure) and $ASCHOPLEX_{tune}$ (ASCHOPLEX with the fine-tuning procedure, using X subjects in the training set). The manual segmentation (GT MSeg) is used as reference.

4. Discussion

4.1. The importance of DNNs for the choroid plexus segmentation task

In this paper we propose a new approach called ASCHOPLEX to obtain an accurate, reliable, and fast semantic segmentation of the human ChP. ASCHOPLEX is a fully data-driven method that has proven to outperform the two state-of-the-art methods for brain segmentation in multiple datasets. Deep Learning based methods have become the state of the art for medical image processing tasks [43,44], with performance equivalent to human trained operators. Recently, the list of DNNs to choose from for medical imaging segmentation tasks has grown. Particularly, self-attention mechanisms of transformers brought interesting results. The UNETR is a novel architecture where a vision transformer is used as encoder to capture long range dependencies, while the U-Net convolutional architecture is employed as a decoder to learn local features [36]. In addition, the nnU-Net is a self-configuring net that raised high performance in the Medical Decathlon Segmentation Challenge [37]. Having too many possibilities makes it difficult to prefer one architecture over the others. From this evidence, ensemble strategies have been proposed. The ensemble consists of agreement procedures (e.g. major voting) between predictions derived from different trained DNNs, making the final segmentation more robust and reliable. Therefore, we followed this evidence to design a novel and generalizable ChP segmentation tool based on DNNs ensemble. Twenty-seven DNNs

[5] In *Supplementary Materials - Supplementary Figures (Figure B)* are shown the performance metrics for $ASCHOPLEX_{tune}$ using five subjects in the training set, FS, FastS, ChPSeg, and GMM.



(a) (b)

Fig. 5. (a) Rainplot of the absolute volume of the predicted segmentation obtained on the 62 testing subjects of the fine-tuning dataset with the compared methods ASCHOPLEX$_{\text{tune}}$ (five subjects in the training set), ChPSeg, Gaussian Mixture Models (GMM), FastSurfer (FastS), FreeSurfer (FS), and manual segmentation (GT MSeg), used as reference. Asterisks indicate the statistical differences between the absolute volumes of the compared methods and GT MSeg. (b) Scatter plot and linear regression model fit with 95% confidence interval bounds of the absolute volume of the predicted segmentation obtained on the 62 testing subjects of the fine-tuning dataset with the compared methods ASCHOPLEX$_{\text{tune}}$(five subjects in the training set), ChPSeg, Gaussian Mixture Models (GMM), and manual segmentation (GT MSeg), used as reference.

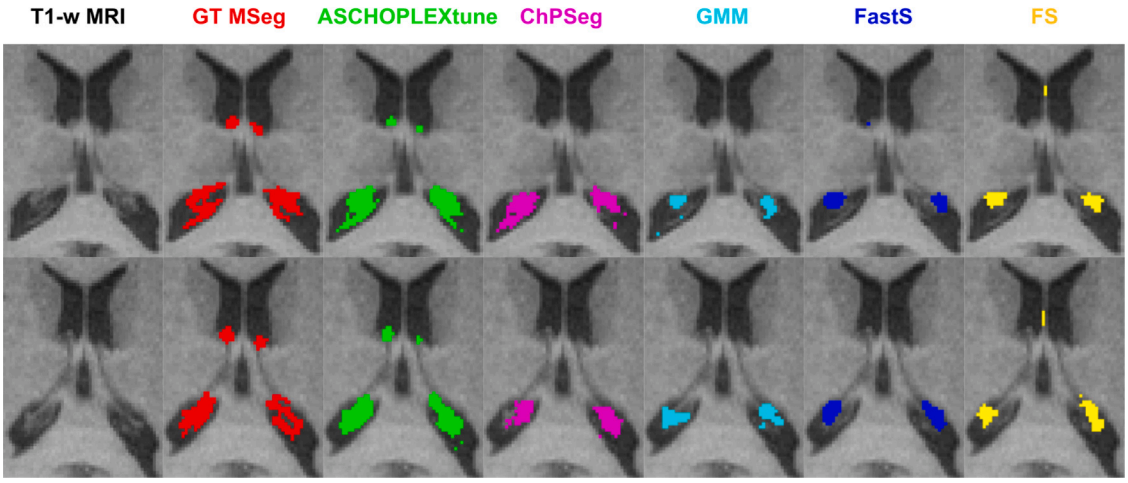


Fig. 6. Predicted segmentation results for two representative subjects of the testing set of the fine-tuning dataset. Zoom-in of 3D T1-w MRI (axial view) with superimposed the ChP segmentations. From left to right, 3D T1-w MRI with superimposed: manual segmentation (GT MSeg) in red, the ensemble result of ASCHOPLEX$_{\text{tune}}$ (five subjects in the training set) in green, FastSurfer (FastS) in blue, ChPSeg in pink, Gaussian Mixture Model (GMM) in light-blue, and FreeSurfer (FS) in yellow. Percentage Volume Difference metric (reference GT MSeg) for the top subject (best case patient): ASCHOPLEX −0.44%, ChPSeg −8.15%, FS −44.70%, FastS −55.36%, GMM −33.66%; for the bottom subject (worst case patient): ASCHOPLEX −18.79%, ChPSeg −20.52%, FS −59.86%, FastS −62.17%, GMM −38.95%.

configurations have been selected, varying both hyperparameters, like the network size and the cost function, and architectures: 3D U-Net, a fundamental tool for medical imaging segmentation, UNETR, that combine transformers with the U-Net, and the self-configuring net nnU-Net.

4.2. ASCHOPLEX core: DNNs common characteristics

The models selected to be part of the ASCHOPLEX ensemble have common characteristics. These were (1) the DNNs architecture type, (2) the dimension of the input patch, and (3) the chosen loss function. The selected DNNs architectures are Dynunet (MONAI implementation of nnU-Net) and UNETR, new generations of DNNs architectures for Medical Image Segmentation, suggesting that the use of self-configuring nets or attention mechanisms improves the results for this segmentation task than simpler 3D U-Net. Regarding the network dimension, only the bigger patch size (128 × 128 × 128) was selected. Large networks lead on one hand to higher performance, and on the other hand, to increased costs, both in terms of increased training time and minimum hardware requirements [37]. The most selected loss function is DiceCE, concordantly as reported in previous studies [36,37]. Including Dice calculation in the loss function helps when the number of background and foreground voxels is unbalanced and, consequently, improves similarity between GT MSeg and prediction [45].

4.3. ASCHOPLEX segmentation performance

ASCHOPLEX improves ChP predictions on the testing set both in terms of similarity indices and ChPV estimate. The mean error committed by ASCHOPLEX in ChPV estimate is 9%, that is 17%, 20%, 32%, and 35% lower than, respectively, ChPSeg, GMM, FastS, and FS. In addition, the volumes of the GT MSeg and the proposed method have similar distributions and the statistical tests revealed there are no statistically significant differences between the two resulting populations. The mean error committed by ASCHOPLEX is randomic and not affected by systematic bias because both the training and testing sets contain both controls and diseased subjects. To our knowledge, the evaluated differences of the ChPV, compared to controls, is around 21.4% for MS [46] and 12.9% for depressed patients [10]. The error committed by ASCHOPLEX is under those thresholds, pointing out the ability of this toolbox to detect a distinction between healthy and diseased populations. These findings highlight the ability of ASCHOPLEX to emulate, better than other methods, the neuroradiologist in the manual segmentation task, and improve the manual approach in speed and reproducibility. The mediocre ability of FS and FastS to emulate the neuroradiologist has already been verified [12,14]. Despite suboptimal performance, the main advantage of deterministic methods like FS and GMM compared to data-driven ones (e.g. DNNs) is the generalizability. However GMM's design was optimized on datasets of healthy controls (Human Connectome Project [47]), mild cognitive impairment, and Alzheimer Disease (Alzheimer's Disease Neuroimaging Initiative [48]) and therefore it might be that different protocols and different pathologies (such as MS or Depression) require a different parameter optimization. ChPSeg is deep-learning based tool with preprocessing steps that allow it to standardize images and theoretically increase its generalization ability. Nevertheless, the cohort used for the training of this tool was composed by both controls and patients with mild cognitive impairment, Alzheimer's disease, Parkinson's disease, and Huntington's disease, that explain why ChPSeg shows unacceptable performance when applied to MS subjects.

4.4. ASCHOPLEX$_{tune}$: a solution for the generalizability problem

ASCHOPLEX can overcome the generalizability issues, both in terms of different populations and in terms of manufacturer shifts, by implementing a fine-tuning step, ASCHOPLEX$_{\text{tune}}$, making the already trained models adaptable to an unseen dataset. Moreover, the procedure is also easily available to non-expert deep learning programmers and can help adapt this model to previously unobserved datasets. The challenge was to investigate if a short additional training combined with a small dataset is sufficient to adapt the tool to a new dataset obtaining better performance than FS, FastS, GMM, and, above all, ChPSeg. The comparison between the direct application of ASCHOPLEX, and ASCHOPLEX$_{\text{tune}}$ has shown that the direct inference of the trained model over an unseen dataset with a different pattern of characteristics brings a drop of performance. A fine-tuning step applied before application on a new dataset improves the results. In particular, results suggest that the best compromise is to have a sufficiently large training set to be able to increase the model's knowledge of the characteristics of the new dataset, and to minimize the number of manual segmentations required to perform the fine-tuning step; here we set the number of subjects in the ASCHOPLEX$_{\text{tune}}$ training set to five. The main disadvantage of implementing a fine-tuning step is the cost, both in terms of hardware requirements (GPU), and the extra time required to run the short training. Based on our investigations, the extra fine-tuning procedure requires manual segmentations of at least ten subjects (five for the training set and five for the validation). However, the added value of the fine-tuning is the invaluable capability of ASCHOPLEX$_{\text{tune}}$ to achieve the same performance of ASCHOPLEX without training the models from scratch [49]. In addition, ASCHOPLEX approach gives the final user the possibility to run a fine-tuning step without writing additional code, and to eventually modify ASCHOPLEX fine-tuning parameters. Although the number of subjects and the pathological status are different between the two testing sets (36 MS subjects for ASCHOPLEX, 62 depressed subjects for ASCHOPLEX$_{\text{tune}}$), segmentation metrics are comparable. Considering the above results, the fine-tuning procedure is recommended when ASCHOPLEX has to be inferred on a new dataset; particularly, ASCHOPLEX$_{\text{tune}}$ with five subjects in the training set guarantees an optimal trade-off between the number of extra subjects to be included and the overall segmentation performance, that are comparable to ASCHOPLEX on the original training dataset.

4.5. Limitations

This study has limitations. Firstly, the entire data sample has been acquired with MRI scanners with magnetic field strength of 3T at a resolution of $1 \times 1 \times 1$ mm. Future directions to be addressed might be the enrichment of the dataset by varying field strength or resolution. In addition, the ChP is merely visible in the third and fourth ventricles with the employed resolution. However, the use of higher field strength acquisitions (i.e. 7T MRI or above) could help in detecting and segmenting the ChP not only in the lateral ventricles but also in the third and fourth ventricles, which might be helpful in clinical evaluations. Nevertheless, the three tested datasets comprehend different neurological disorders (MS, depression) and have been acquired with different scanners (MRI and PET/MRI) derived from different manufacturers (Philips and GE). Both these variables contribute to boost the confidence in the generalizability of the tool. Another limitation of this work is the use of different software for the MSeg depiction, however, testing their influence on the segmentation results is out of the scope of the paper. Moreover, the inter- and intra-rater variability of the neuroradiologists when performing the ChP manual segmentation task was not investigated as it is out of the scope of the paper and has been previously addressed [16]. Lastly, the impact of neuroimaging preprocessing steps (e.g. brain extraction, N4 bias field correction) on ASCHOPLEX performances has not been investigated, as preprocessing steps are not required for ASCHOPLEX to run. Recent studies suggest avoiding preprocessing steps for medical imaging segmentation tasks [50], but we did not explicitly test it in our samples.

4.6. Conclusion

We can conclude that ASCHOPLEX is a valid tool for obtaining an accurate, and reliable segmentation of the ChP with performance that overcomes the literature reference standard in estimating ChPV. The addition of a fine-tuning step improves generalizability and enables accurate ChPV estimates for unseen datasets. ASCHOPLEX is a transdiagnostic, generalizable, reliable, tool for ChP segmentation, which may further facilitate the use of ChPV as a biomarker.

CRediT authorship contribution statement

Valentina Visani: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Mattia Veronese:** Writing – review & editing, Supervision, Resources. **Francesca B. Pizzini:** Resources, Investigation, Data curation, Conceptualization. **Annalisa Colombi:** Data curation. **Valerio Natale:** Data curation. **Corina Marjin:** Data curation. **Agnese Tamanti:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Julia J. Schubert:** Writing – review & editing, Formal analysis, Data curation. **Noha Althubaity:** Data curation. **Inés Bedmar-Gómez:** Data curation. **Neil A. Harrison:** Writing – review & editing, Resources. **Edward T. Bullmore:** Resources. **Federico E. Turkheimer:** Writing – review & editing, Resources. **NIMA Consortium:** Resources, Investigation. **Massimiliano Calabrese:** Supervision, Resources, Investigation, Conceptualization. **Marco Castellaro:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The BIODEP study was sponsored by the Cambridgeshire and Peterborough NHS Foundation Trust, UK and the University of Cambridge, UK, and funded by a strategic award from the Wellcome Trust, UK (104025) in partnership with Janssen, GlaxoSmithKline, Lundbeck and Pfizer. Recruitment of participants was supported by the National Institute of Health Research (NIHR), UK Clinical Research Network: Kent, Surrey and Sussex & Eastern. Additional funding was provided by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, UK, and by the NIHR Cambridge Biomedical Research Centre (Mental Health), UK. ETB is supported by a Senior Investigator award from the NIHR. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Cambridge.[6] We would like to gratefully thank all study participants, research teams and laboratory staff, without whom this research would not have been possible.[7]

J. J. Schubert and F. E. Turkheimer are supported by the NIHR Maudsley Biomedical Research Centre at South London, UK and Maudsley NHS Foundation Trust and King's College London, UK.

Appendix A

NIMA members during the sample collection and data analysis period for the BIODEP Study

Brighton & Sussex University Hospitals NHS Trust Dominika Wlazly

Cambridgeshire & Peterborough NHS Foundation Trust Amber Dickinson, Andy Foster, Clare Knight

Cardiff University Claire Leckey, Paul Morgan, Angharad Morgan, Caroline O'Hagan, Samuel Touchard

GSK Shahid Khan, Phil Murphy, Christine Parker, Jai Patel, Jill Richardson

Janssen Paul Acton, Nigel Austin, Anindya Bhattacharya, Nick Carruthers, Peter de Boer, Wayne Drevets, John Isaac, Declan Jones, John Kemp, Hartmuth Kolb, Jeff Nye, Gayle Wittenberg

Kings College London Gareth Barker, Anna Bogdanova, Heidi Byrom, Diana Cash, Annamaria Cattaneo, Daniela Enache, Tony Gee, Caitlin Hastings, Melisa Kose, Giulia Lombardo, Nicole Mariani, Anna McLaughlin, Valeria Mondelli, Maria Nettis, Naghmeh Nikkheslat, Carmine Pariante, Karen Randall, Julia Schubert, Luca Sforzini, Hannah Sheridan, Camilla Simmons, Nisha Singh, Federico Turkheimer, Vicky Van Loo, Mattia Veronese, Marta Vicente Rodriguez, Toby Wood, Courtney Worrell, Zuzanna Zajkowska

Lundbeck Brian Campbell, Jan Egebjerg, Hans Eriksson, Francois Gastambide, Karen Husted Adams, Ross Jeggo, Thomas Moeller, Bob Nelson, Niels Plath, Christian Thomsen, Jan Torleif Pederson, Stevin Zorn

NHS Greater Glasgow and Clyde Catherine Deith, Scott Farmer, John McClean, Andrew McPherson, Nagore Penandes, Paul Scouller, Murray Sutherland

Oxford Health NHS Foundation Trust Mary Jane Attenburrow, Jithen Benjamin, Helen Jones, Fran Mada, Akintayo Oladejo, Katy Smith

Pfizer Rita Balice-Gordon, Brendon Binneman, James Duerr, Terence Fullerton, Veeru Goli, Zoe Hughes, Justin Piro, Tarek Samad, Jonathan Sporn

Sussex Partnership NHS Foundation Trust Liz Hoskins, Charmaine Kohn, Lauren Wilcock

University of Cambridge Franklin Aigbirhio, Junaid Bhatti, Ed Bullmore, Sam Chamberlain, Marta Correia, Anna Crofts, Tim Fryer, Martin Graves, Alex Hatton, Manfred Kitzbichler, Mary-Ellen Lynall, Christina Maurice, Ciara O'Donnell, Linda Pointon, Peter St George Hyslop, Lorinda Turner, Petra Vertes, Barry Widmer, Guy Williams

University of Glasgow Jonathan Cavanagh, Alison McColl, Robin Shaw

University of Groningen Erik Boddeke

University of Oxford Alison Baird, Stuart Clare, Phil Cowen, I-Shu (Dante) Huang, Sam Hurley, Simon Lovestone, Alejo Nevado-Holgado, Elena Ribe, Anviti Vyas, Laura Winchester

University of Southampton Madeleine Cleal, Diego Gomez-Nicola, Renzo Mancuso, Hugh Perry

University of Sussex Mara Cercignani, Charlotte Clarke, Alessandro Colasanti, Neil Harrison, Rosemary Murray

University of Texas Jason O'Connor

University of Toronto Howard Mount

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.109164.

References

[1] H.H. Damkier, P.D. Brown, J. Praetorius, Cerebrospinal fluid secretion by the choroid plexus, Physiol. Rev. 93 (4) (2013) 1847–1892, http://dx.doi.org/10.1152/physrev.00004.2013.

[2] S. Balusu, M. Brkic, C. Libert, R. Vandenbroucke, The choroid plexus-cerebrospinal fluid interface in Alzheimer's disease: more than just a barrier, Neural Regen. Res. 11 (4) (2016) 534, http://dx.doi.org/10.4103/1673-5374.180372.

[3] R. Spector, R.F. Keep, S. Robert Snodgrass, Q.R. Smith, C.E. Johanson, A balanced view of choroid plexus structure and function: Focus on adult humans, Exp. Neurol. 267 (2015) 78–86, http://dx.doi.org/10.1016/j.expneurol.2015.02.032.

[4] N.A. Jessen, A.S.F. Munk, I. Lundgaard, M. Nedergaard, The glymphatic system: A beginner's guide, Neurochem. Res. 40 (12) (2015) 2583–2599, http://dx.doi.org/10.1007/s11064-015-1581-6.

[5] B. Engelhardt, K. Wolburg-Buchholz, H. Wolburg, Involvement of the choroid plexus in central nervous system inflammation, Microsc. Res. Tech. 52 (1) (2001) 112–129, http://dx.doi.org/10.1002/1097-0029(20010101)52:1<112::AID-JEMT13>3.0.CO;2-5.

[6] J.J. Eisma, C.D. McKnight, K. Hett, J. Elenberger, C.J. Han, A.K. Song, C. Considine, D.O. Claassen, M.J. Donahue, Deep learning segmentation of the choroid plexus from structural magnetic resonance imaging (MRI): validation and normative ranges across the adult lifespan, Fluids Barriers CNS 21 (21) (2024) http://dx.doi.org/10.1186/s12987-024-00525-9.

[7] S. Monaco, R. Nicholas, R. Reynolds, R. Magliozzi, Intrathecal inflammation in progressive multiple sclerosis, Int. J. Mol. Sci. 21 (21) (2020) 8217, http://dx.doi.org/10.3390/ijms21218217.

[8] J.S.R. Alisch, M. Kiely, C. Triebswetter, M.H. Alsameen, Z. Gong, N. Khattar, J.M. Egan, M. Bouhrara, Characterization of age-related differences in the human choroid plexus volume, microstructural integrity, and blood perfusion using multiparameter magnetic resonance imaging, Front. Aging Neurosci. 13 (2021) 734992, http://dx.doi.org/10.3389/fnagi.2021.734992.

[9] J.J. Eisma, C.D. McKnight, K. Hett, J. Elenberger, A.K. Song, A.J. Stark, D.O. Claassen, M.J. Donahue, Choroid plexus perfusion and bulk cerebrospinal fluid flow across the adult lifespan, J. Cereb. Blood Flow Metabol. 43 (2) (2023) 269–280, http://dx.doi.org/10.1177/0271678X221129101.

[10] N. Althubaity, J. Schubert, D. Martins, T. Yousaf, M.A. Nettis, V. Mondelli, C. Pariante, N.A. Harrison, E.T. Bullmore, D. Dima, F.E. Turkheimer, M. Veronese, Choroid plexus enlargement is associated with neuroinflammation and reduction of blood brain barrier permeability in depression, NeuroImage: Clinical 33 (2022) 102926, http://dx.doi.org/10.1016/j.nicl.2021.102926.

[6] Paul A. Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde, Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81.

[7] All members of the NIMA Consortium at the time of data collection are thanked and acknowledged in the *Appendix*.

[11] J.J. Schubert, M. Veronese, L. Marchitelli, B. Bodini, M. Tonietto, B. Stankoff, D.J. Brooks, A. Bertoldo, P. Edison, F.E. Turkheimer, Dynamic ^{11}C-PiB PET shows cerebrospinal fluid flow alterations in Alzheimer disease and multiple sclerosis, J. Nucl. Med. 60 (10) (2019) 1452–1460, http://dx.doi.org/10.2967/jnumed.118.223834.

[12] E. Tadayon, B. Moret, G. Sprugnoli, L. Monti, A. Pascual-Leone, E. Santarnecchi, Alzheimer's Disease Neuroimaging Initiative, Improving choroid plexus segmentation in the healthy and diseased brain: Relevance for tau-PET imaging in dementia, J. Alzheimer Dis. JAD 74 (4) (2020) 1057–1068, http://dx.doi.org/10.3233/JAD-190706.

[13] V. Hubert, F. Chauveau, C. Dumot, E. Ong, L.P. Berner, E. Canet-Soulas, J.F. Ghersi-Egea, M. Wiart, Clinical imaging of choroid plexus in health and in brain disorders: A mini-review, Front. Mol. Neurosci. 12 (2019) 34, http://dx.doi.org/10.3389/fnmol.2019.00034.

[14] A. Yazdan-Panah, M. Schmidt-Mengin, V.A. Ricigliano, T. Soulier, B. Stankoff, O. Colliot, Automatic segmentation of the choroid plexuses: Method and validation in controls and patients with multiple sclerosis, NeuroImage: Clinical 38 (2023) 103368, http://dx.doi.org/10.1016/j.nicl.2023.103368.

[15] O. Senay, M. Seethaler, N. Makris, E. Yeterian, J. Rushmore, K.I.K. Cho, E. Rizzoni, C. Heller, O. Pasternak, F. Szczepankiewicz, C.F. Westin, J. Losak, L. Ustohal, J. Tomandl, L. Vojtisek, P. Kudlicka, Z. Kikinis, D. Holt, K.E. Lewandowski, P. Lizano, M.S. Keshavan, D. Öngür, T. Kasparek, A. Breier, M.E. Shenton, J. Seitz-Holland, M. Kubicki, A preliminary choroid plexus volumetric study in individuals with psychosis, Hum. Brain Mapp. 44 (6) (2023) 2465–2478, http://dx.doi.org/10.1002/hbm.26224.

[16] V. Visani, F.B. Pizzini, V. Natale, A. Tamanti, M. Anglani, A. Bertoldo, M. Calabrese, M. Castellaro, Choroid plexus volume in multiple sclerosis can be estimated on structural MRI avoiding contrast injection, Eur. Radiol. Exp. 8 (33) (2024) http://dx.doi.org/10.1186/s41747-024-00421-9.

[17] E. Tadayon, A. Pascual-Leone, D. Press, E. Santarnecchi, Choroid plexus volume is associated with levels of CSF proteins: relevance for Alzheimer's and Parkinson's disease, Neurobiol. Aging 89 (2020) 108–117, http://dx.doi.org/10.1016/j.neurobiolaging.2020.01.005.

[18] J.D. Choi, Y. Moon, H.J. Kim, Y. Yim, S. Lee, W.J. Moon, Choroid plexus volume and permeability at brain MRI within the Alzheimer disease clinical spectrum, Radiology 304 (3) (2022) 635–645, http://dx.doi.org/10.1148/radiol.212400.

[19] V.A.G. Ricigliano, E. Morena, A. Colombi, M. Tonietto, M. Hamzaoui, E. Poirion, M. Bottlaender, P. Gervais, C. Louapre, B. Bodini, B. Stankoff, Choroid plexus enlargement in inflammatory multiple sclerosis: 3.0-T MRI and translocator protein PET evaluation, Radiology 301 (1) (2021) 166–177, http://dx.doi.org/10.1148/radiol.2021204426.

[20] M. Vercellino, B. Votta, C. Condello, C. Piacentino, A. Romagnolo, A. Merola, E. Capello, G.L. Mancardi, R. Mutani, M.T. Giordana, P. Cavalla, Involvement of the choroid plexus in multiple sclerosis autoimmune inflammation: A neuropathological study, J. Neuroimmunol. 199 (1) (2008) 133–141, http://dx.doi.org/10.1016/j.jneuroim.2008.04.035.

[21] H. Lassmann, Pathogenic mechanisms associated with different clinical courses of multiple sclerosis, Front. Immunol. 9 (2019) 3116, http://dx.doi.org/10.3389/fimmu.2018.03116.

[22] V. Fleischer, G. Gonzalez-Escamilla, D. Ciolac, P. Albrecht, P. Küry, J. Gruchot, M. Dietrich, C. Hecker, T. Müntefering, S. Bock, M. Oshaghi, A. Radetz, M. Cerina, J. Krämer, L. Wachsmuth, C. Faber, H. Lassmann, T. Ruck, S.G. Meuth, M. Muthuraman, S. Groppa, Translational value of choroid plexus imaging for tracking neuroinflammation in mice and humans, Proc. Natl. Acad. Sci. 118 (36) (2021) e2025000118, http://dx.doi.org/10.1073/pnas.2025000118.

[23] N. Manouchehri, O. Stüve, Choroid plexus volumetrics and brain inflammation in multiple sclerosis, Proc. Natl. Acad. Sci. 118 (40) (2021) e2115221118, http://dx.doi.org/10.1073/pnas.2115221118.

[24] P. Lizano, O. Lutz, G. Ling, A.M. Lee, S. Eum, J.R. Bishop, S. Kelly, O. Pasternak, B. Clementz, G. Pearlson, J.A. Sweeney, E. Gershon, C. Tamminga, M. Keshavan, Association of choroid plexus enlargement with cognitive, inflammatory, and structural phenotypes across the psychosis spectrum, Am. J. Psychiatry 176 (7) (2019) 564–572, http://dx.doi.org/10.1176/appi.ajp.2019.18070825.

[25] Y.F. Zhou, J.C. Huang, P. Zhang, F.M. Fan, S. Chen, H.-Z. Fan, Y.M. Cui, X.G. Luo, S.P. Tan, Z.R. Wang, W. Feng, Y. Yuan, F.D. Yang, A. Savransky, M. Ryan, E. Goldwaser, J. Chiappelli, L.M. Rowland, P. Kochunov, Y.L. Tan, L.E. Hong, Choroid plexus enlargement and allostatic load in Schizophrenia, Schizophr. Bull. 46 (3) (2020) 722–731, http://dx.doi.org/10.1093/schbul/sbz100.

[26] Y. Duan, S. Qi, Editorial for a fully-automatic method to segment choroid plexuses in multiple sclerosis using conventional MRI sequences, J. Magn. Reson. Imaging (2023) jmri.28976, http://dx.doi.org/10.1002/jmri.28976.

[27] M. Schmidt-Mengin, V.A.G. Ricigliano, B. Bodini, E. Morena, A. Colombi, M. Hamzaoui, A.Y. Panah, B. Stankoff, O. Colliot, Axial multi-layer perceptron architecture for automatic segmentation of choroid plexus in multiple sclerosis, in: O. Colliot, I.I. sgum (Eds.), Medical Imaging 2022: Image Processing, Vol. 12032, International Society for Optics and Photonics, 2022, 1203208, http://dx.doi.org/10.1117/12.2612912.

[28] B. Fischl, FreeSurfer, NeuroImage 62 (2) (2012) 774–781, http://dx.doi.org/10.1016/j.neuroimage.2012.01.021.

[29] L. Storelli, E. Pagani, M. Rubin, M. Margoni, M. Filippi, M.A. Rocca, A fully automatic method to segment choroid plexuses in multiple sclerosis using conventional MRI sequences, J. Magn. Reson. Imaging (2023) jmri.28937, http://dx.doi.org/10.1002/jmri.28937.

[30] L. Zhao, X. Feng, C.H. Meyer, D.C. Alsop, Choroid plexus segmentation using optimized 3D U-net, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, ISBI, IEEE, 2020, pp. 381–384, http://dx.doi.org/10.1109/ISBI45749.2020.9098443.

[31] K. de Raad, K. van Garderen, M. Smits, S. van der Voort, F. Incekara, E. Oei, J. Hirvasniemi, S. Klein, M. Starmans, The effect of preprocessing on convolutional neural networks for medical image segmentation, in: 2021 IEEE 18th International Symposium on Biomedical Imaging, ISBI, 2021, pp. 655–658, http://dx.doi.org/10.1109/ISBI48211.2021.9433952.

[32] L. Henschel, S. Conjeti, S. Estrada, K. Diers, B. Fischl, M. Reuter, FastSurfer - A fast and accurate deep learning based neuroimaging pipeline, NeuroImage 219 (2020) 117012, http://dx.doi.org/10.1016/j.neuroimage.2020.117012.

[33] S.H. Jeong, C.J. Park, J. Cha, S.Y. Kim, S.K. Lee, Y.J. Kim, Y.H. Sohn, S.J. Chung, P.H. Lee, Choroid plexus volume, amyloid burden, and cognition in the Alzheimer's disease continuum, Aging Dis. (2024) https://www.aginganddisease.org/CN/10.14336/AD.2024.0118.

[34] Consortium MONAI, MONAI: Medical open network for AI, 2022, http://dx.doi.org/10.5281/zenodo.7245821.

[35] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[36] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H.R. Roth, D. Xu, UNETR: Transformers for 3D medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, IEEE Computer Society, Los Alamitos, CA, USA, 2022, pp. 1748–1758, http://dx.doi.org/10.1109/WACV51458.2022.00181.

[37] F. Isensee, P.F. Jaeger, S.A.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211, http://dx.doi.org/10.1038/s41592-020-01008-z.

[38] J.J. Schubert, M. Veronese, T.D. Fryer, R. Manavaki, M.G. Kitzbichler, M.A. Nettis, V. Mondelli, C.M. Pariante, E.T. Bullmore, F.E. Turkheimer, D. Wlazly, A. Dickinson, A. Foster, C. Knight, C. Leckey, P. Morgan, A. Morgan, C. O'Hagan, S. Touchard, S. Khan, P. Murphy, C. Parker, J. Patel, J. Richardson, P. Acton, N. Austin, A. Bhattacharya, N. Carruthers, P. De Boer, W. Drevets, J. Isaac, D. Jones, J. Kemp, H. Kolb, J. Nye, G. Wittenberg, G. Barker, A. Bogdanova, H. Byrom, D. Cash, A. Cattaneo, D. Enache, T. Gee, C. Hastings, M. Kose, G. Lombardo, N. Mariani, A. McLaughlin, V. Mondelli, M. Nettis, N. Nikkheslat, C. Pariante, K. Randall, J. Schubert, L. Sforzini, H. Sheridan, C. Simmons, N. Singh, F. Turkheimer, V. Van Loo, M. Veronese, M.V. Rodriguez, T. Wood, C. Worrell, Z. Zajkowska, B. Campbell, J. Egebjerg, H. Eriksson, F. Gastambide, K.H. Adams, R. Jeggo, T. Moeller, B. Nelson, N. Plath, C. Thomsen, J.T. Pederson, S. Zorn, C. Deith, S. Farmer, J. McClean, A. McPherson, N. Penandes, P. Scouller, M. Sutherland, M.J. Attenburrow, J. Benjamin, H. Jones, F. Mada, A. Oladejo, K. Smith, R. Balice-Gordon, B. Binneman, J. Duerr, T. Fullerton, V. Goli, Z. Hughes, J. Piro, T. Samad, J. Sporn, L. Hoskins, C. Kohn, L. Wilcock, F. Aigbirhio, J. Bhatti, E. Bullmore, S. Chamberlain, M. Correia, A. Crofts, T. Fryer, M. Graves, A. Hatton, M. Kitzbichler, M.-E. Lynall, C. Maurice, C. O'Donnell, L. Pointon, P. St George Hyslop, L. Turner, P. Vertes, B. Widmer, G. Williams, J. Cavanagh, A. McColl, R. Shaw, E. Boddeke, A. Baird, S. Clare, P. Cowen, I.S.D. Huang, S. Hurley, S. Lovestone, A. Nevado-Holgado, E. Ribe, A. Vyas, L. Winchester, M. Cleal, D. Gomez-Nicola, R. Mancuso, H. Perry, M. Cercignani, C. Clarke, A. Colasanti, N. Harrison, R. Murray, J. O'Connor, H. Mount, A modest increase in 11C-PK11195-positron emission tomography TSPO binding in depression is not associated with serum C-reactive protein or body mass index, Biol. Psychiatry: Cogn. Neurosci. Neuroimaging 6 (7) (2021) 716–724, http://dx.doi.org/10.1016/j.bpsc.2020.12.017.

[39] P. Schmidt, V. Pongratz, P. Küster, D. Meier, J. Wuerfel, C. Lukas, B. Bellenberg, F. Zipp, S. Groppa, P.G. Sämann, F. Weber, C. Gaser, T. Franke, M. Bussas, J. Kirschke, C. Zimmer, B. Hemmer, M. Mühlau, Automated segmentation of changes in FLAIR-hyperintense white matter lesions in multiple sclerosis on serial magnetic resonance imaging, NeuroImage: Clinical 23 (2019) 101849, http://dx.doi.org/10.1016/j.nicl.2019.101849.

[40] P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability, NeuroImage 31 (3) (2006) 1116–1128, http://dx.doi.org/10.1016/j.neuroimage.2006.01.015.

[41] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, in: International Conference on Learning Representations, Vol. 6, ICLR, 2019, pp. 4061–4078.

[42] W. Yan, L. Huang, L. Xia, S. Gu, F. Yan, Y. Wang, Q. Tao, MRI manufacturer shift and adaptation: Increasing the generalizability of deep learning segmentation for MR images acquired with different scanners, Radiology: Artif. Intell. 2 (4) (2020) e190195, http://dx.doi.org/10.1148/ryai.2020190195.

[43] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, http://dx.doi.org/10.1038/nature14539.

[44] A.U. Ruby, Binary cross entropy with deep learning technique for image classification, Int. J. Adv. Trends Comput. Sci. Eng. 9 (4) (2020) 5393–5397, http://dx.doi.org/10.30534/ijatcse/2020/175942020.

[45] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: IEEE 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, pp. 565–571, http://dx.doi.org/10.1109/3DV.2016.79.

[46] J. Müller, T. Sinnecker, M.J. Wendebourg, R. Schläger, J. Kuhle, S. Schädelin, P. Benkert, T. Derfuss, P. Cattin, C. Jud, F. Spiess, M. Amann, T. Lincke, M. Barakovic, A. Cagol, C. Tsagkas, K. Parmar, A.K. Pröbstel, S. Reimann, S. Asseyer, A. Duchow, A. Brandt, K. Ruprecht, N. Hadjikhani, S. Fukumoto, M. Watanabe, K. Masaki, T. Matsushita, N. Isobe, J.-I. Kira, L. Kappos, J. Würfel, C. Granziera, F. Paul, Ö. Yaldizli, Choroid plexus volume in multiple sclerosis vs neuromyelitis optica spectrum disorder, Neurol. Neuroimmunol. Neuroinflammation 9 (3) (2022) e1147, http://dx.doi.org/10.1212/NXI.0000000000001147.

[47] M.F. Glasser, S.M. Smith, D.S. Marcus, J.L.R. Andersson, E.J. Auerbach, T.E.J. Behrens, T.S. Coalson, M.P. Harms, M. Jenkinson, S. Moeller, E.C. Robinson, S.N. Sotiropoulos, J. Xu, E. Yacoub, K. Ugurbil, D.C. Van Essen, The human connectome project's neuroimaging approach, Nature Neurosci. 19 (9) (2016) 1175–1187, http://dx.doi.org/10.1038/nn.4361.

[48] R.C. Petersen, P.S. Aisen, L.A. Beckett, M.C. Donohue, A.C. Gamst, D.J. Harvey, C.R. Jack, W.J. Jagust, L.M. Shaw, A.W. Toga, J.Q. Trojanowski, M.W. Weiner, Alzheimer's disease neuroimaging initiative (ADNI): Clinical characterization, Neurology 74 (3) (2010) 201–209, http://dx.doi.org/10.1212/WNL.0b013e3181cb3e25.

[49] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans. Med. Imaging 35 (5) (2016) 1299–1312, http://dx.doi.org/10.1109/TMI.2016.2535302.

[50] B.M. Pacheco, G.D.S.E. Cassia, D. Silva, Towards fully automated deep-learning-based brain tumor segmentation: Is brain extraction still necessary? Biomed. Signal Process. Control 82 (2023) 104514, http://dx.doi.org/10.1016/j.bspc.2022.104514.