UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN BIOINGEGNERIA

"Study of the relationship between choroid plexus volume and CSF derived biomarkers in patients with multiple sclerosis"

Relatore: Dott. Ing. Marco Castellaro

Laureando/a: Martina Cecchinato

Correlatore: Dott. Ing. Valentina Visani

ANNO ACCADEMICO 2022 – 2023

Data di laurea 26/10/23

ABSTRACT

Choroid Plexus (ChP) consists of highly vascularized veil-like organoids located in the brain ventricles and its primary function is the production of the Cerebrospinal Fluid (CSF). ChP structural and functional alterations may contribute to the pathophysiology of neurodegenerative diseases such as Multiple Sclerosis (MS). It can be imaged noninvasively with Magnetic Resonance Imaging (MRI), and recent studies highlighted that ChP is altered and particularly, it is enlarged in MS and other disorders, suggesting its use as a proxy of brain neuroinflammation and raising interest on the ChP volume (ChPV) as potential biomarker of neuroinflammation.

The goal of this study is to investigate whether ChP alterations, measured by the ChPV can be related to biomarkers (cytokines and chemokines) obtained from CSF analysis in patients with MS. The final aim is a better understanding of underlying processes of MS and the potential use of ChPV as new biomarker for faster and non-invasive disease monitoring.

To reach this objective, the ChP of 167 progressive MS subjects from two different acquisition sites has been segmented through ASCHOPLEX, a new approach for ChP automated segmentation. Initially, ASCHOPLEX output was manually validated by a quality check procedure, that highlighted the partial erroneous inclusion of the hippocampus boundaries in the ChP segmented mask in some subjects. This aspect has been deepened through the study on a standardized space (MNI) for comparing the position of ChP segmentations in the brain among the entire cohort. The inclusion of the hippocampus has been corrected by excluding some voxels on the original segmentations through a thresholding operation in MNI space.

ChPV was then divided by the Total Intracranial Volume (TIV) to remove any residual brain-size dependance and finally NeuroCombat approach was used to harmonize ChPV by removing unwanted scan variability.

An additional dataset of 130 CSF molecules derived from the same subjects has also been provided for this study. The processing of these data consisted in a preliminary statistical study to have an overview of distribution, missing values, and collinearity between molecules. Based on this analysis, the number of collected variables has been reduced, considering both the dataset coverage and the presence of missing values A subset of selected molecules was investigated in detail because of previous evidence of a relationship with MS, and they have been considered a hypothesis-driven selection.

NChPV and CSF molecules relationship has been studied using linear regression. To apply it, a previous selection of relevant variables has been performed with Lasso and Stepwise

regression, obtaining two sets of data-driven selections. Lasso Regression reaches an adjusted $R^2 = 0.23$ when using 8 variables derived from the 54 molecules and applied to the entire cohort; Stepwise regression instead, reports an adjusted $R^2 = 0.46$ when applied on a reduced cohort of subjects and variables without missing values, and $R^2 = 0.30$ when applied to the entire cohort. Even if the performance of linear models employed in this study are not optimal, some variables show a significant predictive role (i.e., TNF, CX3CL1, IFNγ and CCL11). These molecules have been previously highlighted to be related to inflammatory aspect within the CNS of MS patients confirming that the alteration of the CSF biological profile might be related to the alteration of ChP.

The main finding of this work is that the molecules derived from data-driven selection give a significant contribution in predicting the NChPV and are consistent with those of the hypothesis-driven selection. Particularly, they are all related to inflammation processes or to immune cells recruiting. To conclude, this study might lay the foundation to a better understanding of the relationship between inflammatory mechanisms in MS and ChPV, showing the promising usefulness of ChPV as a new non-invasive neuroimaging biomarker.

Keywords: Choroid Plexus, Multiple Sclerosis, CSF biomarkers, segmentation

Summary

1. INTRODUCTION

1.1 Multiple Sclerosis

Multiple sclerosis (MS) is a chronic autoimmune-mediated Central Nervous System (CNS) disease that affects more than 2 million people around the world [1], characterized by chronic inflammation, demyelination, gliosis and neuronal loss [2].

MS is threefold more common in females than in male and it has the highest prevalence among those between the ages of 20 and 40 years [1], but it can present at any age and almost 10% cases present before the age of 18 [2].

Clinical manifestations of MS are characterized by a high inter-individual variability [3] and by broad range of symptoms that reflect lesion burden, location, and degree of injury.

Lesions can be identified through Magnetic Resonance Imaging (MRI), however, it is essential to consider the neural plasticity mechanism in this context. Typical manifestations may include visual symptoms (like visual loss or double vision), vestibular symptoms (vertigo, gait imbalance), bulbar dysfunction, motor symptoms (weakness, tremor), cognitive and psychiatric symptoms like memory impairment, anxiety, depression. MS can manifest as either a progressive course with a gradual worsening of a patient's condition, or relapses, that are acute episodes of the aforementioned symptoms that can occur sporadically with a stability between episodes.

The two pathological processes seen in MS patients are limited to the primary CNS and consist in focal inflammation, resulting in plaques and injury to Blood-Brain Barrier (BBB), and neurodegeneration with microscopic injuries involving axons, neurons, and synapse [2].

Lesions from focal inflammation center around small veins and show sharp margins; during active inflammation, BBB disruption corresponds to gadolinium enhancement on T2-weighted MRI images, and when inflammation subsides an astrocytic scar remains. Microscopically, innate immune cells (monocytes and macrophages) stimulate T-cell migration across the BBB, resulting in an injury of BBB and in an entry of systemic immune cells [4].

Diagnosis of MS involves a comprehensive assessment encompassing medical history, physical examination, MRI, evoked potentials, as well as Cerebrospinal Fluid (CSF) and blood studies [2]. It can be made if it fulfills any of the “McDonald Criteria” that are updated over the years (the last update dates back to 2017). In general, diagnosis can be made with evidence of two or more relapses, for example through objective clinical evidence of two or more White and Gray Matter (WM, GM) lesions that demonstrate inflammation with mixed pathology of neuro-

axonal damage and demyelination. Since the first definition of diagnostic criteria, it has emerged the need to identify lesions Dissemination In Space (DIS, which is the development of lesions in distinct CNS areas) and Dissemination In Time (DIT, the development of new CNS lesions over time) [3]. MRI is the most reliable and used diagnostic tool for MS [1] and has a role in establishing DIS of T2 hyperintense lesions, if they are found in at least two of the four typical regions: spinal cord, infratentorial, juxtacortical, periventricular. DIT instead, according to the 2010 McDonald criteria for MS diagnosis, can be established when new lesions appear in an MRI follow-up scan, and from 2017, in some cases also CSF-specific oligoclonal bands allow a diagnosis of MS [5].

Unlike other diseases, there is no specific serologic diagnostic test for MS.

MS can have different disease courses that group in the following categories (Figure 1):

- Relapsing-remitting (RRMS): 70-80% of MS patients demonstrate initially a RR course, presenting symptoms consistent with MS that develop over days to weeks and last 24 to 48 hours, with stability between episodes.
- Primary progressive (PPMS): 15-20% of patients show a gradual deterioration from the onset, with an absence of relapses.
- Secondary progressive (SPMS): after an RR initial course of 10-20 years, patients demonstrate a more gradual deterioration.
- Progressive relapsing (PRMS): 5% of patients demonstrate a gradual deterioration with superimposed relapses.

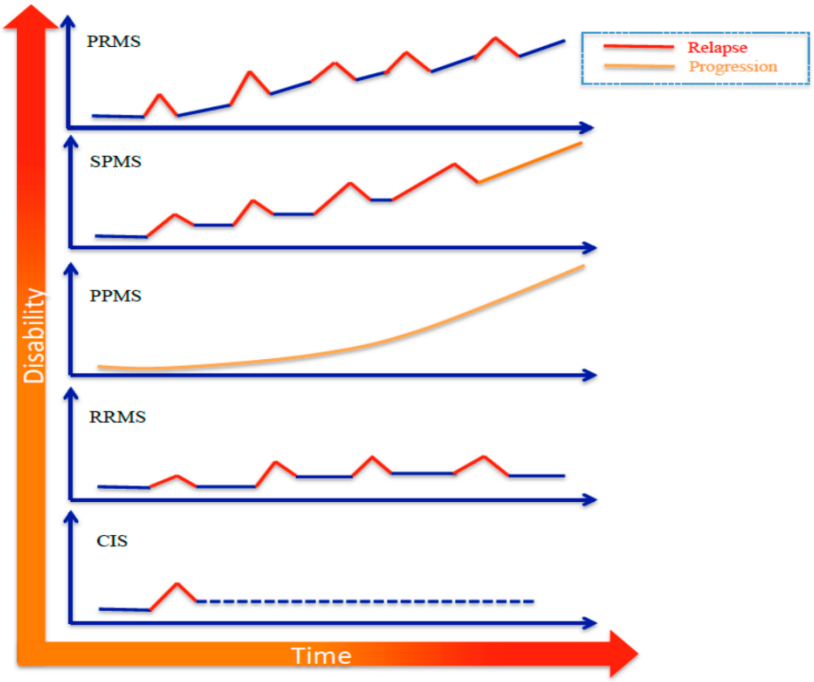


Figure 1 Possible MS disease courses. PR (progressive relapsing) MS is characterized by gradual deterioration with superimposed relapses; SP (secondary progressive) MS shows a gradual deterioration after an initial relapsing course; PP (primary progressive) MS shows only a progressive course from the beginning; RR (relapsing remitting) MS demonstrate relapses with stability between them; CIS (clinically isolated syndrome), is a single episode of demyelination. [3]

It is included in the spectrum of MS also clinically isolated syndrome (CIS), a single episode of inflammatory CNS demyelination [2]. If an MRI exam highlights a past episode, a MS diagnosis can be made.

The disability in patients with MS is quantified by the Expanded Disability Status Scale (EDSS), that is used to monitor disease progression but is quite insensitive [1].

1.2 Cerebrospinal Fluid classic markers

Even though etiology and specific pathogenesis of MS are still unclear, it is certain that the immune system plays an important role [3]. Investigation of CSF in the diagnostic work-up in suspected MS patients has regained attention in the latest version of the diagnostic McDonald criteria due to its good diagnostic accuracy [6]; CSF biomarkers indeed reflect accurately the inflammatory profile of the CNS and are more sensitive compared to clinical or MRI assessments (Figure 2).

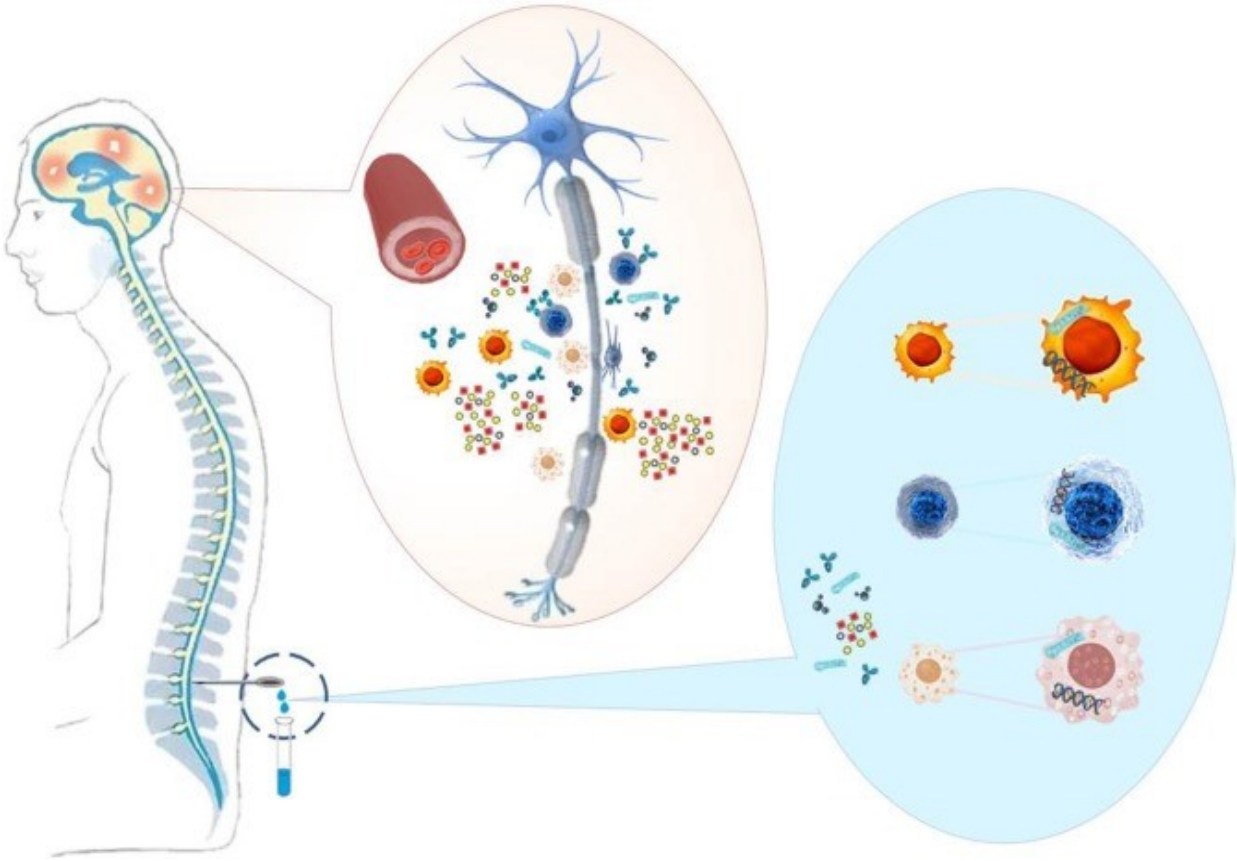


Figure 2 Immune cells, pathological antibodies, adhesion molecules, cytokines, chemokines, and nuclear acids, that reflect CNS inflammation, are present in patients CSF and can serve as biomarkers to support MS diagnosis and monitoring. [6]

Typically, in MS the CSF is characterized by elevated protein and myelin basic protein, leukocytes (usually mononuclear cells), increased total IgG, free kappa light chains (that contribute to the assembly of immunoglobulins), and oligoclonal bands [2].

If the immune response by T cells and plasma cells is altered, then the intrathecal synthesis of immunoglobulins, in particular the most abundant IgG class, will cause CNS inflammation, loss of myelin, and oligodendroglia death [3]. Several studies have also documented the presence of B cells and related cytokines, like CXCL13, C-X-C motif chemokine ligand 13 specific for B cells, and BAFF, a B cell-activating factor of the tumor necrosis factor family.

The hallmark of MS-specific changes in CSF is the detection of IgG Oligoclonal Bands (OCB) by Isoelectric Focusing (IEF) followed by IgG specific immunoblotting or immunofixation. This qualitative analysis aims to compare CSF IgG and serum IgG, explaining whether these immunoglobulins come from intrathecal synthesis or through BBB; if CSF and serum have identical bands, there is no pathological immunoglobulin synthesis in CNS because CSF bands originate in systemic circulation [6]. The examination of CSF also includes the quantitative measurement of Ig and albumin in order to get the IgG index, that is defined as follows:

$$IgG\ index = \frac{CSF\ IgG / serum\ IgG}{CSF\ albumin / serum\ albumin}$$

An IgG index > 0.7 contributes to MS diagnosis.

It is well known though that the detection of specific oligoclonal profile in CSF, and not in serum, and IgG index are thought to indicate chronic immune activation in the CNS and, therefore, are not exclusively found in MS [6]. Plus, they are not ideal for predicting relapse and progression and, regarding OCB, the protocol is time consuming, and the interpretation is operator dependent.

1.3 Choroid Plexus and its involvement in Multiple Sclerosis

The Choroid Plexus (ChP) is a vascular tissue placed inside the brain ventricular system, particularly in laterals, third and fourth ventricles, and is the main source of CSF secretion [7], that reaches a rate of 0.3-0.4 ml/min [8] . They consist in a highly vascularized network surrounded by a single layer of ciliated cuboidal epithelial cells, with a loose connective stroma separating them. ChP epithelium, richly endowed with mitochondria because of the high energy demand for transepithelial transport [9], has a general barrier function [10] between blood and CSF (BCSFB) together with the arachnoid membrane and the circumventricular organs [11]. This is thanks to the linking tight junctions of ChP epithelial cells that create a physical barrier between blood and CSF [7], but also thanks to a wide variety of transporter proteins involved in entry of molecules to CSF or removal of waste products. Thus, the ChPs regulate inflammatory cell transit between these two compartments, acting as a key neuroimmune checkpoint within the CNS [8] and having a potential involvement in neurological diseases [12] with a neuro-inflammatory component such as MS. The immune cell trafficking through ChPs [13] plays a crucial role in MS: there is prove of initial access of pathogenic T-helper 17 lymphocytes into CNS through ChPs (chemokine receptor 6 and C-C motif chemokine ligand 20 axis [14]) in experimental autoimmune encephalomyelitis (EAE) MS models [15], that could precede parenchymal inflammatory infiltrates and demyelinating WM lesions [16]; along with EAE models, also nonimmune mediated models (cuprizone intoxication) of demyelination showed lymphocytic and innate cell infiltration at the acute stage, associated with ChP enlargement on MRI [17].

It has been described that on MRI T1- weighted acquisitions ChPs are enlarged in patients with MS [16], raising interest in these structures as potential novel MRI biomarker of

neuroinflammation. From a histopathologic point of view, enlargement and inflammation of the choroid plexus are due to edema that presents high concentration of T lymphocytes, dendritic cells and activated macrophages, but also of vascular adhesion protein 1 that indicates active recruitment of peripheral inflammatory cells [18].

1.4 Aim of this study

Diagnosis of MS is still hard to reach and time consuming due to the high variability of the disease manifestation and the specific mechanisms underlying this disease are still unclear [3]. A hypothesis is finding a biomarker that is ideally non-invasive, safe, accurate, reproducible, cost-effective, and easily detectable in patients. Research though is moving toward a more promising approach that exploits a combination of different biomarkers to diagnose MS and to have a complete and personalized description of the disease. The aim of this thesis is to study the correlation between the Choroid Plexus Volume (ChPV) and biomarkers derived from the CSF in a dataset of PMS patients. As a matter of fact, the relationship between these variables might be relevant to identify a corresponding biological relationship between an enlarged ChP and the concentration of some CSF variables. The ChPV is a promising possible biomarker of MS that has raised increasing interest for its relationship with brain inflammation [17] and its capability to distinguish patients from healthy subjects [16]. The CSF molecules instead are representative of a broad range of MS manifestations, like inflammation and axonal damage [2], and it is still partially unknown how their modifications are involved in the disease and how they could contribute to its diagnosis and monitoring.

To reach this main objective, firstly, a recent toolbox based on Deep Neural Networks (ASCHOPLEX) was applied to the T1-w MRI scans of 169 PMS patients from two acquisition sites to obtain ChP segmentations; the first step was then to validate the algorithm on this new dataset. The segmentations were checked in the MNI standard space to get a common space reference and compare the subjects, and through this technique it was detected an erroneous inclusion of hippocampus in some subjects that was more emphasized in a dataset than the other. So, ChPs are corrected by removing voxels belonging to hippocampus and then the ChPV was computed. After that ChPV was normalized for the Total Intracranial Volume (TIV) to use a more robust measure not affected by brain dimension; the final normalized volumes (NChPV) have been obtained harmonizing the volumes to remove scanner-variability.

At the same time, a statistical analysis of the CSF derived biomarkers was carried out. The objective was to have an overview of the variables and to process them for the study of the relationship with the ChPV. After studying the relationships between them, Lasso Regression and Stepwise regression have been used to select the most relevant molecules to predict NChPV. Finally, these molecules are used as predictors in linear regression models.

The main objective of finding a link between the ChPV and the CSF variables is translated in this thesis as follows. Given a subset of molecules, selected from previous studies for their relationship to MS (hypothesis-driven selection) the objective is to observe if data-driven methods (Lasso and Stepwise regression) are able to identify a similar significant subset from the entire dataset.

2. STATE-OF-THE-ART OF THE RELATIONSHIP BETWEEN CHOROID PLEXUS VOLUME AND CEREBROSPINAL FLUID DERIVED BIOMARKERS IN MULTIPLE SCLEROSIS

In the last few years ChPV is drawing attention as a potential new biomarker for brain inflammatory diseases, and the same goes for a wide range of molecule concentrations from CSF [1]. In this section there will be at first an overview of the state-of-the-art regarding the ChPV, its automatic segmentation, and its relationship with MS inflammation process. Finally, the last section is dedicated to the latest findings about potential CSF biomarkers in patients with MS.

2.1 Choroid Plexus volume and its role in Multiple Sclerosis: state-of-the-art

In the last few years, some studies confirmed an enlargement and inflammation of the ChP in patients with MS compared to HCs. For example, Ricigliano et al. [16] assessed ChPs volumetric and inflammatory changes in a secondary analysis of 97 MS patients (61 patients with RRMS 36 patients PMS and 44 HCs (Figure 3). MS subjects were assessed with the EDSS and the number of relapses in the two years before and after inclusion were collected. ChP was manually segmented on 3D T1- weighted magnetization-prepared rapid gradient-echo (T1-w MPRAGE) MRI images by consensus of trained neurologists using software ITK-SNAP [19]. The ChPV estimate was normalized for TIV. The study also showed mild correlation between manual and Freesurfer (FS) [20] segmentation. It was discovered that in MS patients ChPV was 35% increased than in HC. Adjusting also for ventricular volume, ChPV was enlarged in patients than in HCs. Considering MS subgroups, only RRMS patients showed augmented ChP than HCs, maybe due to the low statistical power of PMS given the smaller sample size. The study also found that enlarged ChP is associated with reduced brain volume, relapses and brain inflammation, which has been assessed in 37 patients and 19 HCs using ^{18}F-DPA-714 (translocator protein fluorine) Positron Emission Tomography (PET) scans. Finally, enlarged

ChP was associated also with T2-weighted WM volume in the whole MS group. They concluded that the enlarged ChPV may be related to an inflammatory component that could peak in the relapsing-remitting phase and could subsequently become less pronounced at later progressive stages.

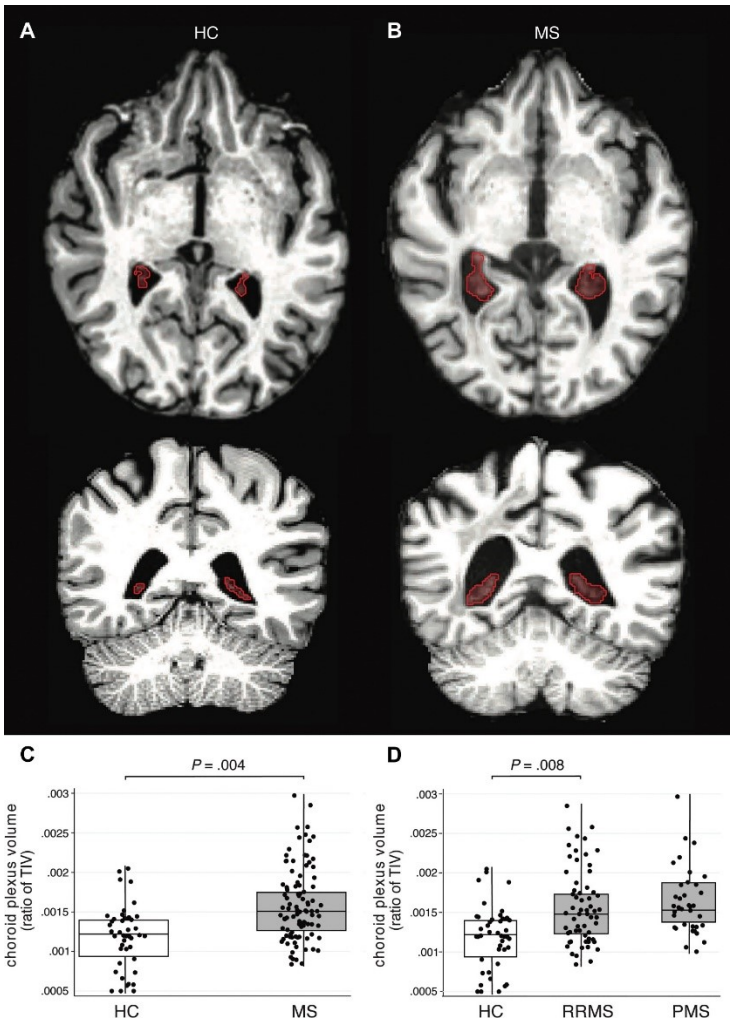


Figure 3 3D T1w MPRAGE MRI scans show segmentation of left and right Choroid Plexus (ChP) in axial (top) and coronal (bottom) planes in Healthy Controls (HC) participant (A) and in MS patient (B). (C, D) Box plots show higher ChP volume in HCs versus MS patients (C) and in HCs versus Relapsing-Remitting (RRMS) and Progressive PMS patients (D) [16].

In a later study [21] , it was demonstrated that a cohort of 27 patients with presymptomatic MS, referred to as radiologically isolated syndrome (RIS), showed ChPV 32% increased than HCs. At the same time, it was highlighted that in patients with a CSF-positive pattern, defined as high IgG index or positive oligoclonal bands, ChPV was comparable to that of CSF-negative pattern patients. Finally, in postmortem studies it was found an excess of CD163, an acute phase-regulated transmembrane protein on monocytes, in MS patients. These results suggest a possible involvement of ChPs at the early subclinical phase of MS.

Muller et al. [22], aimed to compare ChPV in MS and neuromyelitis optica spectrum disorder (NMOSD). To reach this, they used a Multi-Dimensional Gated Recurrent Units deep learning algorithm to segment the ChP in lateral ventricles on 3D T1-weighted MRI MPRAGE sequences, then reviewed by experienced raters and manually corrected. The ChP maps of the 125 training manual segmentations were quality checked by an expert radiologist and used as ground truth for the algorithm; TIV and lateral ventricle volume were measured using FS [20] on white matter lesion filled T1-w MPRAGE sequences. They studied 180 patients with MS and 98 with NMOSD and used 94 healthy individuals and 47 patients with migraine (non-autoimmune inflammatory CNS disease) as controls. Among the other results, ChPV expressed as a ratio of TIV was 21.4% higher than HCs ($12.13 \times 10^{-4} \pm 2.75 \times 10^{-4}$ vs $9.99 \times 10^{-4} \pm 2.51 \times 10^{-4}$), and they found a strong correlation between ChPV and T2-w lesion burden.
This difference between the two studies results has been hypothesized to be due to the inclusion of less active MS patients in the second study: they had 33% of RRMS patients with a relapse in the 2 years before the MRI (80% in the first study) and only 2 of 180 subjects had gadolinium enhancement MRI scans (32% in the first study); this differences show that ChPV is associated with both clinical disease activity, represented by MS stage and relapses rate, and MRI disease activity, represented by lesion burden and gadolinium enhancement.

Other works applied measurements of ChP to find a relationship between the ChPV and other characteristics of MS disease. A recent research study [23] aimed to assess the relationship between measurements of ChP inflammation extracted from pseudo T2 maps of the whole brain, hypothesizing that it could be reflected by ChPV, and it investigates their association with clinical disability progression in MS. The ChP segmentation has been initially obtained from lesion-filled T1-w images acquired at baseline and after 5.5 years they have been depicted using FS, refined with a Gaussian Mixture Model method (GMM) [24] and manually corrected, and finally normalized multiplying by SIENAX scaling factor [25]. SIENAX is a package that estimates the total brain tissue volume, normalized by skull size, finding a scaling factor between the image and the standard space, and multiplying it to the brain volume to reduce head-size variability. They also calculated pseudo-T2 value of the ChP for quantifying inflammation; this index was found to be significantly greater in patients with MS and highly associated with future disease progression. The results also confirmed that patients presented increased ChPV without considering brain volume or ventricular volume, and no difference has

been found in ChPV between RRMS and PMS patients, suggesting that ChP enlargement could be an early phenomenon of the disease.

The relationship between ChP characteristics and neuroinflammation has been investigated also by Fleischer et al. [17]. After the extraction of ChP segmentation from T1-w images using FS, they found enlarged ChPV in 330 MS patients compared to HC (Figure 4). Moreover, in MS patients larger ChP was associated with the severity of clinical disability (EDSS score). Within the MS cohort it was highlighted that patients with evidence of disease activity (EDA) displayed larger ChPV at baseline compared to patients with no evidence of disease activity (NEDA). It was hypothesized the ChPV as a strong prognostic factor for EDSS scores applying Structural Equation Model (SEM), in contrast to T1 contrast-enhancing lesions and enlarging T2 lesions that didn't reach significance. In a subset of patients for whom albumin content in CSF and serum was available, associations between ChPV with CSF albumin and albumin quotient (plasma to CSF, indicator of blood-brain barrier dysfunction) were attested, but not for serum albumin. For 42 patients on DMF, a disease modifying therapy based on dimethyl fumarate, treatment immune cells subsets were available (CD4+, CD8+, CD19+, CD56+), but no association with ChPV was found. A limitation of this study was the delineation of the ChP using FS, which may not capture its entire structure.

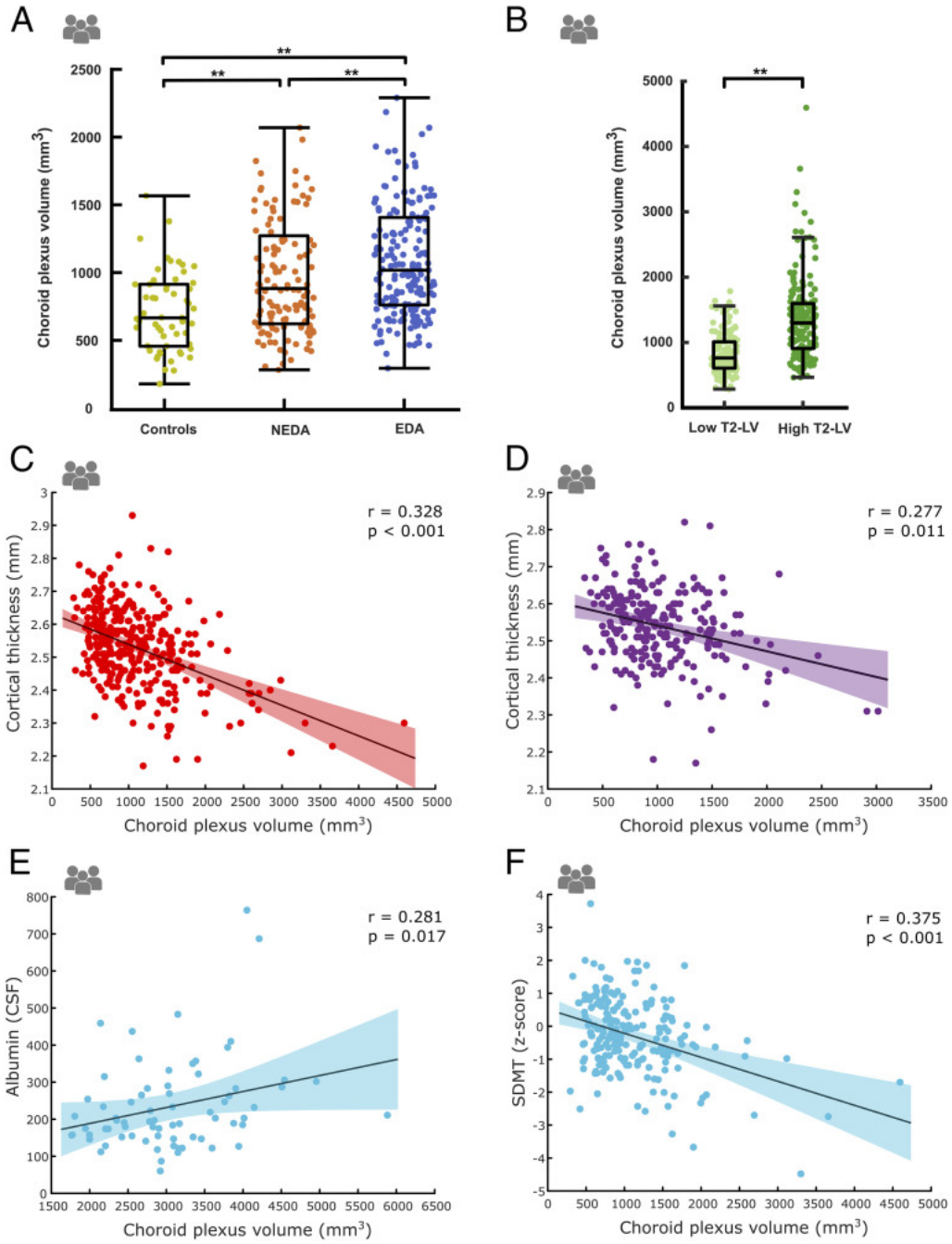


Figure 4 (A): ChP volumetric differences depicting enlarged volume in patients with MS with EDA compared to NEDA, each compared to healthy controls. (B) ChP volume comparison between patients with a high T2 lesion load in MRI in comparison to patients with low lesion load. Scatter plots that show association between ChP volume and cortical thickness in the study cohort (C) and in replication cohort (D), and scatter plots depicting association with albumin in the CSF (E) and with cognitive performance (measured with Symbol Digit Modalities Test) in the study cohort [17].

Slow-burning inflammatory and demyelination at the edge of chronic MS lesions has been implicated as major factors that contribute to disease evolution, including neurodegeneration, brain atrophy and disability worsening. A study [26] investigated the association between ChPV and subsequent lesion expansion in 49 patients with RRMS. ChP was semi-automatically segmented on T1-w images using JIM 9 software (https://www.xinapse.com/j-im-9-software/)

and manually adjusted by a trained analyst. It was confirmed another time that ChPV/TIV ratio was significantly larger in RRMS patients than the 40 HCs (95.8 ± 32.1 vs. 75.2 ± 12.1), and that this ratio remains stable also during 4 years of follow up. In addition, it was observed a strong association between ChPV/TIV ratio and subsequent rate of lesion expansion, hypothesizing a causal relationship between them. The study concluded that the size of ChP has strong correlations with subsequent expansion of chronic periventricular MS lesions, associated axonal loss within established lesions and central brain atrophy.

2.2 Automatic segmentation of Choroid Plexus

Even though the gold standard technique is still manual segmentation from T1-weigthed MRI scans, this complex procedure is time-consuming and affected by inter- and intra-operator variability [27].

The increasing interest in the ChP segmentation task has raised the need to find a completely automatic method to segment the ChP.

The automatic methods proposed in literature are two: FS and GMM. FS is the most popular freely available software for automatic brain segmentation, but it doesn't perform well on ChP compared to manual segmentation concerning volume correlation [20]. GMM method pipeline is described as follows: 1) create ChP, CSF and brain ventricles wall masks with FS; 2) apply a Bayesian GMM with two components; 3) obtain two clusters of voxels: CSF (low intensity, near zero), ChP + ventricles wall (high intensity, near one); 4) take ChP + ventricles wall cluster and apply 3D Susan smoothing; 5) both CSF and ventricles wall voxels are set near zero, while ChP ones to one; 6) apply a Bayesian GMM with three components and classify the voxels of the cluster with the higher mean intensity value as ChP. GMM is more robust and outperforms FS in terms of Dice Coefficient [24]. A recent improvement of GMM [28] showed moderate-high segmentation accuracy on co-registered 3D T1-w+FLAIR (Fluid Attenuated Inversion Recovery) acquisitions (Mean Dice = 0.66). Automatic algorithms proposed in literature have poor results compared to the ground truth manual segmentation, consequently, new automatic segmentation tools are required to accurately segment the ChP to allow an accurate volume estimate in large cohorts. With the exception of Storelli et al. [28], recent research has mainly focused on deep learning-based methodologies.

Zhao et al. [29] developed and optimized a 3D U-net for ChP segmentation based on a standard U-net. The proposed method was trained on T1-w MRI images of 10 healthy subjects, and it consists of an encoding step and a decoding step. The model provides an accurate tool to study

ChP structure and showed about 25% higher accuracy in Dice and sensitivity scores compared to FS (FIG). The proposed 3D U-Net reaches a Dice score of 0.73 using a non-uniform patch-extraction, data augmentation and weighted binary cross-entropy, but it was trained on a small cohort that included only healthy subjects.

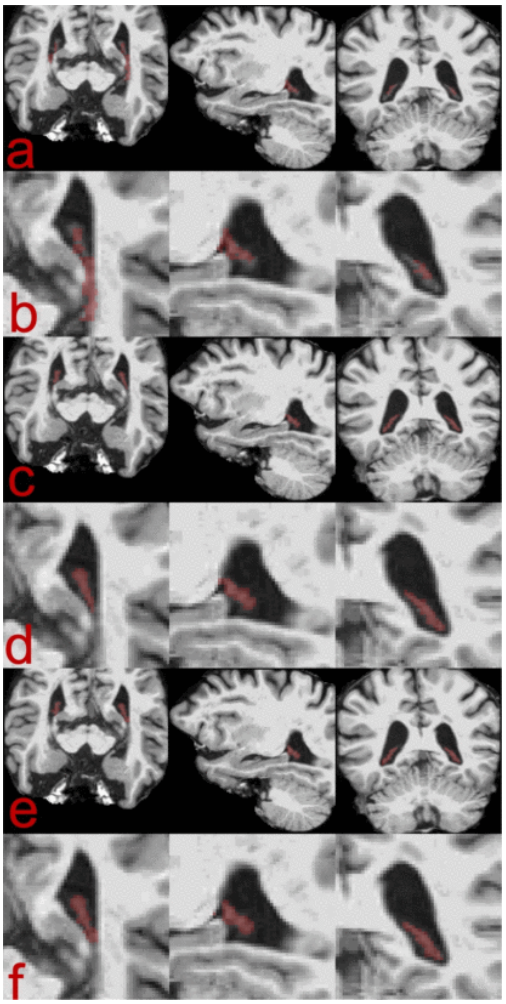


Figure 5 Choroid Plexus (ChP) segmentation results given by (a)(b) FreeSurfer (Dice = 0.58), (c)(d) the optimized 3D U-Net with preprocessed T1-w images (Dice = 0.73), and (e)(f) manual segmentation (ground truth).

Yazdan-Panah et al. [30] proposed a simple solution that requires minimal preprocessing steps and that consists in a 2-step 3D U-Net. The dataset was composed of a research dataset (44HC, 61 RRMS patients, 36 PMS patients) and a clinical dataset (pre-symptomatic MS, fulfilling diagnostic criteria for RIS). On both datasets, the ground truth was considered the manual segmentations of the ChP portion within the two lateral ventricles performed by a trained neurologist and corrected from a senior neurologist. On the second dataset, an independent segmentation by a trained neurologist has been performed. The preprocessing consisted in a

reorientation to canonical voxel orientation, resampling to have an isotropic 1 mm^3, padding to get a uniform image dimension and an intensity rescaling.

Step 1 of the method takes a down-sampled image and selects a maximum of 500 patches around voxels that have a probability higher than 80% to be ChP. Step 2 takes the patches as input to segment the ChP, and segmented patches are merged to create the final segmentation mask, using a Hann windowing to reduce edge-effects. The architecture, shown in Figure 6, uses group normalization and a leaky ReLu as activation function to avoid "dead ReLu" effects.

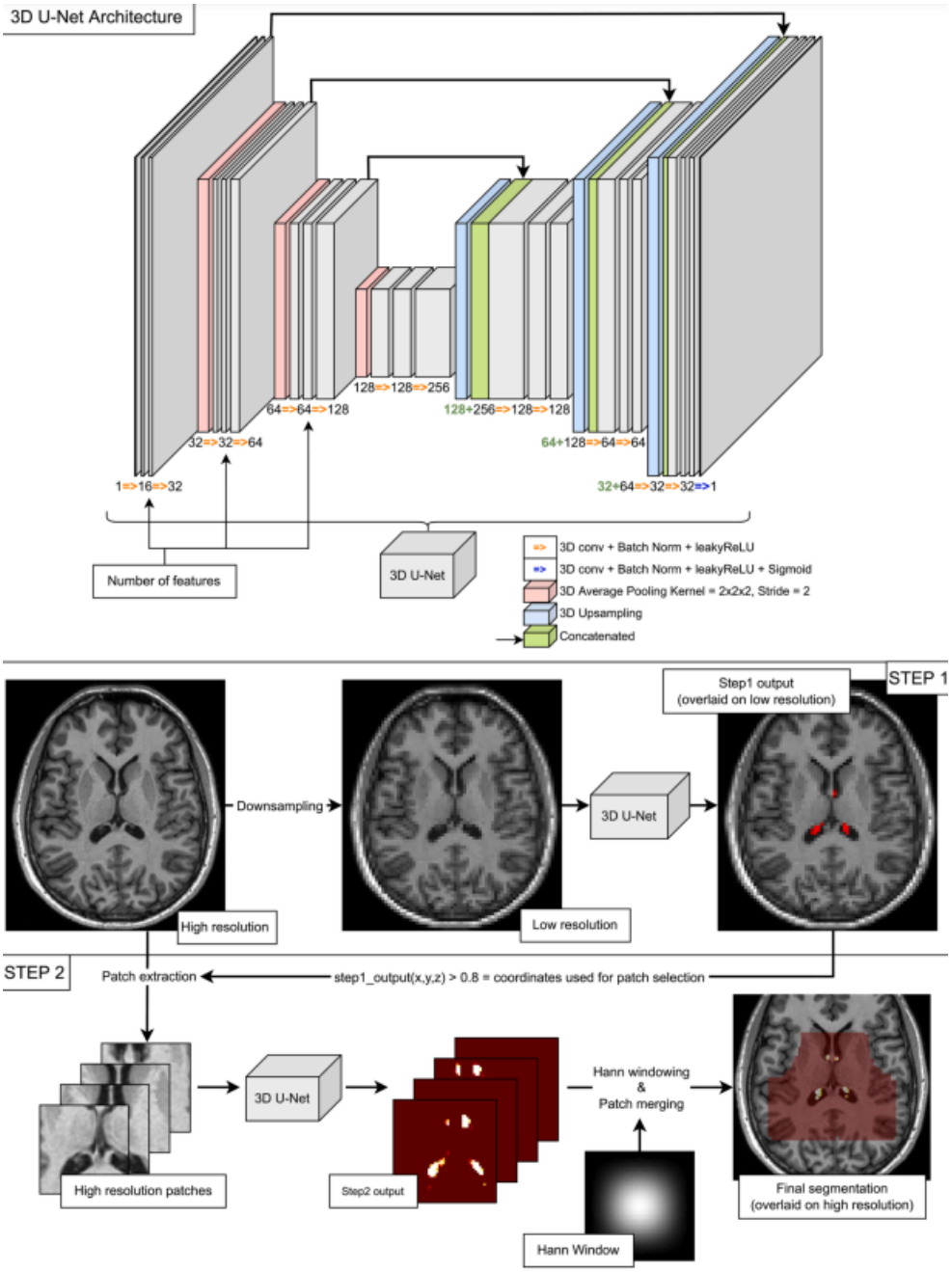


Figure 6 Top row: 3D U-net architecture. Bottom rows: two-step method. Step 1 down samples the initial scan and gets a low-resolution image, then selects maximum 500 patches that are likely to be Choroid Plexus (ChP). Step 2 takes the high-resolution selected patches as input to segment the ChP. Finally, segmented patches are merged, and a Hann windowing is applied to reduce edge effect. Step 1 output with threshold at 0.1 for visualization.

To evaluate the performances, Dice Coefficient, Recall, Precision, Volume Error Rate, and absolute Volume Error Rate have been used; performances in the second dataset were in general worse. For dataset 1, Dice was around 0.7, both with a 2-step and a 1-step architecture and both with and without data augmentation, considerably better performances compared to FS. On dataset 2, Dice coefficients were around 0.61 – 0.67 considering annotator 1, and between 0.56 -0.59 for annotator 2. This method was demonstrated to be effective on heterogeneous clinical data with both controls and patients. The obtained results are better than FS, but the performance improvement is not optimal.

Eisma et al. [31] expanded on the method proposed by Zhao et al. by including 3D T2-weighted and 2D T2-weighted-FLAIR MRI as well as T1-w images. The model based on 3D U-Net was trained on the three different datasets composed by the three different MRI sequences of 50 adult participants with and without clinically diagnosed neurodegenerative diseases; all images were registered to the International Consortium for Brain Mapping-Montreal Neurological Institute (ICBM-MNI) 152 T1-w template to reduce morphological variability. The encoding and decoding steps were composed of three blocks each composed of two 3D convolutional layers with batch normalization, followed by a rectified linear unit. Looking at the T1-w dataset, results show a significant difference between performances of 3D U-Net (Dice Coefficient = 0.72) and FS (Dice Coefficient = 0.19).

Another model called “Axial-MLP”, based on an assembly of Axial multi-layer perceptrons (MLPs) for ChP segmentation has been introduced by Schmidt-Mengin et al [27]. They trained the model in a cross-validation fashion on 91 subjects from the same MS dataset as Ricigliano et al. [16] , while the remaining 50 subjects have been tested.

Axial-MLP consists of dividing the input image into non-overlapping patches and down sampling it to a multiple of the patch size; a fully connected layer was applied on the resulting tensor to expand it along the channel axis, consequently passing it through 6 fully connected layers,each one applied on a specific dimension. The results are passed through a leaky ReLu activation function, summed, and normalized. Finally, the patches are rearranged into an image (Figure 7).

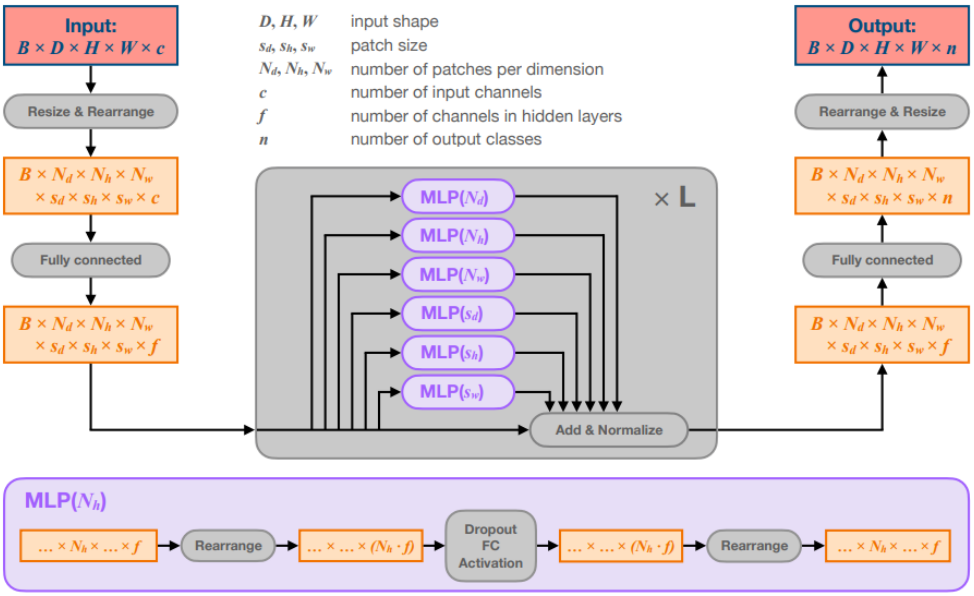


Figure 7 Structure of the proposed Axial-MLP for ChP segmentation. The input image is divided in $N_d \times N_h \times N_w$ patches of dimension (s_d, s_h, s_w), then passed through a fully connected layer to expand the image in channel axis. Then it enters L layers consisting of 6 fully connected layers (MLP, multi-layer perceptrons) that correspond to the number the image dimensions, a after that a Leaky ReLu activation function. After the sum and normalization of the output, it is passed through a fully connected layer to transform the channel dimension equal to the number of outputs class and resized to get the output image [27].

The results confirm that FS and FastSurfer, a faster version of FS, give poor results when applied to ChP segmentation. Their method gave results comparable to other standard deep learning methods based on 3D U-net and nnU-net (which gave slightly best results, 0.71 vs 0.72 for Dice coefficient, but still statistically significantly different). The main disadvantage of this approach is its complex structure.

Every proposed approach has both pros and cons, but the balance between the two is highly unstable. This has arised the need for a new method based on the ensemble by majority voting of the predictions derived by a pool of selected DNNs trained algorithms for the ChP segmentation task. The new approach is called ASCHOPLEX and it is described in the *Materials and Methods* chapter.

2.3 Cerebrospinal Fluid molecules: potential biomarkers for neuroinflammation

The CSF analysis has a great potential as a source of emerging biomarkers for neurogenerative diseases because of the opportunity to directly evaluate specific CNS inflammatory processes and changes in the immunologic pattern due to the progression of the disease [3].

The search of biomarkers aims at covering the heterogeneity of mechanisms involved in the disease. Currently, an axonal injury marker known as Neurofilament Light (NFL) is one of the most promising CSF biomarkers [6]. NFL is a cytoskeletal protein released from damaged axons into the CSF and the bloodstream. Its concentration reflects disease activity and progression, not limited to MS patients alone. Additionally, it could be used to monitor therapeutic efficiency. Efficiency treated patients with agents such as natalizumab [32] exhibit a normalization of CSF NFL levels, reducing them to the levels observed in HC [6]. Studies have found that NFL levels of MS patients are increased during active relapse and acute relapse, and that are correlated with CD4+ T lymphocytes, implicated in inflammation and progression from RR to SPMS [1]. Magliozzi et al. [33] found elevated NFL in MS patients with high meningeal inflammation compared to those with lower levels of meningeal inflammation. In both these groups NFL levels were significantly higher than those in HCs. Moreover, NFL positively correlated with the number and the volume of cortical lesions. Other proteins described in the following lines exhibited increased levels in MS patients with high inflammation compared to controls; CXCL13, CXCL12, which are chemokines involved in regulation of B cell migration in lymphoid tissue [34] ; CCL25, a C-C motif chemokine ligand also known as TECK (Thymus-Expressed Chemokine); tumor necrosis factor (TNF), involved in inflammatory process and its regulation [35], and its receptor sTNF-R1; interferon-γ (IFNγ), secreted from T-cells and belonging to immune system [36], IFN-α2, secreted both from T-cells and B-cells, and IFN-λ2; IL6 (interleukin-6), IL8, IL10 all linked linked to proinflammatory patterns; BAFF; APRIL, a proliferation-inducing ligand [37]; LIGHT, that belongs to TNF family and stimulates T-cell proliferation; TWEAK, TNF-like weak inducer of apoptosis; GM-CSF, granulocyte-macrofage colony stimulating factor; matrix metalloproteinase 2 (MMP-2); pentraxin3, an acute-phase protein [38]; sCD163. They also developed a linear multivariate model with IFNγ, TNF, CXCL12, CXCL13, IL6, IL10, and LIGHT as independent variables, that were able to explain 89% variance of cortical lesion volume and 87% of cortical lesion number.

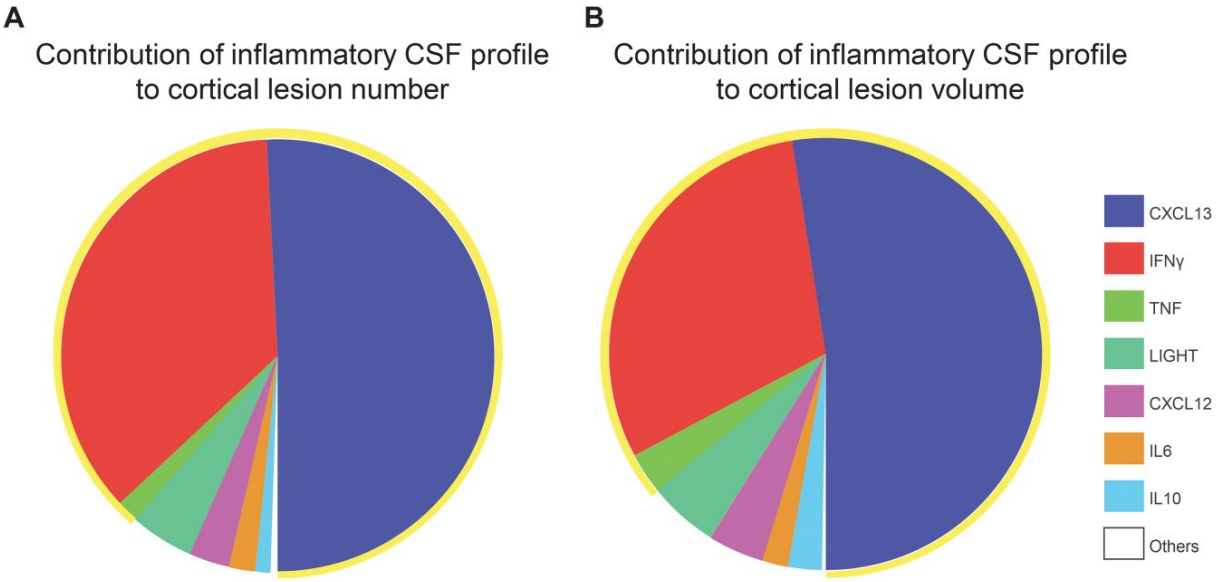


Figure 8 Contribution of variables in the linear multivariate models that aim at describing the variance of cortical lesion number and cortical lesion volume. The biggest contribution is given by CXCL13, IFNg and TNF, that together explain approximately 88% of cortical lesion number (A) and 85% of cortical lesion volume (B). [33]

In the realm of investigating potential biomarkers for immunomodulation and inflammation, several studies have documented the presence of B cells in WM lesions of MS subjects. These cells are a pivotal component of the adaptive humoral immune system, responsible for producing antigen-specific immunoglobulins. Notably, depletion of B cells has been shown to suppress inflammatory processes in RRMS and impede the progression into PMS. Additionally, researchers have found an elevated concentration of B cell-related cytokines, like CXCL13 and BAFF, as well as myelin-specific antibodies in the CSF [3]. Specifically, CXCL13 has been correlated with a worsened prognosis in RRMS and the conversion of CIS to MS. However, similar to NFL, its specificity is limited, as it might also be presented in patients with infection. Furthermore, other potential inflammation biomarkers reflecting disease type and differentiating underlying disease pathology and activity could be the products of Th (T helper) 1 and Th17 cells. These products include cytokines such as IL-17, interferon (IFN)-γ, and tumor necrosis factor (TNF)-α [1].

TNF, a major inflammatory mediator, can be produced by both immune cells and CNS cells. It plays a crucial role in immune system development, regulation of immune response and T-cell-mediated tissue injury.

In particular, Magliozzi et al. [39] conducted a study on CSF of 122 patients that aimed at evaluating the clinical and MRI correlates of TNF and its soluble receptors sTNF-R1, present in most tissues, and sTNF-R2 present in cells of the immune system. Their findings revealed significantly elevated levels of these molecules in MS patients compared to the control group

(comprising 36 patients with other neurological diseases). Notably, there was a greater overexpression of TNF-R1 in patients in comparison to controls, as well as in relation to the TNF-R1 and TNF-R2 ratio under physiological conditions (5:1). This ratio will be examined as an additional variable in this thesis. The researchers concluded that assessing the balance between TNF and its receptors, together with MRI data, could aid in identifying patients with a worse prognosis. Moreover, they suggested that modulating this balance could represent an important avenue for MS treatment.

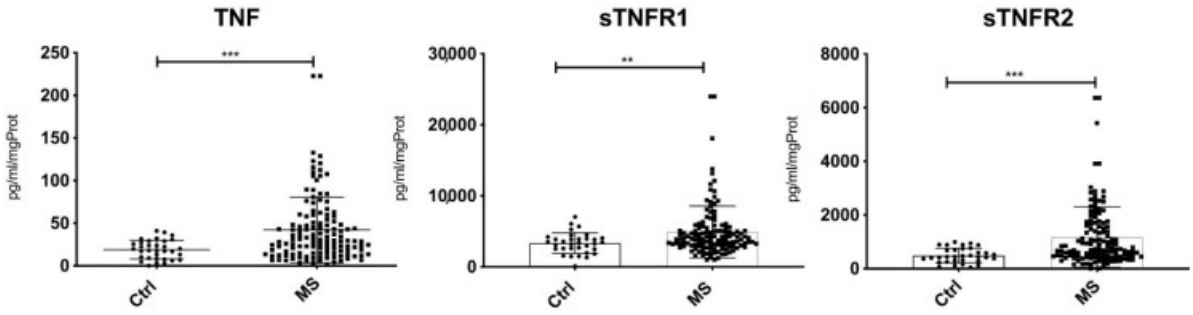


Figure 9 Levels of TNF, sTNF-R1 and sTNF-R2 in MS cohort vs the control group. The difference is significant in all cases (respectively, p<0.001, p<0.01, p<0.001) [39].

In conclusion, numerous potential alternative CSF biomarkers exist for clinical practice. Further research is imperative to identify the specific biomarkers or combinations thereof that can effectively diagnose and monitor MS, as well as predict disease progression and treatment-related side effects.

Previous studies have explored correlations among CSF biomarkers. Magliozzi et al. [40], in their investigation aiming to profile CSF and reflect cortical damage in early stages of MS, analyzed the levels of 69 inflammatory mediators in CSF samples from both patients and HCs. They employed a combination of immune-assay multiplex techniques and identified 29 proteins that were significantly elevated in MS patients compared to controls. Alongside these variables, they measured levels of NFL, soluble sCD14, Haptoglobin, free-hemoglobin, and fibrinogen. Through the analysis of these molecules, noteworthy findings emerged, including a positive correlation between NFL and sCD163, and a negative correlation between NFL and sCD14. Furthermore, sCD163 exhibited positive correlations with several inflammatory mediators such as MMP2, interleukin 10 (IL10), CXCL13, and CXCL12, and also with free-Hb, haptoglobin and fibrinogen (Figure 10).

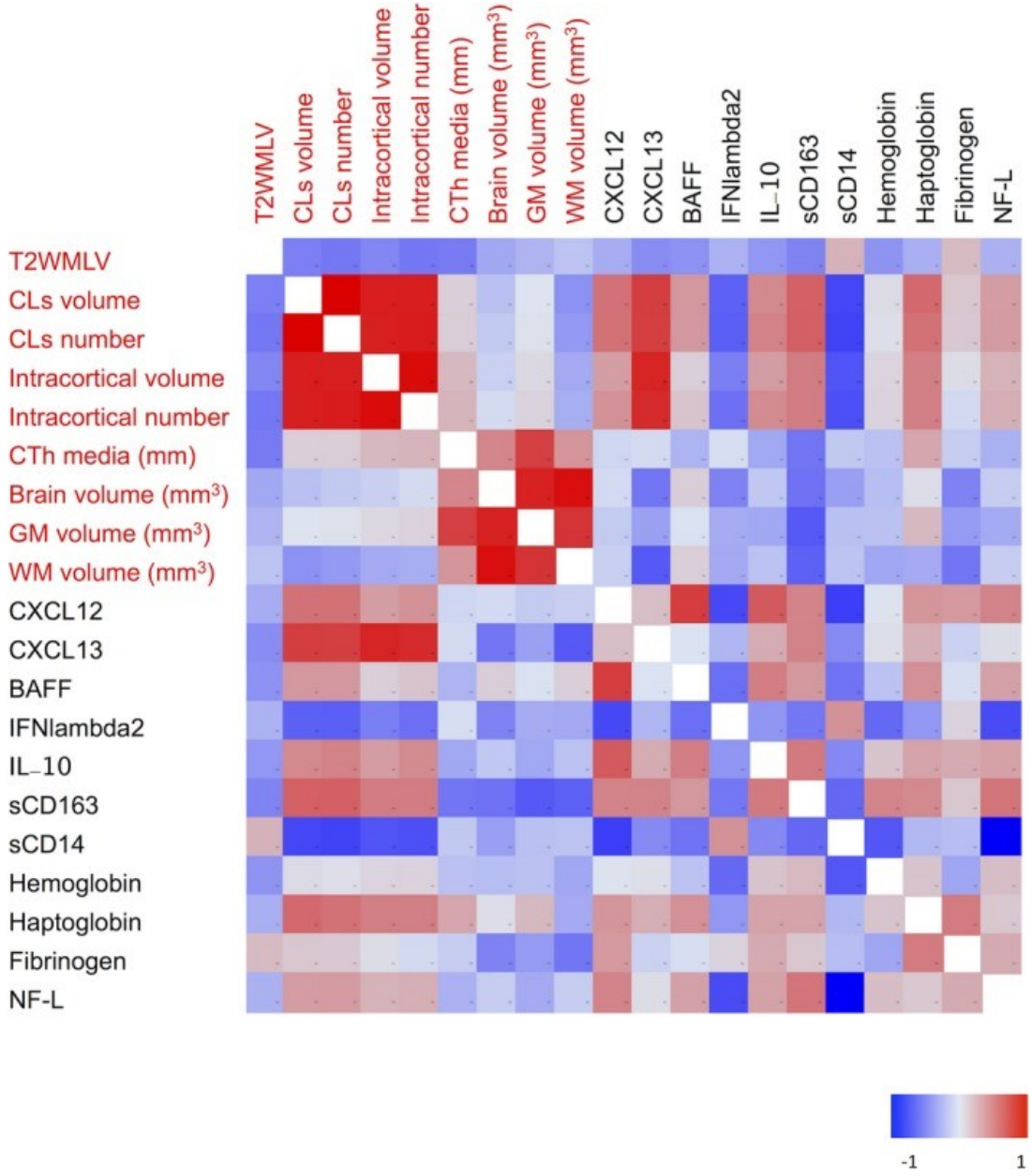


Figure 10 Correlation matrix (Pearson) between MRI parameters and CSF variables. The figure shows only significant correlations (positive in red). The most relevant are between sCD163 and MMP2, interleukin 10 (IL10), CXCL13, and CXCL12 and NFL, and a negative correlation between NFL and sCD14.

In another study, Magliozzi et al. [39] demonstrated the correlation between TNF, TNF-R1 and TNF-R2 and other significant molecules (Figure 11). Using Pearson correlation analysis to assess CSF inter-molecular concentrations, elevated levels of TNF were found to be significantly associated with increased concentrations of BAFF, IFN-γ, IL-1β, IL-10, IL-8, IL-16, CCL21, haptoglobin and fibrinogen. Furthermore, heightened levels of TNF-R1, which can be expressed in several cell types, were significantly associated with elevated concentrations of CXCL13, TWEAK, LIGHT, IL-35, osteopontin, pentraxin-3, sCD163 and chitinase-3-L1 (a molecule that plays a major role in tissue injury and repair, inflammation, and remodeling

responses [41]). increased levels of TNFR2, expressed exclusively in endothelial cells, certain immune cells and specific CNS cells, were significantly associated with high CSF levels of IFN-β, IFN-λ2, and sIL-6Rα.

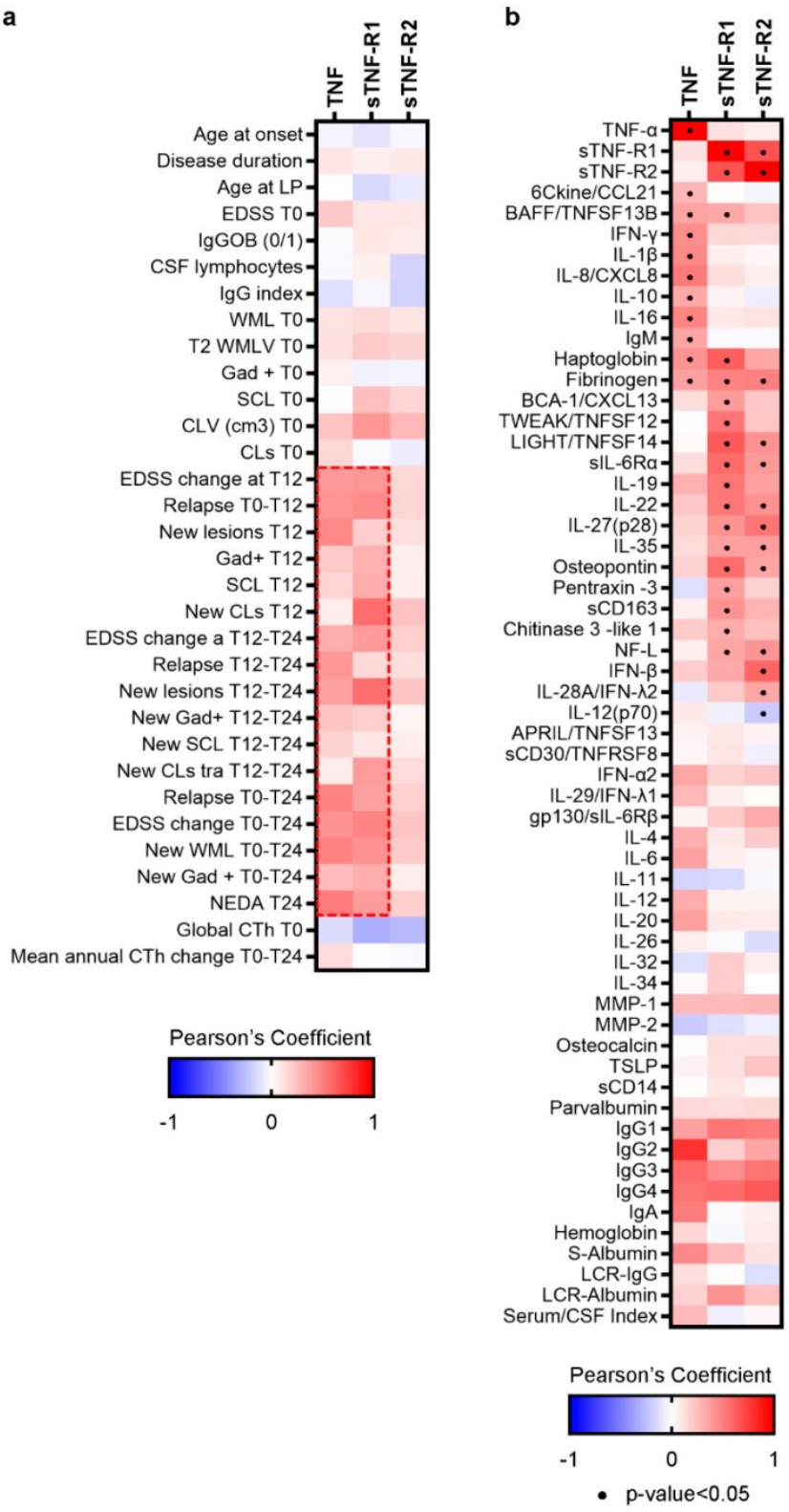


Figure 11 Correlation (Spearman) matrix between TNF and its receptors, (a) and clinical and MRI parameters, (b) and other CSF variables levels. Points indicate significant correlation (p-value < 0.5)

3. MATERIALS AND METHODS

This chapter is intended to describe the data that has been provided for this thesis and the methods applied to elaborate this information. So, after a short outline of the dataset, it is described the deep learning algorithm used to get the ChPV and the techniques used to evaluate and improve the segmentation quality. After that, the statistical analysis tools are presented with a final focus on methods used to assess the links between ChPV and the CSF biomarkers.

3.1 Materials

3.1.1 MRI Dataset

Dataset is composed of 169 subjects affected by PMS at an early stage that underwent both MRI and liquor (CSF) analysis. The dataset was provided by the Multiple Sclerosis Centre of the University of Verona and was acquired between May 2014 and December 2022. 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 three datasets.

A. Dataset 1 – It is composed of 131 subjects (age 39.2 ± 12 years). T1-w MRI images were acquired at Borgo Trento site 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 1x1x1 mm; SENSE acceleration factor: 2.5; TE/TR: 3.7/8.4 ms; FA: 9°; total acquisition time: 4min 50s. It will be referred to as “BT”.
B. Dataset 2 - It is composed of 34 patients (age 40.1 ± 12.3 years). T1-w MRI images were acquired at Borgo Roma site 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: 1x1x1 mm; compressed SENSE acceleration factor: 4; TE/TR: 3.7/8.4 ms; FA: 8°; total acquisition time: 3min 20s. It will be referred to as “BR”.
C. Dataset 3 - It is composed of 4 patients (age 33.8 ± 19.5 years). T1-w MRI images were acquired at Altavilla on a Philips Elition S with 32-channels head coil (Software version R5.7.2.1) with the same parameters of Dataset 2. For this reason, the 4 subjects will be included in the BR dataset.

75% of the subjects are female patients (n = 127, mean age 38.5 years), while men are n = 42, mean age 41.1 years. The mean age of the whole dataset at the date of the diagnosis is 39.1 ± 12.1 years.

3.1.2 Cerebrospinal Fluid Variables

The liquor analysis provided 130 variables. From these variables, 46 have been excluded by this study because they had equal or less than 30 samples. The definitive number of analyzed CSF variables is 84.
One subject has been excluded from the correlation study because it presented two MRI scans but three CSF acquisitions and it wasn't possible to go back to its clinical history, so the total number of subjects is 167.

Some previous studies introduced in section 2.3 are the starting point of the CSF biomarkers selection process. These studies found that some variables have a high predictive power when considering cortical lesion volume and cortical lesion number, suggesting that they are related to MS pathological activity.
The biomarkers in this pre-established subset (from now on it will be referred to as Subset1, shown in Table 1) are known to have some sort of connection to a MS mechanism or to another type of MS biomarker, such as MRI derived ones, and so it has been investigated more in depth in terms of correlation between each other and with ChPV.

MCP_4_CCL13_28	MIP_1delta_CCL15_66_	IL_35
IL12p70	APRIL_TNFSF13	IL_28A_IFNlambda2
IL_29_IFN_lambda1	Fractalkine_CX3CL1	FibrinogenNg_ml_mgProt
SCYB16_CXCL16_64_	Gro_alpha_CXCL1	CSF_Albumina_mg_L
EOTAXIN_CCL11	x6Ckine_CCL21	IFN_alfa2
MPIF_1_CCL23_37	TECK_CCL25	IL_16
ENA_78_CXCL5	MIF_35_	SDF_1____CXCL12
TNF	MIP_1alpha_CCL3_55_	LIGTH_TNFSF14
Gro_Beta_CXCL2	BAFF_TNFSF13B	Parvalbumin
MCP_3_CCL7	NFL	IFN_beta

IFNgamma	BCA_1_CXCL13	sCD163

Table 1 Biomarkers connected to multiple sclerosis mechanisms in previous studies (Subset1).

For the selection of the variables for linear models (Lasso and Stepwise regression, Section 3.7.2 and Section 3.7.3), it was required to have a dataset without missing values. To do so, 44 variables have been selected, to keep the 80% (133) of the subjects. These 44 molecules have been merged with other 11 molecules from Subset1 that have been excluded from the selection. These 55 molecules then form the Subset2.

3.2 ASCHOPLEX: a Deep Learning based approach for the automatic segmentation of Choroid Plexus

The ChP segmentations have been obtained using a novel automatic segmentation approach based on the ensemble of DNNs. The tool called ASCHOPLEX (Automatic Segmentation of CHOroid PLEXus) has been implemented in MONAI version 1.0.1 [42]. MONAI (Medical Open Network for Artificial Intelligence) is a set of PyTorch-based (version 1.13.1) open-source collaborative frameworks built for clinical collaboration in Medical Imaging. It aims to gather the best practices of AI development for healthcare researchers, providing reproducibility of experiments for comparison against state-of-the-art implementations. Notably, MONAI incorporates a streaming loading modality to enhance training efficiency, integration of implemented DNNs, as well as domain-specific metrics and loss functions.

Specifically, ASCHOPLEX is a modified version of Auto3DSeg [43] a MONAI application designed for 3D medical image segmentation. Auto3DSeg, as shown in Figure 12, adopts a structured framework that organizes algorithm folders for each selected DNN based on the MONAI bundle. A bundle, in this context, refers to a self-descriptive network integrating its architecture definition with metadata, including comprehensive data-related information and scripts for model training and inference. Auto3DSeg offers several advantages, including simplified network distribution, user-friendly operations, and result reproducibility.

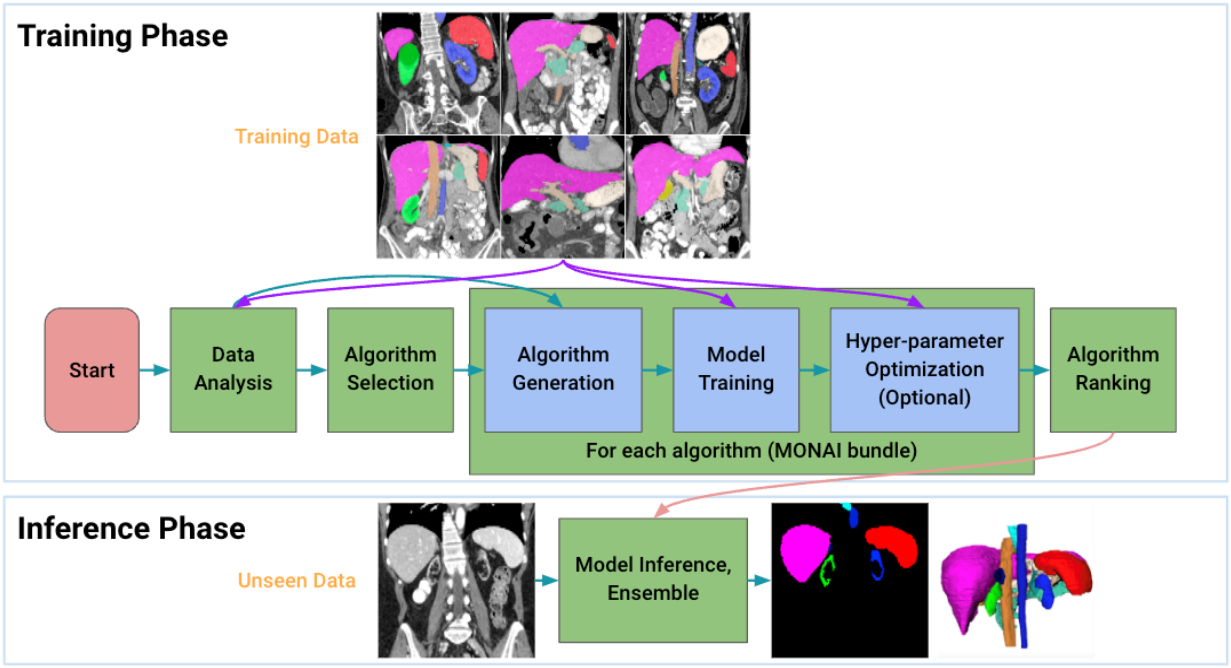


Figure 12 Auto3dSegm method (https://github.com/Project-MONAI/tutorials/tree/main/auto3dseg). The segmentation of training dataset starts with an analysis of global information of the dataset to generate the algorithm folders. For each algorithm in MONAI bundle the initiation of model training occurs, and the ensemble module selects the best models via algorithm ranking. Finally, the inference phase consists in using the ensemble to make predictions in unseen data.

Furthermore, Auto3DSeg introduces an ensemble module that evaluates trained algorithms based on validation accuracy from checkpoints. It selects the top N models and generates predictions using ensemble methods, allowing users to choose between mean and majority voting. This feature addresses the significant variability observed in output segmentations derived from various DNN models. Implementing an ensemble procedure becomes imperative to consolidate and integrate single model predictions, ensuring a more robust and reliable outcome.

ASCHOPLEX was trained on T1-w MRI derived from the same two scanners as the whole dataset of this study. Particularly, 61 RRMS subjects derived from Borgo Roma scanner, while 67 (24 HC, 43 RRMS) from Borgo Trento one. The 128 subjects were divided into a training set (92 subjects) and a testing set (36 subjects) balancing the number of HC, RRMS and the two scanners. Figure 13 shows the ASCHOPLEX implementation pipeline.

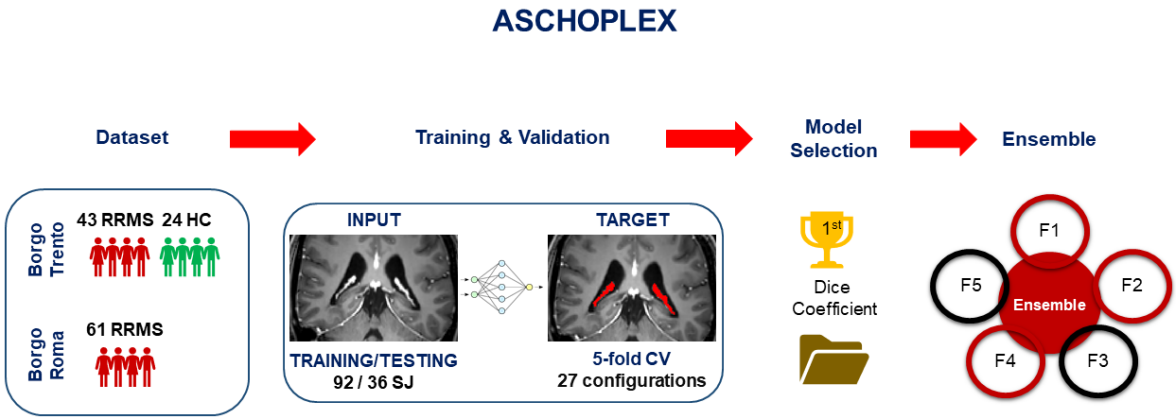


Figure 13 ASCHOPLEX implementation pipeline. 1) T1-w MRI data derived from two scanners were used as input (RRMS: Relapsing-Remitting Multiple Sclerosis; HC: Healthy Controls) 2) Training and validation of twenty-seven DNNs configurations using five-fold cross-validation training procedure 3) Selection 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 output of ASCHOPLEX.

Twenty-seven configurations of DNN were trained in a five-fold cross-validation fashion, Each configuration was a combination of the following architecture and hyperparameters: three possible architectures, which were 3D U-Net [44], DynUnet (that is MONAI implementation of nnU-net) [45], and UNETR [46]; three patch sizes (64×64x64, 96×96x96, and 128×128x128); three loss functions, named Generalized Dice, Cross Entropy (CE) and the combination of the two, Dice+CE.

MONAI bundle was modified to add the selected configurations. Data augmentation transforms were applied to the training set to increase generalizability and decrease overfitting. Adam-Weighted optimizer [47] was used with parameters: learning rate (1e-04), weight decay (1e-05), maximum number of iterations (2e04), batch size (1). 16 GB NVIDIA Tesla V100 GPU was employed for the training execution. The overall time required for the training procedure was 20 days (483 hours).

The Dice Coefficient was used to save the best model parameters in the training procedure, and it was also employed to identify the optimal configuration for each fold, resulting in the selection of the top five models. These models were the inputs to the ensemble by majority voting procedure to enhance the robustness of the predicted segmentation on the testing set. The five selected models were UNETR or DynUnet, all 128 patch-based, with Dice or Dice+CE loss function.

The predictions of ASCHOPLEX were compared to FS and GMM outcomes. Literature based segmentation metrics evaluated were Dice Coefficient and Percentage Volume Difference. Percentage Volume Difference calculates the error committed by the prediction in estimating ChP volume compared to the reference and is defined as follows [27]:

$$PercentageVolumeDifference = 100 \times \frac{|Volume_{Prediction} - Volume_{GT}|}{Volume_{GT}} \ [\%]$$

Where $Volume_{GT}$ stands for absolute volume of the Ground Truth (mm^3), that in this case was the manual segmentation by consensus of two neuroradiologists, and $Volume_{Prediction}$ refers to the volume of the prediction (mm^3).

The Dice Coefficient instead measures the similarity between two binary segmentations, indicating the spatial overlap of the two, and it is defined as follows:

$$Dice(G, P) = \frac{2 \times |G \cap P|}{|G| + |P|}$$

Where G is manual segmentation (ground truth), and P is the predicted segmentation. More precisely, the intersection of G and P is the True Positive Rate (TPR) of the number of voxels correctly classified as ChP with respect to the reference, while the denominator corresponds to the sum of TP, False Positive (FP) and False Negative (FN) Rate.

The Dice Coefficient evaluated on the testing set was: ensemble 0.79 ± 0.11, FS 0.31 ± 0.09, GMM 0.54 ± 0.1. The Percentage Volume Difference was: ensemble -1.84 ± 15.76 %, FS -44.95 ± 22.07 %, GMM -19.85 ± 25.59 %.

Results have highlighted ASCHOPLEX provides a robust and reliable ChP segmentation compared to the state-of-the-art algorithms and for this evidence it was selected to obtain ChP segmentations on this study dataset.

3.3 Visual inspection of Choroid Plexus segmentation: Quality Check

Python version 3.9 was used also to visualize the ChP segmentation in the axial plane, with the objective of checking if the ChP was correctly identified by the model. For each patient, a

mosaic image that represented 12 slices of overlapping anatomical images and segmentation was created, with a constant step between the slices, followed by a view of respective anatomical scans only. Each subject therefore had two mosaic images of the selected slices: the first with ChP segmentation highlighted, and the second only with T1-w images. The comparison between these two mosaic images revealed that some segmentations comprised also a part of the hippocampus and brought to the decision of deepen the topic. ITK-Snap version 3.6.0 was then used to have a more complete overview of the segmentations: looking at the entire segmentation of each subject, the inclusion of hippocampus was confirmed as a systematic error that was present in a considerable number of subjects.
In order to quantify the subjects with this issue and to correct the segmentations, it was decided to investigate them in a standardized space.

3.4 Registration to the standard space and Total Intracranial Volume evaluation

After finding out that some segmentations included hippocampus, it was necessary to have a common reference to compare subject's segmentations and to quantify the number of subjects affected by the error. In fact, in a standard space it is assumed that the ChP starts from and ends in the same slices for all subjects. Therefore, the registration of all the ChP segmentations in the standard space allows a quantitative analysis of the inclusion of the hippocampus.
The procedure consists in bringing the images into Montreal Neuroimaging Institute (MNI) space [48], that is an atlas of the human brain provided by Advanced Normalization Tools (ANTs) software package (Figure 14). This procedure is achieved by rigid registration, i.e., rotation and translation, by affine registration, that includes also shearing and scaling, and by non-linear algorithms. The MNI template selected was the ICBM 2009c Nonlinear Symmetric shown in Figure 14.

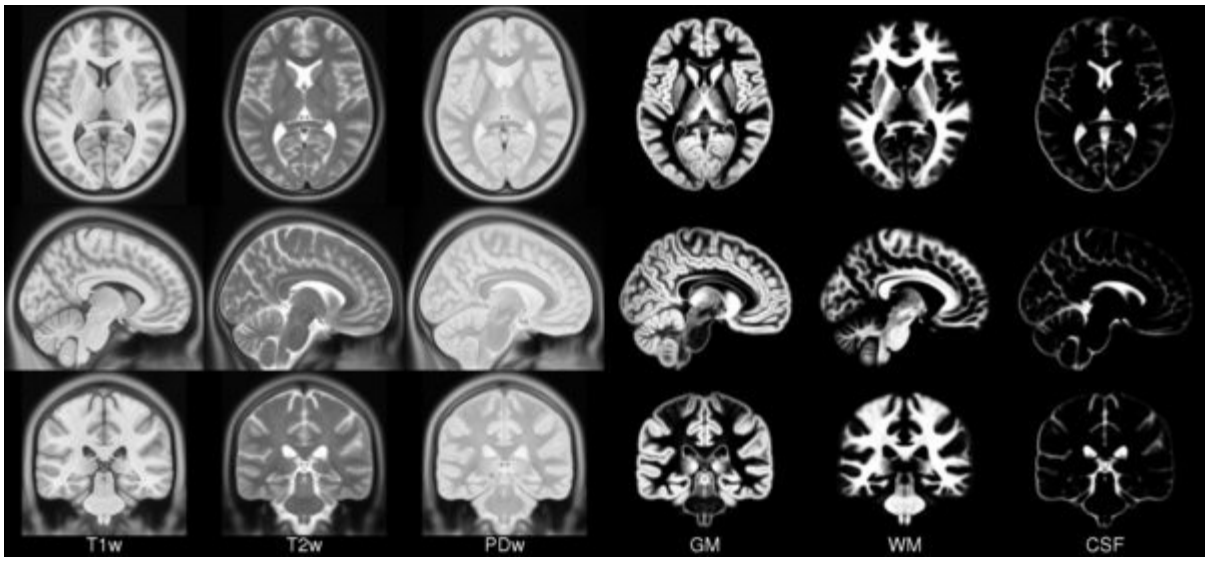


Figure 14 MNI space atlas (https://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLin2009)

In this process, the algorithm also uses whole-brain extraction tools such as ROBEX and Deep Brain. ROBEX (Robust Brain Extraction, https://www.nitrc.org/projects/robex) [49] uses a Random Forest classifier to detect the brain boundary, and a generative model to ensure the plausibility of the results by finding the contour with the highest likelihood. Deep Brain (https://github.com/GUR9000/Deep_MRI_brain_extraction) [50] instead uses a 3D U-Net trained on manual verified skull-stripping datasets with normal images and images with pathologies. It then selects the better brain mask between the ones derived from these tools or the intersection of them, and it uses it to perform the registration. The selected mask has also been exploited to evaluate the Total Intracranial Volume (TIV), as it excludes the skull. The TIV was used to normalize the ChPV; this procedure has already been used in literature for studying the enlargement of ChP in MS [16], in order to get a more robust measure that is not affected by brain volume. So, dividing the ChPV for the TIV, the Normalized Choroid Plexus Volume (NChPV) is obtained:

$$NChPV = \frac{ChPV}{TIV}$$

MNI has also be exploited to correct the segmentations that wrongly included hippocampus in the ChP segmentation. The ChP segmentation obtained after the MNI registration is characterized by levels referring to the slice index in MNI space. We focused only on axial plane (MNI z-coordinate). To correct the ChP segmentation, a threshold was set accordingly to the anatomical reference in MNI space to exclude the hippocampus. The hippocampus is in the lower sliced of the brain, looking at z-axis, as consequence the voxels of the ChP segmentation

classified as ChP belonging to slices under the threshold value in MNI space are turned to zero. Finally, the ChP segmentation was registered to the original subject space, thanks to the reversible transformation, and then normalized, obtaining the NChPV.

Ultimately, MNI has been used to make quantitative comparisons between different scanner acquisitions. The inferior slice and the slice range of the segmentation in the MNI z-coordinate was studied also separating the dataset by the acquisition site: it highlighted a large bias between the two groups that has been corrected using NeuroCombat [51].

3.5 NeuroCombat: a tool for MRI dataset harmonization

NeuroCombat [51] is a widely utilized tool for harmonizing extracted features from neuroimages, enabling the entire cohort to be processed as if all images were acquired from the same scanner. It is adopted from genomics literature, and it models and removes unwanted scan variability but preserves biological associations in the dataset [52], allowing the analysis of large-scale imaging databases and increasing the power and reproducibility of statistical analysis to test common biological responses.

NeuroCombat was implemented in the R statistical software (version 4.3.1) (https://github.com/Jfortin1/ComBatHarmonization).

The algorithm consists in estimating scanner-specific location and scale parameters for each feature from post-processing data. So, in this thesis it was applied on NChPV, after the ChP segmentation threshold selection in MNI space, including Dataset 3 (ALT) in Dataset 2 (BR), and considering age and sex as biological associations to preserve.

NeuroCombat requires as default data inputs the measurements derived from the imaging acquisitions to harmonize (NChPV), and the scanner group (or "batch") to which they belong; regarding the batches, there is the possibility to indicate a reference batch that is considered more reliable, and BT was set to reference batch because of the smaller number of subjects with hippocampus inclusion in the segmentations. Together with that, it gives the chance to add biological variables that should be protected from the removal of site effects, that in study were age and sex. As default, it runs an empirical Bayes framework to improve the variance of parameter estimates pooling information across all the features. In this case it was preferable not to use empirical Bayes because the number of features was substantially smaller than the number of participants ($1 << 167$). The NChPV in output from NeuroCombat is the final volume used for the subsequent analysis.

3.6 Statistical Analysis on Cerebrospinal Fluid derived molecules and Choroid Plexus volume

In this section there is a description of methods used to search for statistically significant relationships between CSF molecules and NChPV, used to investigate the difference between two variables, like demography data or ChPV before and after each processing step, or used to assess normality of data distribution.

3.6.1 Correlation Analysis

The study of correlation between CSF molecules has been implemented to observe significant relationships that could negatively affect the subsequent linear regression models. Correlation has been studied also between CSF molecules and ChPV, with the objective of finding potential significant regressors.

The Pearson's Correlation coefficient *r* is the most popular measure for the strength of a statistical relationship, and it is essential for the study of the joint behavior of two or more variables. It can assume values between –1 and +1 (strong negative linear association, strong positive linear association), and it is defined as follows:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{x})^2]}}$$

where x and y are the two variables of interest, and $\bar{x}$ and $\bar{y}$ their means respectively.

The main problem that affects Pearson coefficient is the susceptibility to the influence of outliers and data distribution.

Spearman correlation is used to confirm Pearson results. Spearman correlation coefficient instead is a non-parametric measure that needs to arrange *x* samples from smallest to largest and to assign a rank R_i for each value, and the same thing for *y* assigning to each value a rank S_i. The statistic is the same as Pearson, but samples values and means are replaced with ranks:

$$\rho = \frac{\sum (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{[\sum (R_i - \bar{R})^2][\sum (S_i - \bar{S})^2]}}$$

The two coefficients are close if there are a few or no outliers.

All the p-values derived from correlation analysis have been adjusted for multiple comparisons. For adjusting the p-values for multiple comparisons, the selected method is the one from

Benjamini & Hochberg (1995) [53], that controls the false discovery rate, a less stringent condition than the common family-wise error rate and therefore a more powerful tool than others.

3.6.2 Comparative Analysis

The two-sample t-test compares two population means using data from two independent samples and in this thesis has been used to compare male and female subject's volumes and features from different scanners. Defining μ_1 and μ_2 the mean of the two populations, the null hypothesis is [54]:

$$H_A: \mu_1 = \mu_2$$

and the *t* statistic is defined as follows [54]:

$$t = \frac{\overline{x_1} - \overline{x_2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}(\overline{x_1} - \overline{x_2})}$$

Where:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$\overline{x_1}$ and $\overline{x_2}$ are sample means, n_1 and n_2 are sizes of the two samples, and s_1 and s_2 are the sample variances.

Wilcoxon Signed-Rank Test is an alternative test to the aforementioned t-test, that can be generalized also to non-independent samples. It is a non-parametric procedure that uses ranks to compare paired population means like ChP volumes before and after a processing step. It is more robust and less affected by extreme observations, and it evaluates the null hypothesis that the two medians are identical. The samples of both populations are combined from smallest to largest and assigned a rank.

The statistic is defined as [54]:

$$z = \frac{R - \mu_R}{\sigma_R}$$

Where:

$\mu_R = \frac{n_1(n_1 + n_2 - 1)}{2}$, $\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$, and R is the sum of the ranks from the sample with size n_1.

3.6.3 Methods To Assess Normality

Many of the data processing methods used in this study assumes a gaussian distribution of the variables in the dataset. The Shapiro-Wilk test (W test) has been implemented in R and used to assess normality and lognormality of CSF variables. The null hypothesis of the test is that the samples have a normal distribution, so if the p value is equal to or less than 0.05, the null hypothesis is rejected and therefore the variable is not normal. In contrast, p value > 0.05 means that the test couldn't find a significant difference from normality, so the W test is passed, and the variable is considered to have a gaussian distribution. This test is less reliable when many missing values are present, so the result has to be analyzed together with the number of Missing Values (MV). This test has been applied to the ChPV both before and after the threshold, to every variable with at least 30 non-missing values, and to the base-2 logarithm of all these variables since it was suspected a lognormal distribution of most of the variables.
Different plots are used for visual inspection of variables to assess normality. Given the fact that tests for normality often apply stringent methods to assess gaussian distribution, a visual inspection of a variable distribution can bring to consider it normal even if the test is negative and therefore to apply processing that assumes gaussian distribution.

3.7 Methods to assess the relationship between Choroid Plexus volume and Cerebrospinal Fluid derived variables

The CSF dataset comprised a large number of variables, even when removing variables with too MVs (left number of molecules: 84). In order to apply linear models to predict the NChPV, Lasso and Stepwise regression are used to reduce the number of molecules to use. These models are computed:

- Lasso Regression on a full matrix of 44 variables and 133 subjects (80%), without missing values (referred to as Selection1).

- Lasso Regression on a full matrix of 54 variables (Subset 2) and 84 subjects, without missing values (referred to as Selection2). A linear model is based on this selection.
- Lasso Regression on a full matrix of 26 variables from Subset 1 and 91 subjects, without missing values (referred to as Selection3). A linear model is based on this selection.
- Stepwise Regression on a full matrix of 54 variables (Subset 2) and 84 subjects, without missing values. This model is then:
    - added age, getting a second linear model.
    - optimized reducing the number of predictors, getting a third linear model.
    - Generalized for the 167 subjects.
    - Optimized for the 167 subjects.

Finally, logistic regression, is applied to IgG_OCB, the only categorical variable.

3.7.1 Linear Regression

Linear regression aims to study how much an independent variable, or a combination of independent variables, called *x* can predict a real-valued output, i.e., the dependent variable Y. In this thesis, this method is applied to see how much the NChPV (in formulas it is referred to as NV_i^{ChP}) is explained by some selected predictors among the CSF molecules. The model is described by the following formula [54]:

$$Y_i = \beta_0 + \sum_{j=0}^{k} \beta_j x_{ji} + \varepsilon_i$$

where Y_i is the dependent variable from the i^{th} subject, which in this thesis corresponds to the ChP volume; $\beta_0, \beta_1, \ldots, \beta_k$ are the $k+1$ unknown coefficients to estimate; x_{ji} is the value of the *j*th independent variable from the *i*th subject, that corresponds to the CSF derived biomarkers and other covariates from population demographics. In this thesis, every variable has been normalized, getting a zero-mean and a unitary variance population; consequently, the intercept β_0 of every regression model will be set to zero, so at the end the formula can be re-written as:

$$NV_i^{ChP} = \sum_{j=0}^{k} \beta_j x_{ji} + \varepsilon_i$$

ε_i is the model random error term that is distributed as a normal with zero mean and unknown variance σ^2 [54].

The estimation of parameters β_0,β_1, …, β_k is done by least squares method, that minimizes the sum of squared deviations S:

$$S = \sum_{i=1}^{n} (NV_i^{ChP} - \sum_{j=0}^{k} \beta_j x_{ji})^2$$

The VIF (Variance Inflation Factor) method has been used in combination with the correlation to exclude the potential presence of collinearity between the covariates used for the regression, which could lead to problems in fitting and interpreting the model. It is computed by measuring how much the variance of a regression coefficient is inflated due to multicollinearity, with the following computation [55]:

$$VIF = \frac{1}{1 - R^2}$$

There is no fixed VIF value that implies collinearity, and usually the threshold is set between 5 and 10. In this thesis, since it was a preliminary study of the relationship between NChPV and CSF molecules, it was set that a VIF>10 was considered an indication of collinearity.
To assess the quality of the regression model the determination coefficient R^2 is used. It is the square of the correlation coefficient, and it can be interpreted as the proportion of observed Y variation that can be explained by the model. It can be expressed as:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

where SST is the total sum of squares and measures the total amount of variation in Y that can be written as SSR = SSR + SSE; SSR is the regression sum of squares, defined as $SSR = \sum (\hat{y}_i - \bar{y})^2$, that measures the variation explained by the linear relationship; SSE is the part of variation that remains unexplained and it is defined as $SSE = \sum (y_i - \hat{y}_i)^2$. R^2 then ranges from 0, that means that all the coefficients are zero, to 1, and it indicates how much the dependent variable is explained by the predictors. Naturally, adding a new covariate, the R^2 would increase, and it wouldn't be a good measure of the improvement brought by the new explanatory variable. $\bar{R}^2$ (R^2 adjusted) in this case can "adjust" the evaluation of the model by using degrees of freedom $df_e = n - k$ and $df_t = n - 1$. The formula becomes:

$$\bar{R}^2 = 1 - \frac{SSE/df_e}{SST/df_t}$$

The accuracy of the coefficients is used to evaluate the goodness of the model and it has been expressed in terms of coefficient of variation (CV) that are defined as follows:

$$CV_{\beta_i} = \frac{sd(\beta_i)}{\beta_i}$$

The acceptable range established for evaluating linear models in this thesis is CV < 100%.
The goodness of the fit is evaluated also in terms of standardized residuals. Residuals are defined as the difference between observed values and estimated values using the model. The residuals have been transformed to get a unitary variance and a zero mean, so they have been referred as standardized residuals; their evaluation consists in visually checking if they are comparable to white noise distribution.
The linear models have been fitted with multiple combinations of variables. The starting points for choosing linear model predictors are the molecules that showed a non-zero coefficient in Lasso (Section 3.7.2), or Stepwise (Section 3.7.3) regression. The models are then optimized by removing the predictors with high values of CVs or VIF, or with high correlation with other predictors.

3.7.2 Lasso Regression

Lasso regression is a least absolute shrinkage method used to select the most meaningful variables of a model. It aims at shrinkage regression coefficients by applying a L1 penalty on their size, defined as $\sum_{j=1}^{p}|\beta_j|$, that makes the solutions nonlinear in y_i. The lasso estimate [56] is a quadratic problem without closed form defined by:

$$\beta^{Lasso} = arg\min_{\beta} \sum_{i=0}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_{ij})^2$$

If the predictors are standardized, β_0 is equal to $\bar{y}$, so it becomes a model without intercept.
In contrast with other shrinkage methods, Lasso will cause some coefficients to become exactly zero, so in fact it performs a kind of variables subset selection.

The $\lambda > 0$ tuning parameter controls the amount of shrinkage, so a bigger λ (or a small enough t) means a bigger regulation and so β^{Lasso} leading exactly towards 0.

In this work, Lasso regression has been used to determine the subset of CSF derived variables that could better describe the normalized ChP volume Starting from the largest possible dataset without missing values, Lasso was performed and only the variables with non-zero coefficients have been used then in linear models to predict NChPV.

The same procedure has been performed starting from the molecules from Subset1 (without IL12p70, CCL23, NFL, CSF_Albumin IL_28, Fibrinogen and Parvalbumin, for the large number of MVs): in this case, it is used to get a reduced number of molecules considered significant from literature, as a starting point to select linear model predictors.

Lasso has been implemented in R and the function takes only the subjects with no missing values; to avoid considering a very limited number of subjects, the following algorithm has been applied:

a) The variables with at least 30 non-missing values are ordered based on the increasing number of missing values.

b) The variable with fewer missing values is added to the subset of selected variables.

c) The number of subjects with no missing values in this subset is computed.

d) If the number of subjects excluded because of the presence of MV exceeds the threshold of 20%, the algorithm stops. Instead, the cycle from point b) is repeated until the subjects with all numeric values are less than 80%.

Using this procedure, a subset of 133 subjects has been used for this analysis. These are the input variables of the function in R `glmnet()`, that is firstly used to find the best λ by means of a 10-fold cross-validation that minimizes the test Mean Square Error (MSE), and then it is re-applied specifying it to find the coefficients of the best model. The output of the function is then the set of the coefficients in which some are not visualized because they have been forced to zero.

3.7.3 Stepwise Regression

Stepwise regression identifies a small subset of factors from a large dataset that relate significantly to the outcome, excluding predictors that have no real relationship to the outcome. In this thesis, StepWise regression is used to select the most significant predictors of NChPV

between CSF molecules. It was applied on Subset2, and the 22 coefficients in output from the model have been used for 4 linear models (two on 84 subjects, and two on the entire cohort; without and with correcting for CVs and VIF). The procedure consists in a forward selection procedure with reexamination of the variables included in the previous steps, which aims at avoiding redundant information. Stepwise regression can be seen as a sequence of forward and backward steps, and it can be summarized through the following steps [54]:

1. Fitting of a simple linear regression model to each factor.
2. Selection of the most important factor according to a predetermined criterion.
3. Test for significance of the factor selected and determine according to the Akaike Information Criterion (AIC) whether to add the factor to the model.
4. Test for significance of the factors included in the model, and elimination of the variable from the model if it meets the elimination criterion.
5. Repetition of steps 2 and 3 until no variable meets the criterion.

3.7.4 Logistic Regression

In contrast with linear regression, logistic regression objective is to predict a qualitative categorical response or, more precisely, the posterior probabilities of the K response variable classes [54]. In this thesis, the only CSF variable with a categorical (0-1) response is the IgG_OCB, that is the presence (1) or absence (0) of the oligoclonal band corresponding to the immunoglobulin IgG. As described in section (1.5) the IgG in OCB is a hallmark of specific MS changes in CSF, and therefore it is interesting to investigate a possible relationship between this biomarker and ChPV. In particular logistic regression it's been used to know whether ChP volume is able to predict or not IgG_OCB value; assuming π the probability of IgG_OCB = 1, and therefore assuming $1 - \pi$ the probability of IgG_OCB = 0 (and therefore having a Bernoulli variable), the estimation of parameter β_1 can be done iteratively starting from the following linear model in the log-scale [54]:

$$ln\frac{\pi_i}{1 - \pi_i} = \beta_1 x_i$$

derived from the general logistic function:

$$f(\beta_1 x_i) = \frac{1}{1 + e^{-\beta_1 x_i}}$$

which ranges from 0 to 1 and has an incrementing S-shape with a threshold.

4. RESULTS

This section aims to present the results obtained from the work exposed in the previous chapter. The first two paragraphs illustrate the results of the validation of the automatic segmentation tool and the outcomes of the segmentation processing. The following sections instead provide a report of the results of the statistical analysis of biomarkers acquired from the CSF and of the study of the relationship with the NChPV.

4.1 Choroid Plexus Automatic Segmentation: Quality Check and Volume computation

After the segmentation of the ChPs of the entire dataset through the algorithm described in paragraph 3.2, a preliminary overview of the segmented volumes was performed (Figure 15) in order to visualize the slices that contain ChP in the axial plane. The objective of this initial step was to identify possible significant anomalies that were worth further investigating; this step was performed by comparing the segmentation superimposed to the anatomical image to the original images in the range where the segmentation was present with fixed intervals. During this analysis, it emerged that the algorithm could have included in the ChP also a portion of the hippocampus in the inferior area of the segmentation in some subjects (an example is shown in Figure 16). In this step also the indexes of the slices in the axial plane have been visualized to check potential lack of continuity of segmentation.

The ChPV is computed for each patient considering 1 voxel = 1 mm^3. The mean volume was 2531.7 ± 692.8 mm^3.

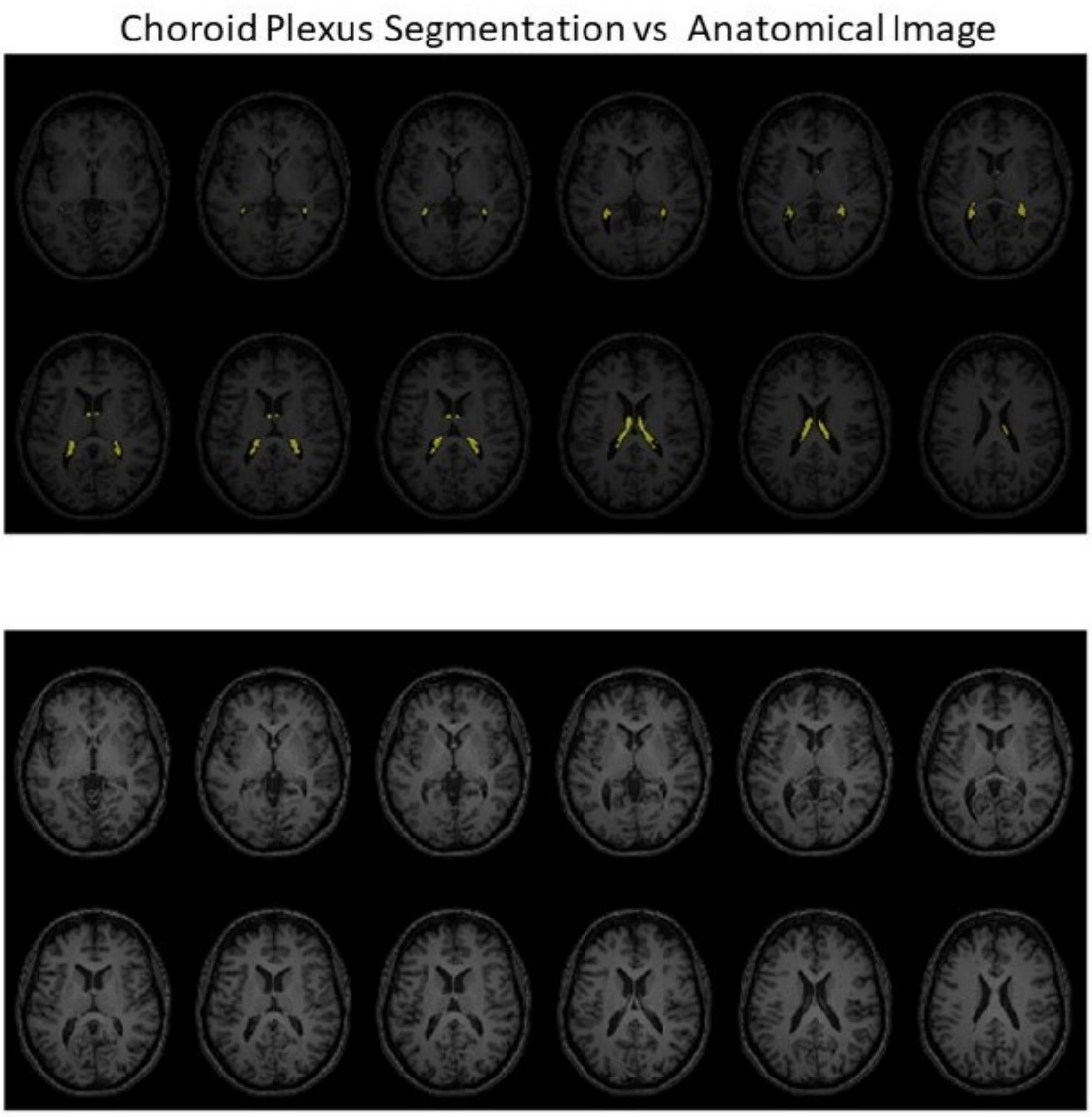


Figure 15 Quality check of Choroid Plexus (ChP) segmentation for a representative subject. The upper figure shows a selected number of slices of the ChP over imposed on the T1-w MRI image, while the lower figure represents only the anatomical scan. In this subject the ChP is correctly identified.

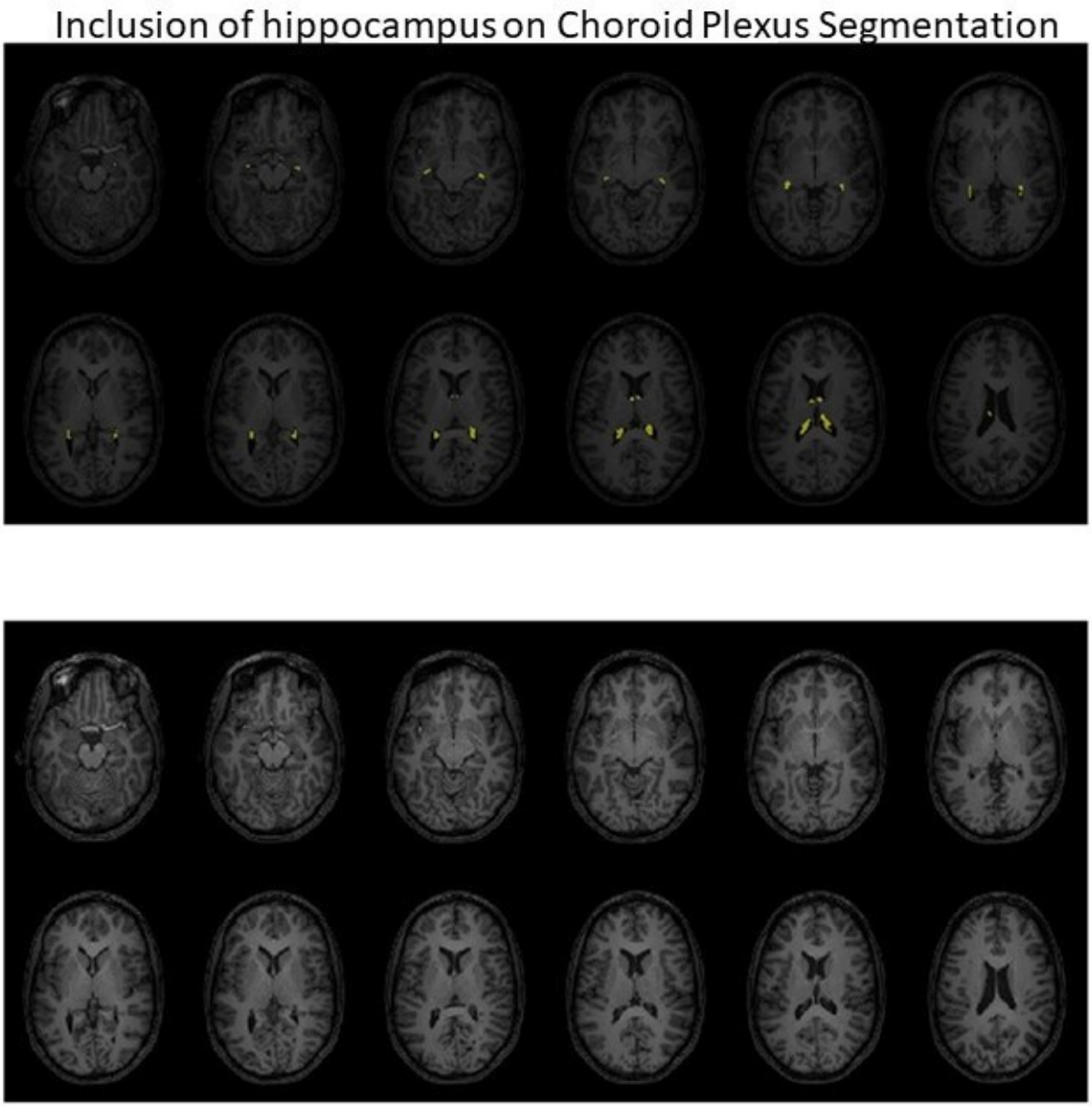


Figure 16 Quality check of Choroid Plexus (ChP) segmentation for a representative subject. The upper figure shows a selected number of slices of the ChP over imposed on the T1-w MRI image, while the lower figure represents only the anatomical scan. In this subject ChP segmentation exceeds hippocampus; It can be seen in the slices the upper row.

Therefore, an in-depth analysis has been performed by way of ITK-SNAP, that, thanks to all three planes visualization, allowed to confirm the inclusion of the hippocampus in the segmentation of identified subjects (approximately one third of the dataset).

4.2 Choroid Plexus Segmentation processing after standard space registration

The registration of the segmentations into standard MNI space allowed to compare the position of the segmented ChPs within the whole dataset. Table 2 shows the results of the MNI slice analysis.

The histogram of the inferior ChP segmentation slices in MNI space shows a bimodal trend (Figure 17) with a local minimum of the distribution at slice 66th. This is confirmed by the histogram of the range of slices that contain ChP, that reflects the inclusion of the hippocampus in the volume for a considerable portion of subjects (Figure 18).

Superior ChP segmentation slices instead have more consistent indexes and don't show a significant difference between scanners (t = -1.2251, df = 57.177, p-value = 0.2256) (Figure 19).

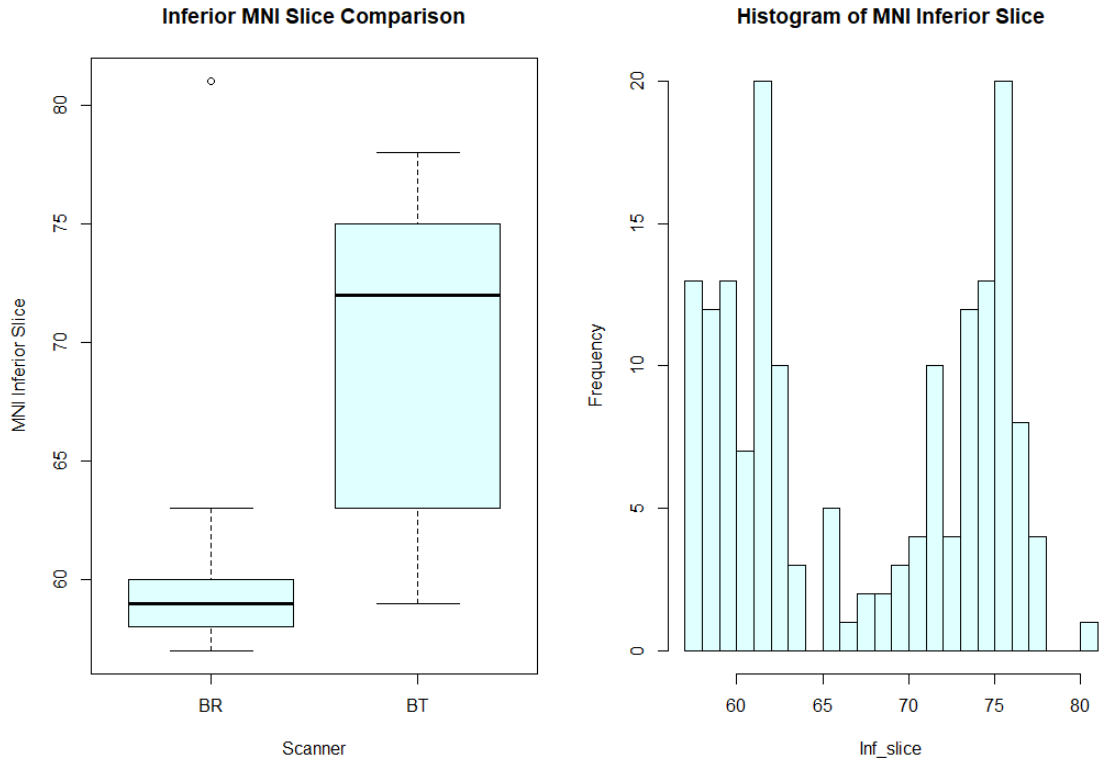


Figure 17 Choroid Plexus (ChP) inferior slice index in MNI space. Left: boxplot shows differences between scanners of the index of the first inferior slice containing ChP segmentation in the axial plane. Right: There is an evident bimodal distribution in the histogram representing ChP inferior slice index in MNI space, that suggests an erroneous inclusion of hippocampus in the segmentation.

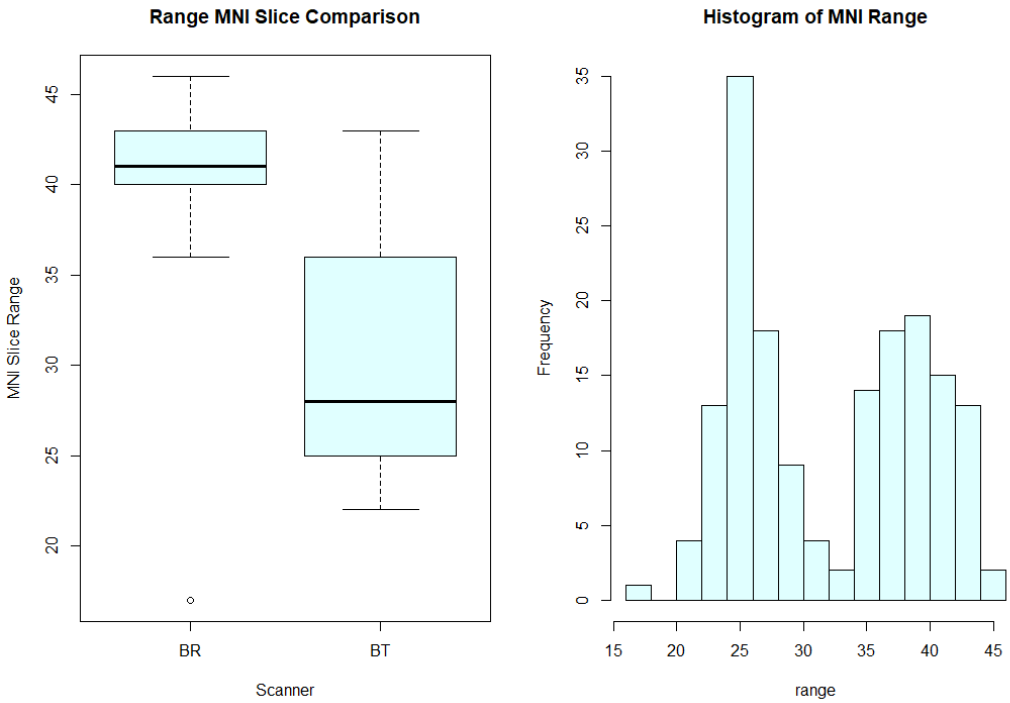


Figure 18 Choroid Plexus (ChP) slice range in MNI space. Left: boxplot shows differences between scanners of the range of slices containing ChP segmentation in the axial plane. Right: there is an evident bimodal distribution in the histogram representing ChP slice range in MNI space, that suggests an erroneous inclusion of hippocampus in the segmentation.

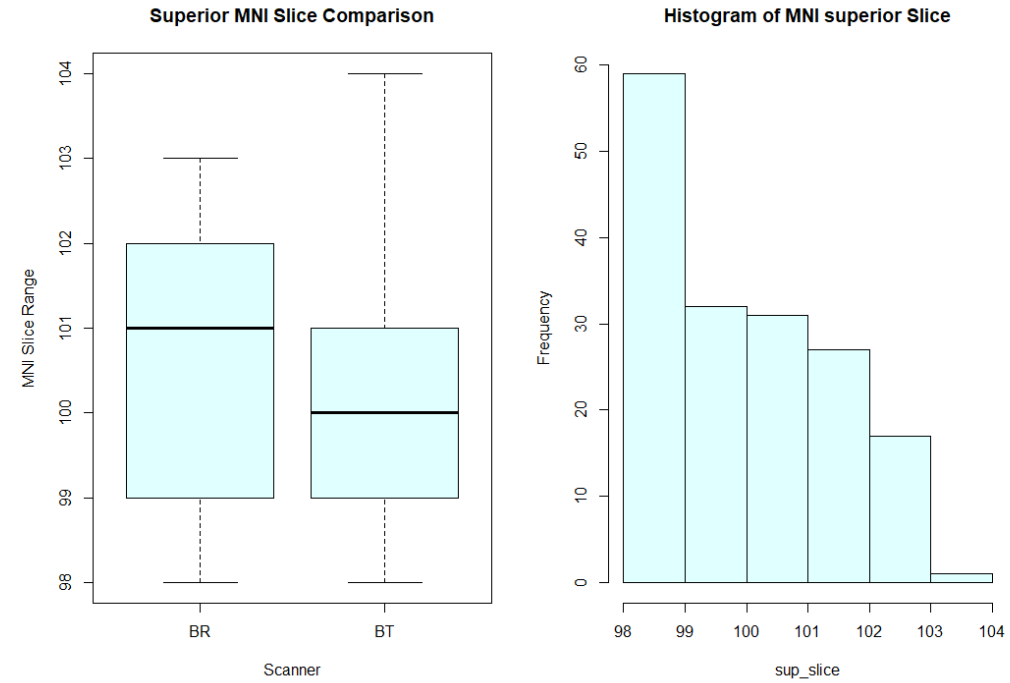


Figure 19 Choroid Plexus (ChP) superior slice index in MNI space. Left: boxplot shows the absence of difference between scanners of the index of the last superior slice containing ChP segmentation in the axial plane. Right: Superior Slice assumes a limited range of values, without a bimodal distribution.

	Dataset 1: BT	Dataset 2: BR+ALT
MNI Inferior Slice	69.95 ± 6.16	59.29 ±1.57
Slice Range MNI	30.33 ± 6.02	40.82 ± 4.54
MNI Superior Slice	100.29 ± 1.52	100.65 ± 1.63

Table 2 Segmentation analysis in MNI space. The results are displayed as mean ± standard deviation.

4.2.1 Threshold of Choroid Plexus segmentation in standard space

The following step consisted in cutting the segmentation portion under the 67th slice in MNI because slice 66th corresponds to the local minimum of the inferior slice distribution. 83 patients presented an inferior slice under the 67th slice. Thresholding operation is done through the reverse transform from MNI space to the original T1-w MRI space of the subject's brain. The resulting post-threshold volume is plotted in Figure 20 together with the initial volume extracted from ASCHOPLEX output. The major differences are encountered mainly in the most recent scans acquired from Dataset 2.

The paired t-test states a significant difference between the volume before and after the thresholding (t = 7.1615, df = 168, p-value = 2.382e-11).

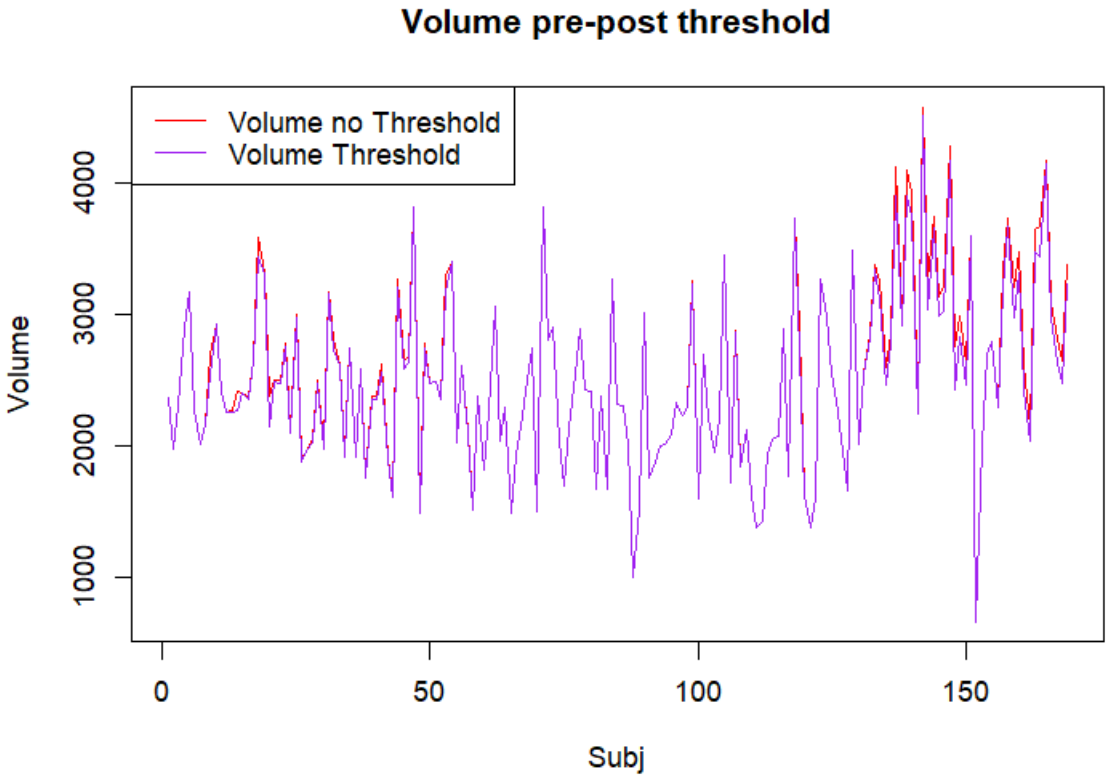


Figure 20 Choroid Plexus Volume (ChPV) comparison before (red) and after (purple) the threshold on the 67th slice in MNI space. The major differences are encountered mainly in the most recent scans acquired from Dataset 2.

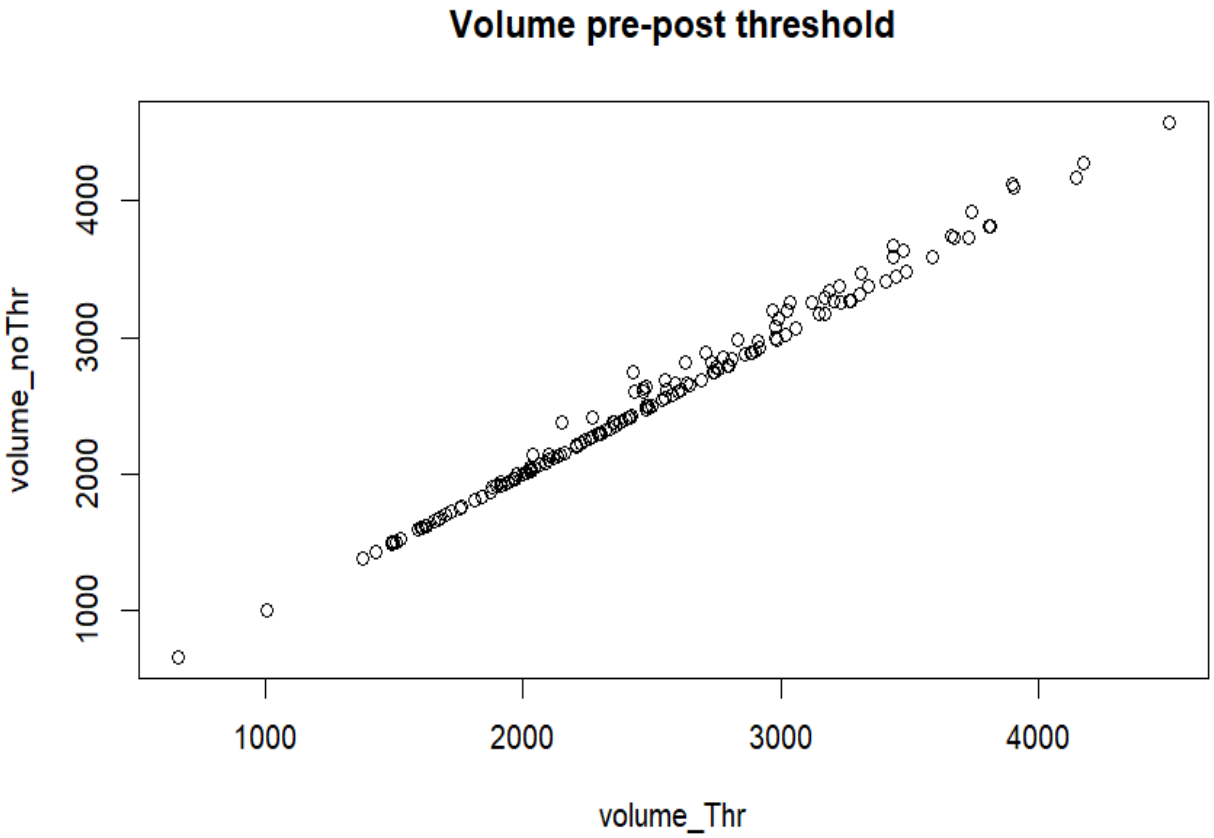


Figure 21 Scatter Plot representing Choroid Plexus Volume (ChPV) before and after the application of the threshold. Subjects that don't rely on the diagonal included hippocampus in the segmentation.

Both the original volume and the volume after the application of the threshold result normally distributed using the Shapiro-Wilk normality test (respectively, p-value = 0.0792, p-value = 0.09398), and the scatterplot (Figure 21) suggests a non-exact correlation. After the threshold, the histogram shows a more regular profile (Figure 22B) compared to the initial volume (Figure 22A). The mean ChPV after the threshold was 2495.4 ± 663.9 mm^3. Using Wilcoxon Signed-Rank test, the volume before and after the threshold result statistically different (V = 3106, p-value = 8.94e-14)

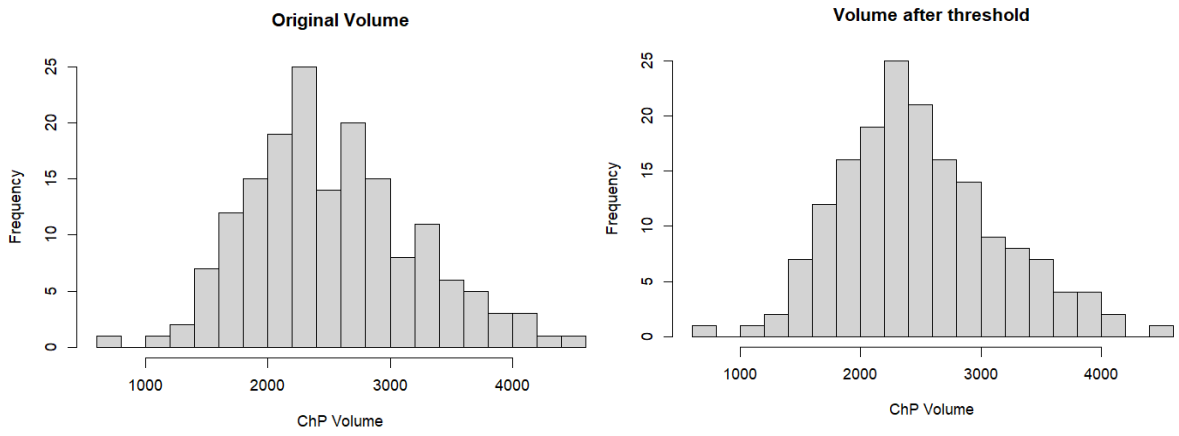


Figure 22 A: histogram of Choroid Plexus volume (ChPV) distribution obtained from the DNN ensemble. B: histogram of ChPV after the removal of voxels under the 67th slice in the axial plane in MNI space.

The TIV computed for the whole cohort (from the Brain Extraction Mask selected from all the skull-stripping methods) has a mean of 1358978 ± 137023mm^3.

4.2.2 Choroid Plexus volume estimation differences between scanners

A comparison between the two scanners has been performed to investigate a possible relationship with the inclusion of the hippocampus in the segmentation. The ChPV of the two populations has been compared before the application of threshold (Figure 23), after the threshold operation (Figure 24), and after the normalization (Figure 25) performed using TIV acquired from MNI mask (Table 3). Both before the threshold (t = -6.75, df = 50.95, p-value = 1.34e-08) and after the threshold (t = -5.77, df = 50.85, p-value = 4.81e-07), subjects from different scanner have a significantly different distribution. Moreover, after the normalization the distributions are statistically different (t = -6.85, df = 54.83, p-value = 6.62e-09).

	Dataset 1: BT	**Dataset 2: BR+ALT**
Volume pre threshold (mm^3)	2338.4 ± 556.4	3187.8 ± 714.1
Volume post threshold (mm^3)	2331.1 ± 552.7	3053.6 ± 711.8
Normalized volume	1.71e-03 ± 3.99e-04	2.28e-03 ± 4.56e-04

Table 3 Comparison between the two scanners before the threshold under the 67th slice, after the threshold, and after normalizing for total intracranial volume (TIV).

Boxplots in Figure 23, Figure 24 and Figure 25 show the three cases described in Table 3. Before the threshold the difference between the two scanners is evident: BR shows significant larger values of volume than BT. BR also shows an outlier with a small ChPV. After the threshold, the difference is smoothened, but still present. Finally, after the normalization, NChPVs show difference similar as the ChPV before the threshold.

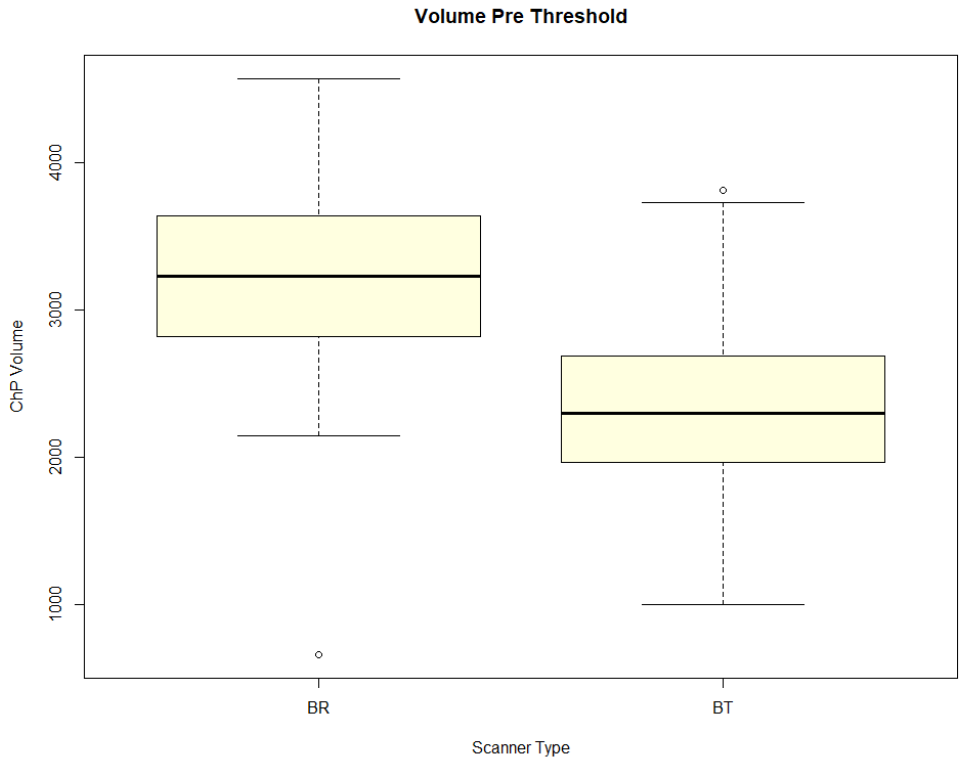


Figure 23 Boxplot of Choroid Plexus volume (ChPV) computed directly to the output segmentation of the algorithm. It highlights the differences between the two scanner types in terms of average volume.

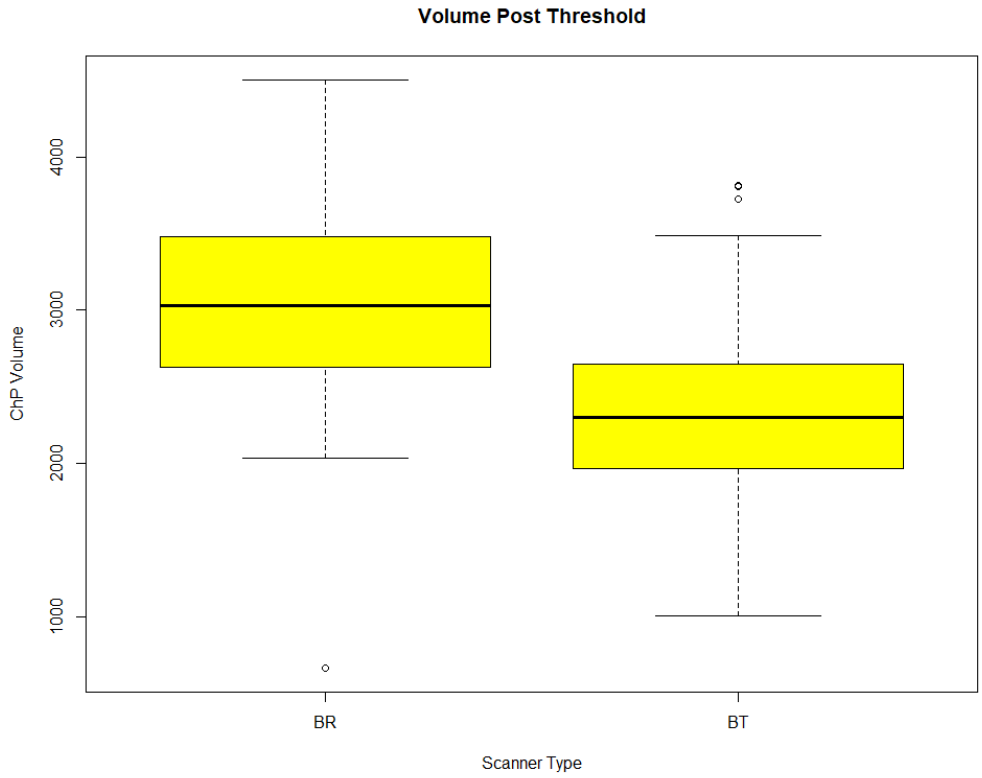


Figure 24 Boxplot of ChP volume computed after the slice threshold applied into the MNI space. Regarding the average segmentation volume, the population shows less difference than Figure 22 between the two scanners.

TIV for BT patients has a mean value of 1365364 ± 140254mm^3, in BR+ALT instead the TIV mean value is 1337297 ± 124731mm^3. The difference between the two populations isn't significative (t= 1.1841, p = 0.2406).

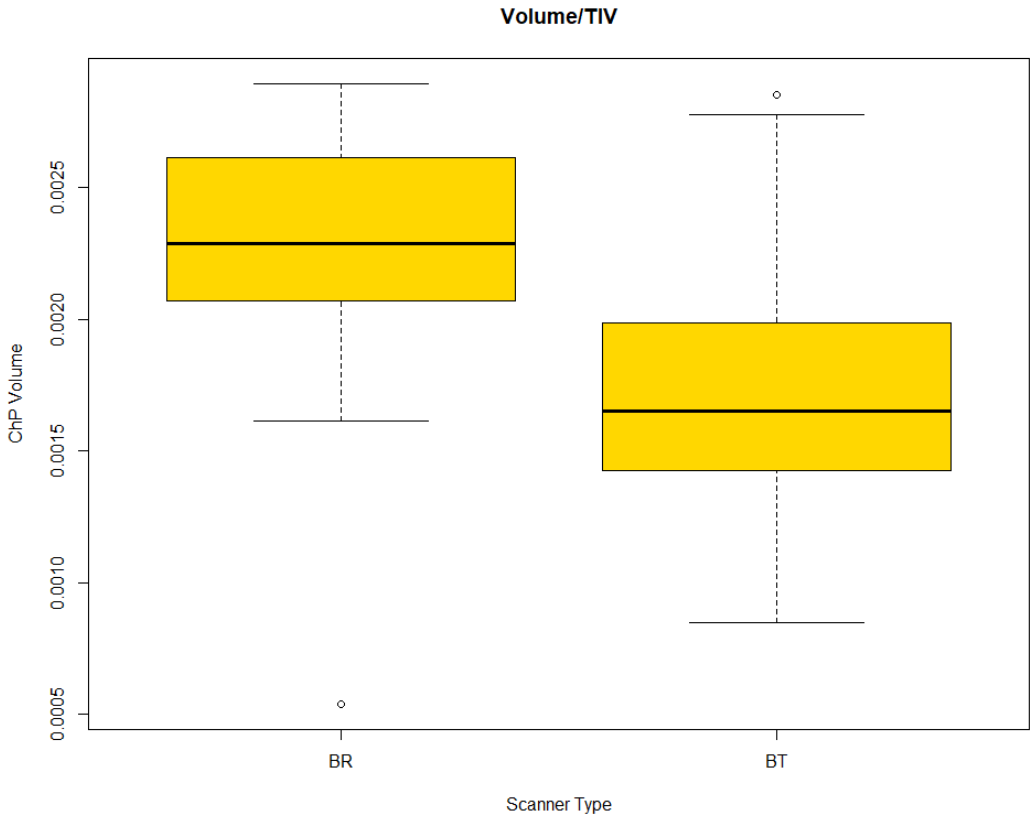


Figure 25 Boxplot of Choroid Plexus volume (ChPV) computed after the slice threshold applied into the MNI space and after the normalization of this volume using TIV. The difference has returned more prominent, so the pattern has returned more similar to Figure 23

4.3 Harmonization of Choroid Plexus normalized volume

Harmonization with NeuroCombat was applied on NChPV because there was no significant difference in TIV between BT and BR sites (Figure 26).

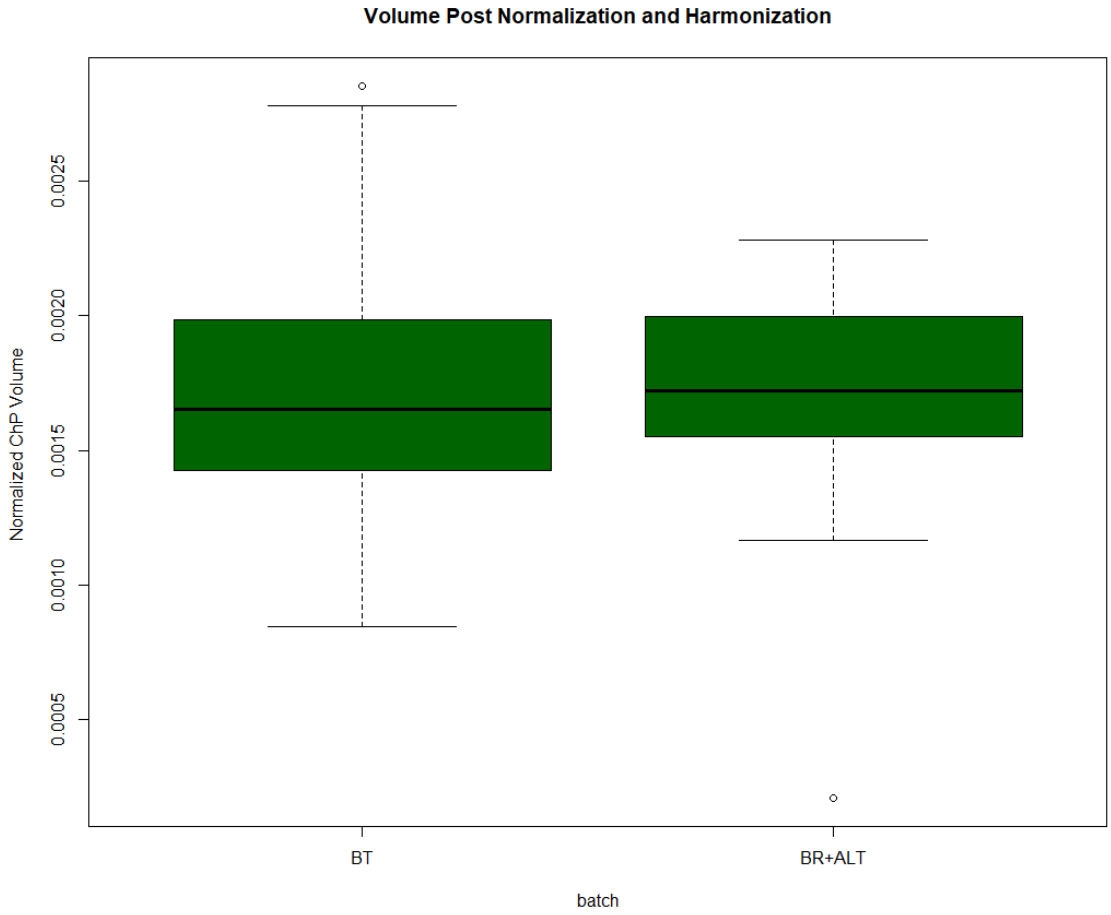


Figure 26 Normalized Choroid Plexus Volume (NChPV) after harmonization through Neurocombat. No significant differences are found between the two scanners (batches), BT and BR+ALT.

After harmonization of NChPVs, the two distributions are no longer significantly different. (BT: 1.71e-03 ± 3.99e-04, BR+ALT: 1.72e-03 ± 3.94e-04, t = -0.039, p-value = 0.9687).

4.4 Statistical Analysis on Cerebrospinal Fluid derived molecules and Choroid Plexus volume

The first objective regarding the dataset of the CSF derived biomarkers was to have an overview of the available information, so every variable extracted from the liquor analysis has been analyzed counting how many samples were available and only molecules with at least 30 samples were taken into account in the following analyses. As anticipated in section 3.1, the initial number of variables was 130; considering only molecules with enough samples, the number of left CSF variables is 84.

After this selection, as said in section 3.5.3, the W-test has been performed to assess normality both to raw data and to the base-2 logarithm of the variables. IgG_OCB has been excluded from this analysis because it is categorical.

No variable resulted having a gaussian distribution, but the following 14 molecules on Table 4 are resulted normally distributed considering the base-2 logarithm:

Molecule normally distributed	**MV**
IgM	78
IgG1	78
HemoglobinNg_ml_mgProt	138
TSLP	0
IL_20	6
IL_12_p70_	41
IL_10	0
BAFF_TNFSF13B	0
MIP_1alpha_CCL3_55_	0
MIF_35_	0
EOTAXIN_2_CCL24	1
CSF_IgG_mg_L_	59
EOTAXIN_CCL11	7
ProteineTotaliCSF_g_L_	0

Table 4 Normally distributed variables (using W-test) with correspondent number of missing values (MV). Molecules in bold below to Subset1. A high number of MVs might affect the result.

Among these molecules resulting normally distributed, IL_12_p70_, BAFF_TNFSF13B, MIF_35_ and EOTAXIN_CCL11 are included also in the Subset1 of molecules considered of interest from literature.

Using `lillietest` implemented in Matlab, also sCD14Pg_ml_mgProt, MMP_2, IL_32, CSF_Albumina_mg_L_CC resulted normally distributed.

By visual inspection of the variables histogram, every CSF variable could be approximated to a normal distribution and therefore the function `scale()` has been applied to the entire CSF dataset.

For the following analysis, only the NChPV after the harmonization has been considered.
For what concerns Subset1, a correlation analysis has been performed among them and with NChPV, considering the entire cohort. No significant correlation with NChPV has been found both with Spearman and Pearson correlation and adjusting p-values for multiple comparisons (Figure 27). On the contrary, there are high correlation values between lot of CSF molecules.

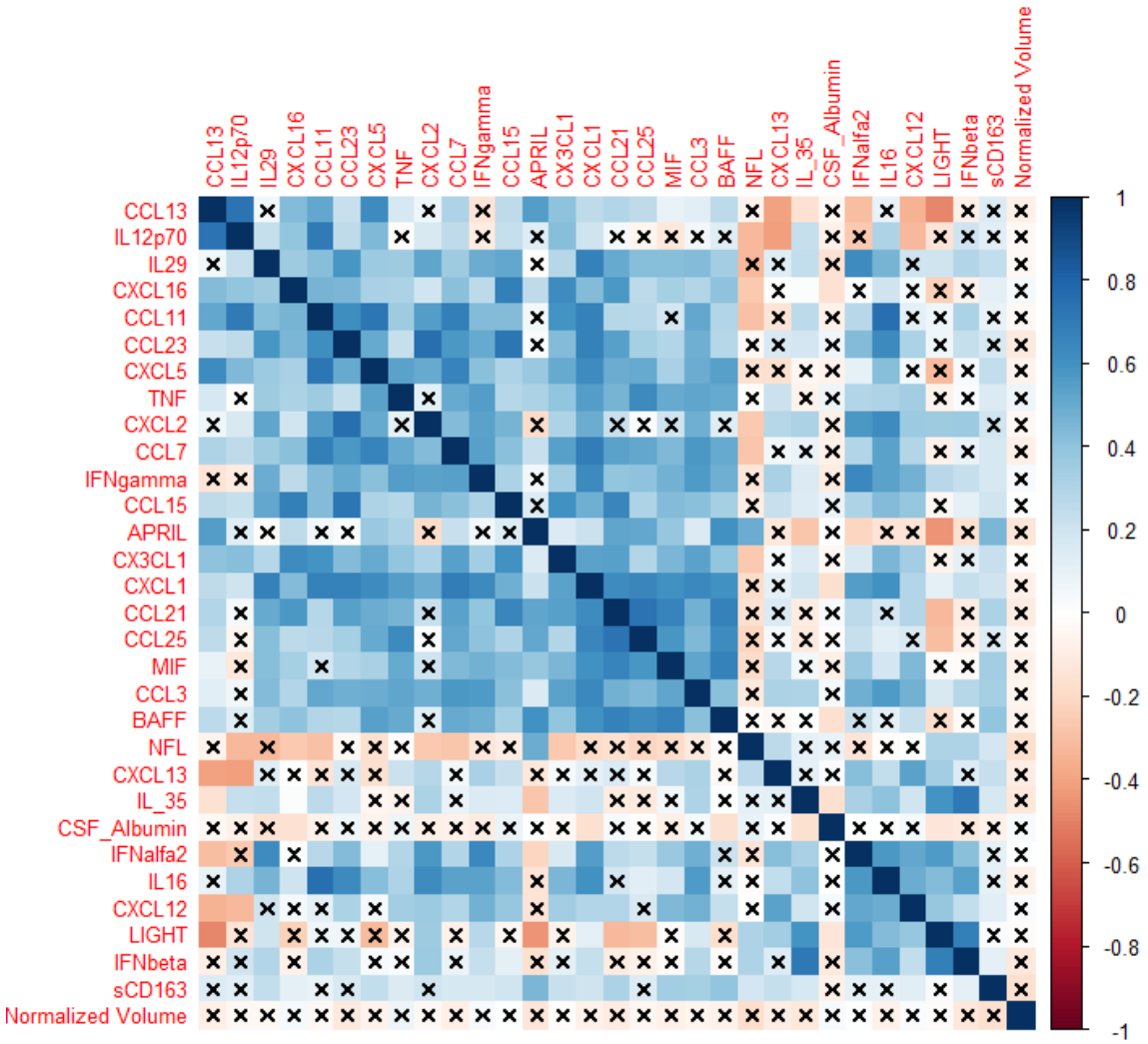


Figure 27 Correlation Matrix between the molecules from Subset1, meaning that they are all related to multiple sclerosis. The X indicates non-significant correlation, considering p-values adjusted for multiple comparisons. Normalized Choroid Plexus Volume shows no significant correlation, while lot of molecules are highly correlated.

The following Figure 28 is used to study the normality of variables. It shows the histogram of Subset1 variables after the computation of the base-2 logarithm, together with the p-value of Saphiro-Wilk test and the number of MVs. The highest number of MVs are from NFL (96),

CSF_Albumin (57), and IL12p70 (40). Most of the variables are comparable to a normal distribution.

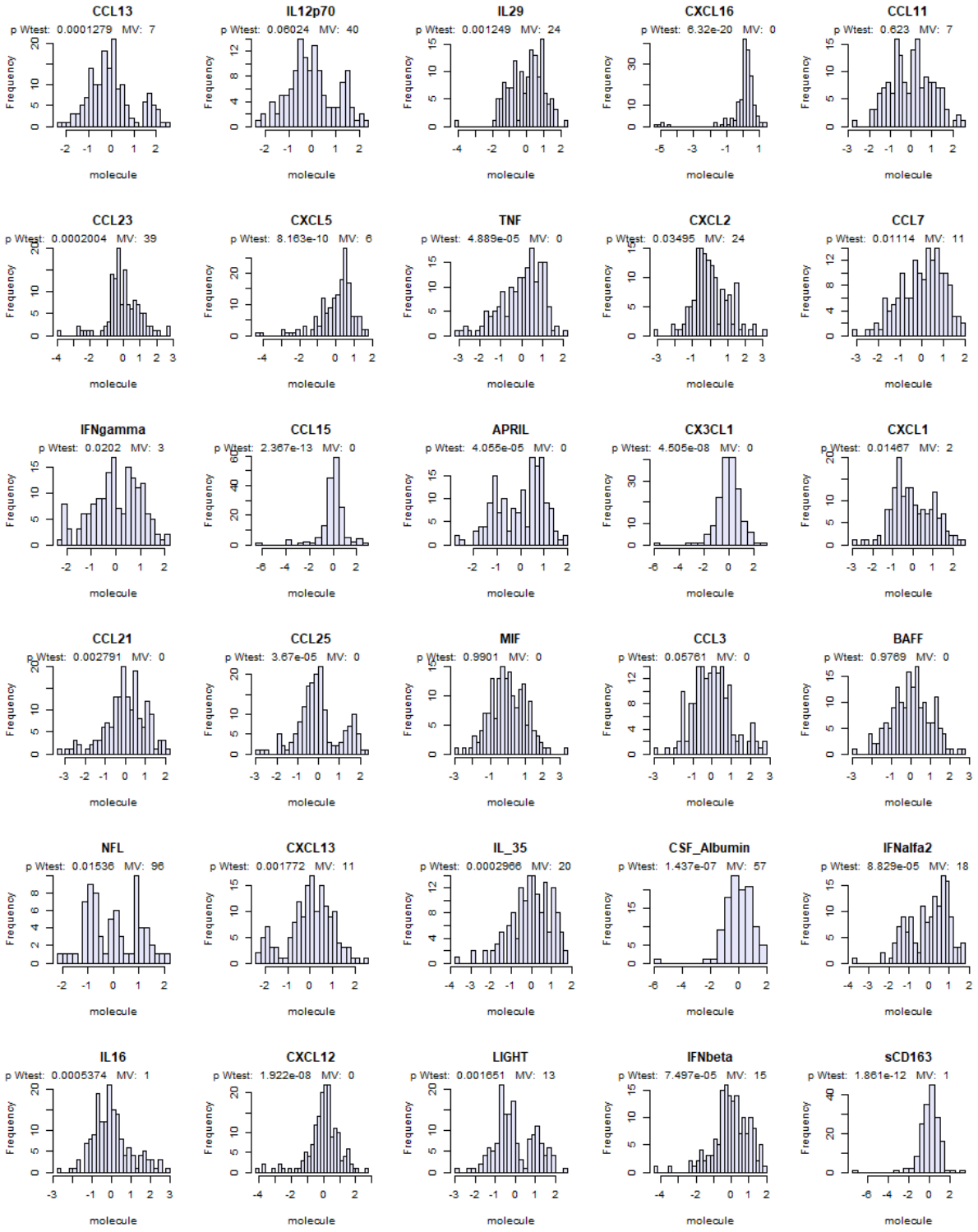


Figure 28 Representation of base-2 logarithm of the variables from Subset1, together with the number of Missing Values and the result of the Saphiro-Wilk Test (W-test) for normality (p-value). IL_28 Fibrinogen and Parvalbumin have been excluded because of the too missing values to use the normality test.

The 84 molecules from the CSF have been ordered based on the number of MVs, and the first 70 molecules have been used to perform a correlation plot together with the NChPV. This selection made it possible to consider variables with at least 95 samples and more robust pairwise correlations. Adjusting p-values for multiple comparisons no significant correlation with NChPV have been found, but several significant correlations are found between the CSF variables (Figure 29).

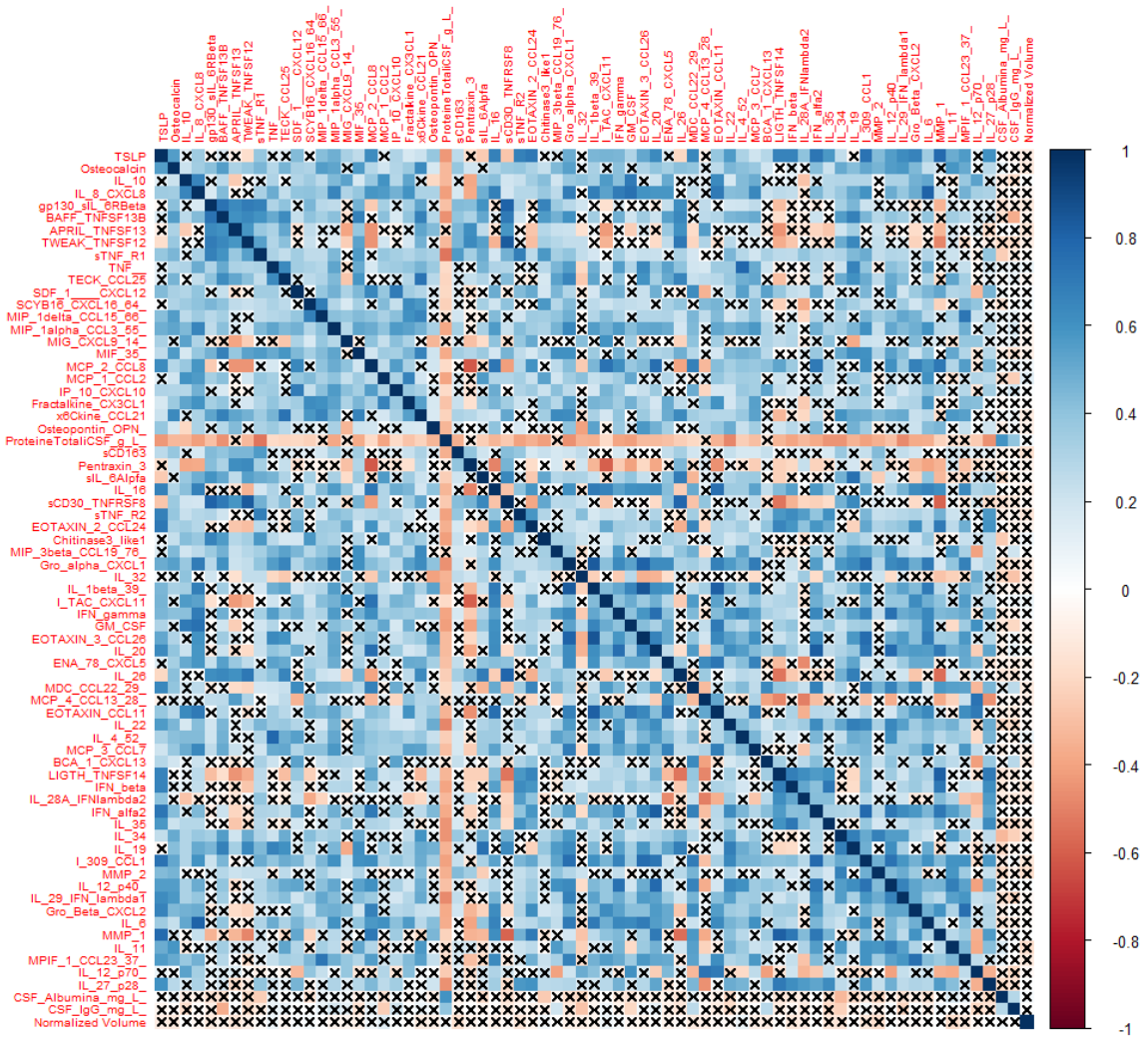


Figure 29 Correlation Matrix including the 70 molecules with the smaller number of missing values and Normalized Choroid Plexus Volume (ChPV). The X indicates non-significant correlation, considering p-values adjusted for multiple comparisons. NChPV shows no significant correlation, but some molecules are highly correlated.

4.4.1 Study of the full Cerebrospinal Fluid matrix

The 44 molecules in the corrplot represented in Figure 30 are the output of the selection described in section 3.6.3 in order to get a full matrix with the 80% of subjects (133) keeping the largest number of variables.

In addition to the 44 selected molecules, other 11 molecules have been included in the subset (Subset 2). These variables were included in the Subset1 and therefore considered of interest based on previous studies but were excluded from the selection based on MVs. The number of subjects included in Subset 2, deleting subjects with at least one MV, is then 84, and the number of molecules is 55.

The correlation analysis in this case shows different results: even if adjusting for multiple comparisons, using Pearson coefficient the following molecules resulted negatively correlated with NChPV: IL6beta, BAFF, APRIL, TWEAK, sCD163, IL6alfa, sCD30. Using Spearman, IL6beta and sCD163 are still negatively correlated with NChPV.

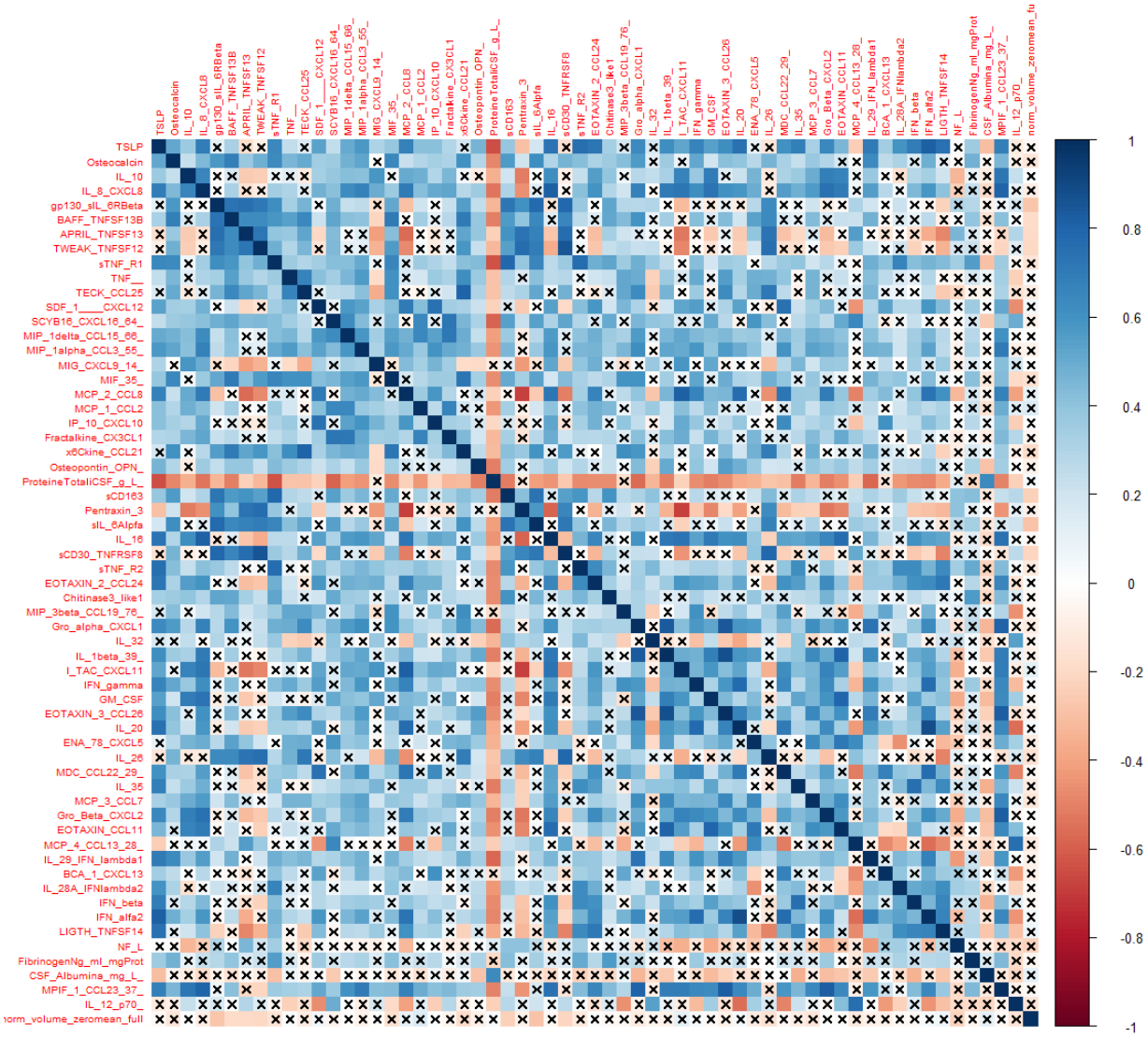


Figure 30 Correlation Matrix including Normalized Choroid Plexus Volume (NChPV) and the 54 molecules from Subset2, that includes the 44 molecules with the smaller number of missing values (MV) that keep the 80% of the subjects when removing subjects with MVs. The X indicates non-significant correlation, considering p-values adjusted for multiple comparisons. NChPV shows negative correlation with IL6beta, BAFF, APRIL, TWEAK, sCD163, IL6alfa, sCD30. Some molecules are highly correlated.

4.4.2 Single Molecule Analysis

NFL, Fibrinogen, Parvalbumin and CSF Albumin have a few available samples and couldn't be included in the regression analysis. TNF_R1/TNF_R2 (referred to as R1_R2) instead is a variable that wasn't present in the dataset as a ratio but was considered of interest from literature [39] and easy to compute. Therefore, they are processed as the others (with base 2-logarithm and scaling) and singularly analyzed to search for relationships with ChPV. For each of these variables it is presented the Pearson correlation with NChPV and the Multiple R^2 of the linear

regression model with age, considering that in a model with only age as predictor, R^2= 0.09 (Table 5). Only NFL showed a significant contribution in predicting NChPV.

	Correlation with Normalized Choroid Plexus Volume (r)	**R^2 of linear model with age**
Fibrinogen	0.05	0.15
Parvalbumin	0.01	0.09
NFL	-0.18	0.13.
CSF_Albumin	0.03	0.10
R1_R2	0.13	0.09

Table 5 Single-molecule analysis for variables with a large number of Missing Values (MV) and for TNF_R1/TNF_R2, a ratio between two available variables. No variable showed a relevant correlation with Normalized Choroid Plexus Volume (NChPV), and only NFL gave a significant contribution in predicting NChPV.

4.5 Selection of the relevant molecules to predict Choroid Plexus volume

Lasso Regression and StepWise Regression are the two methods used to filter the large number of CSF variables and to keep the most significant variables for predicting NChPV through linear regression models. In no linear model sex has been added as a predictor, because there was no significant difference in NChPV between the male and female (p = 0.41).

4.5.1 Selection based on Lasso Regression

As described in section 3.5.2, Lasso regression has been used to extract the most relevant variables among those with the larger number of samples to describe the NChPV, avoiding collinearity and optimizing linear regression models. It has been applied into three different sets of molecules: firstly, it has been applied on Subset2 before (Selection1) and after (Selection2) the addition of the 11 molecules from Subset1 that are excluded from Subset2, so it could apply an initial filter among molecules. In addition to this, it has been applied to the Subset1 (Selection3).

In the first case, considering the entire full matrix the following coefficients have not been pushed to zero and so identified relevant molecules are described as follows.

Considering the Selection1 among 44 variables (λ=0.073) Lasso has extracted the coefficients reported in Table 6, that are the outcome of Selection1, while the coefficients of the variables

not shown here have been pushed to zero. Among the reported coefficients, the molecules in bold belong to Subset1. However, generally, coefficients are low.

Molecule	**Coefficient**
IL_8_CXCL8	-0,09
gp130_sIL_6RBeta	-0,13
TNF	0,07
MCP_1_CCL2	0,16
sCD163	2e-05
sIL_6Alpfa	0,16

Table 6 Lasso coefficients from Selection1, therefore considering 44 molecules from 80% of subjects. Molecules in bold belong to Subset1. Coefficients have low values.

Considering the Selection2 among the entire Subset2 with 55 variables (λ=0.056), Lasso has extracted the 13 coefficients presented in Table 7, that are the outcome of Selection2, while the coefficients of the variables not shown here have been pushed to zero. Selection2 comprises a larger number of variables from Subset1 (molecules in bold) and shows higher coefficients. In particular, the most relevant molecules are TNF and CX3CL1.

Molecule	**Coefficient**
gp130_sIL_6RBeta	-0,26
TNF	0,34
CCL2	0,16
CX3CL1	-0,23
CCL21	-0,13
ProteineTotaliCSF_g_L	-0,12
sCD163	-0,03
sIL_6Alpfa	-0,16
Chitinase3_like1	0,17
IL35	-0,03

CCL7	-0,06
IL29_IFN_lambda1	-0,04
CXCL13	-0,06

Table 7 Lasso coefficients from Selection2 considering Subset2, and therefore 84 subjects. A large number of molecules from Subset1 are present (in bold), and coefficients are higher than Selection1.

Finally, for what concerns the last case, in Table 8 are shown the 11 Lasso non-zero coefficients from Selection3, considering 91 subjects and molecules from Subset1 without IL12p70, CCL23, NFL and CSF_Albumin because of the too large number of MVs, with $\lambda = 0.051$. The largest coefficients belong to TNF, CCL21, CCL15 and CXCL13.

Molecule	**Coefficient**
CXCL5	0,02
TNF	0,34
CCL7	-0,04
IFNγ	0,13
CCL15	0,17
APRIL	-0,12
CX3CL1	-0,14
CCL21	-0,36
CXCL13	-0,16
IL35	-0,10
sCD163	-0,02

Table 8 Lasso coefficients considering Subset1 (Selection3), without IL12p70, CCL23, NFL and CSF_Albumin. The largest coefficients belong to TNF, CCL21, CCL15 and CXCL13.

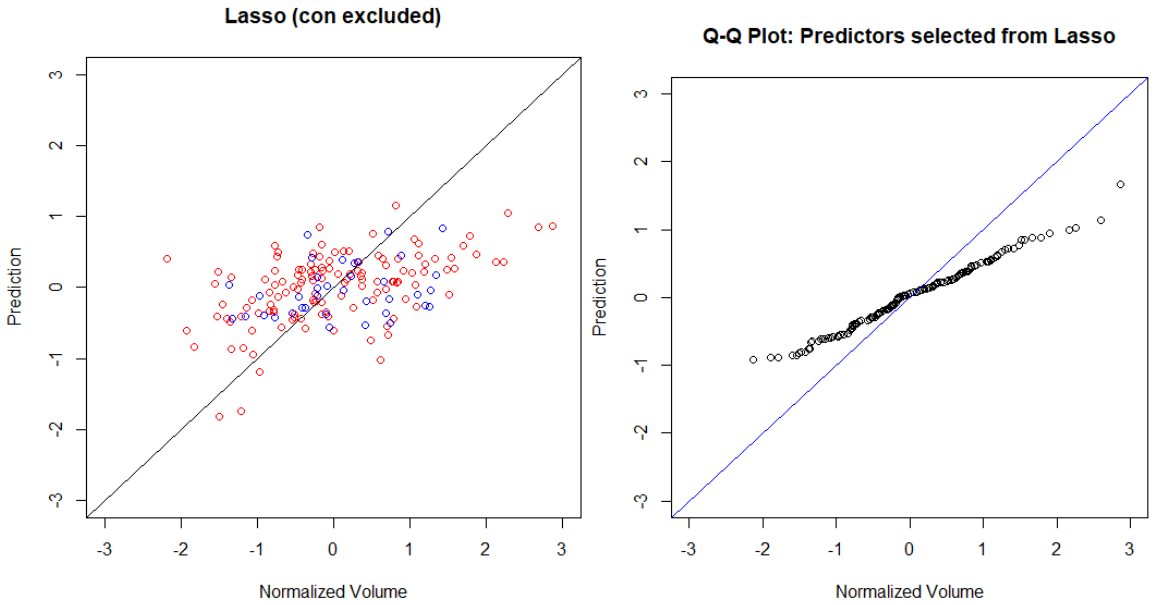
These subsets of molecules from Selection2 and Selection 3 are used in linear regression models together with age for trying to predict NChPV volume.

The linear model with predictors derived from the Lasso Regression with 54 molecules from 84 subjects in Table 7 (Selection2), applied to all the 167 subjects, reaches a Multiple R^2 = 0.277 (Adjusted R^2 = 0.232) (Figure 31). CXCL13, IL_29_IFN_lambda1, sIL_6alfa, sCD163,

CX3CL1 and CCL21 are excluded because of the high values of the CV of the estimates. Age and gp130_sIL_6RBeta give the highest contribution, followed by CCL7, Chitinase3_like1, and TNF. When comparing the prediction of the linear model to the NChPV, the results show a bad fitting of the model, and a low adhesion of the prediction to the original data. Residuals are comparable to a gaussian distribution (Figure 31).

	Estimate	**Std. Error**	**P-value**	**$CV_{estimate}$ (%)**	**VIF**
Age	0,32	0,08	7.73e-05 ***	24.49	
CCL7	-0,27	0,09	0.00253 **	32.47	1.51
TNF	0,20	0,09	0.02862 *	45.18	1.60
gp130_sIL_6RBeta	-0,39	0,09	4.91e-05 ***	23.81	1.86
ProteineTotaliCSF	-0,17	0,09	0.07411 .	55.54	1.52
Chitinase3_like1	0,31	0,11	0.00482 **	34.87	2.25
CCL2	0,11	0,09	0.19950	77.55	1.64
IL35	-0,14	0,08	0.10229	60.77	1.38

Table 9 Summary of the model derived from Lasso Selection2



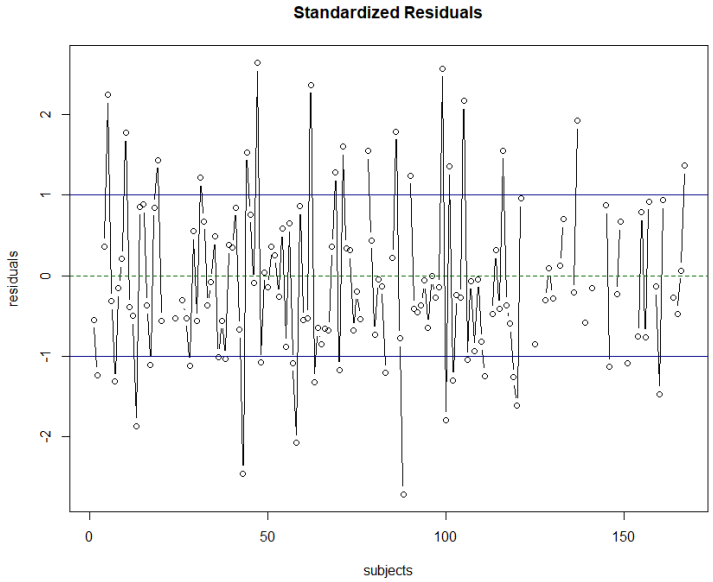


Figure 31 Upper level - Left: Scatter Plot of Normalized Choroid Plexus Volume (NChPV) vs the prediction of the Lasso regression model from Selection2 applied to the entire dataset; BT and BR are equally distributed. Right: Q-Q plot of NChPV compared to model prediction, that shows a bad fitting of the model. Lower Level: standardized residuals of the model.

In the second case, considering the Selection3 of molecules with non-zero coefficients in Lasso Regression applied in Subset1 (Table 8) applied to the entire dataset, the linear model reaches a Multiple $R^2 = 0.275$ (Adjusted $R^2 = 0.1939$). CXCL5 and sCD163 are excluded from the model because of the too large CVs. The predictors and their relative coefficients estimates are illustrated in Table 10. P-values, that indicates the contribution of the variables, are lower than the model before (Table 9), and the more significative are IFNγ, CCL15 and CXCL13. The fitting of the model is not good, and several subjects from BR site are excluded from the model because of MVs. Residuals are comparable with a gaussian distribution (Figure 32).

	Estimate	**Std. Error**	**P-value**	**$CV_{estimate}$ (%)**	**VIF**
Age	0,18	0,10	0.0835 .	57.13	
TNF	0,22	0,13	0.0961 .	59.45	1.66
CCL7	-0,18	0,15	0.2315	83.01	2.51
APRIL_TNFSF13	-0,23	0,12	0.0511 .	50.7	1.60
IFNγ	0,26	0,12	0.0359 *	46.95	2.18
CCL23	-0,31	0,16	0.0577 .	52.00	3.46
CCL15	0,37	0,15	0.0180 *	41.50	3.90
CXCL13	-0,24	0,10	0.0150 *	40.31	1.40

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
Age	0,18	0,10	0.0835 .	57.13	
CCL21	-0,21	0,18	0.2643	89.02	3.34
IL35	-0,15	0,10	0.1552	69.75	1.26

Table 10 Summary of the model derived from Lasso Selection3. The most significative molecules are IFNγ, CCL15 and CXCL13.

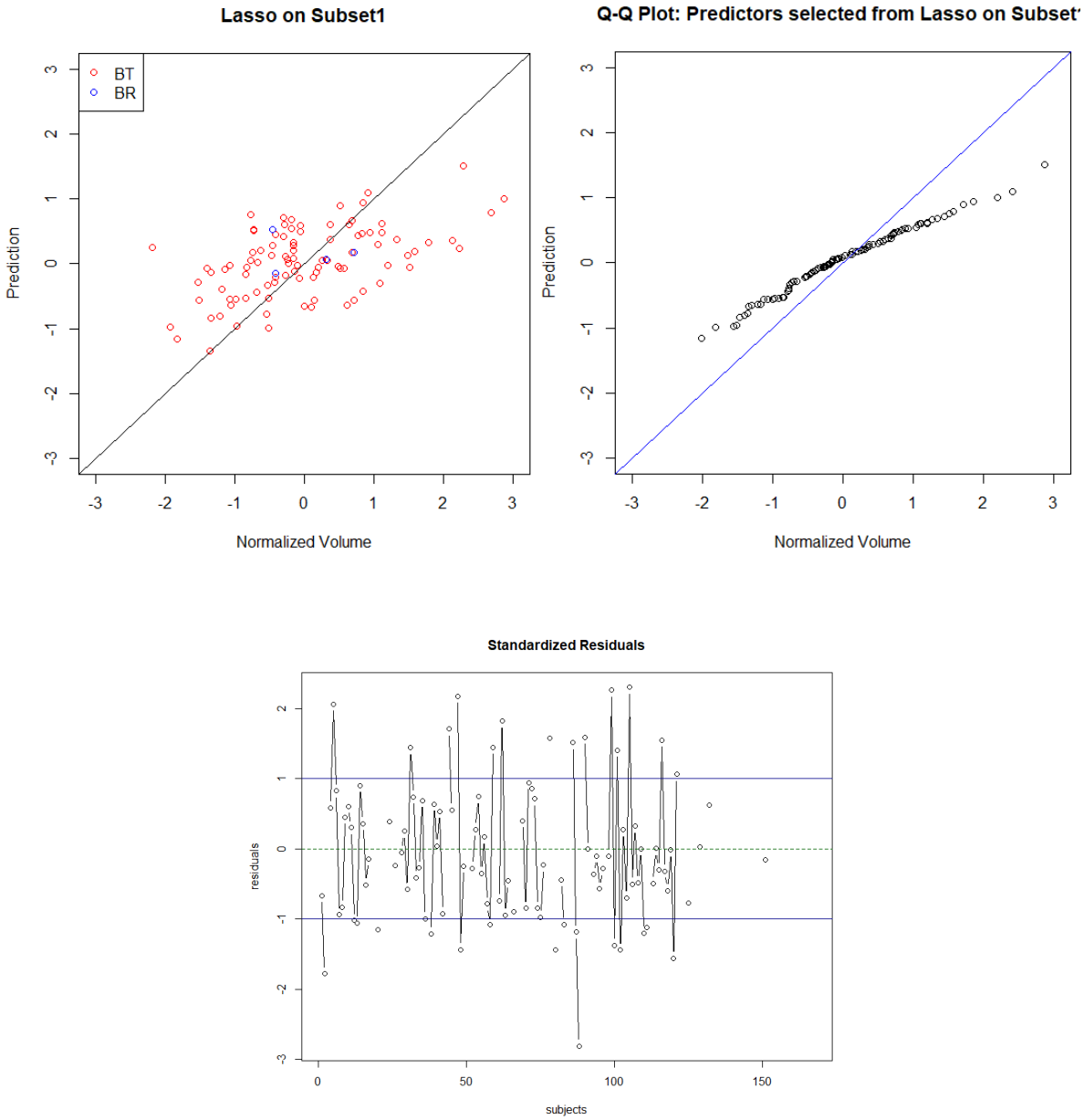
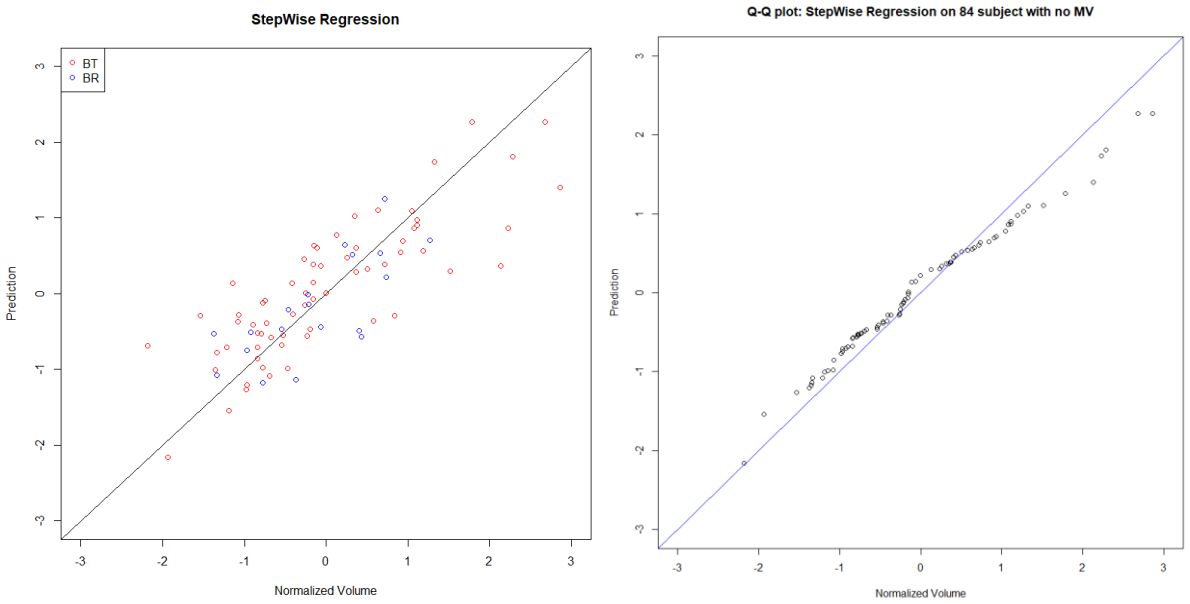


Figure 32 Upper Level – Left: Scatter Plot of Normalized Choroid Plexus Volume (NChPV) vs the prediction of the Lasso regression model from Selection3 applied to the entire dataset; BT and BR are equally distributed, but most of BR dataset is excluded. Right: Q-Q plot of NChPV compared to model prediction. Lower Level: standardized residuals of the model.

4.5.2 Selection based on Stepwise Regression

Stepwise regression was applied on 84 subjects with Subset2 variables and age as predictors. From the selected molecules from stepwise regression, 4 linear regression models have been implemented.

The Stepwise-resulting linear model has 22 regressors (Table 11) and reaches a Multiple R^2 = 0.656 (Adjusted R^2 = 0.534). The most relevant contribution is given by TNF, Pentraxin_3, CXCL11 and CCL11; also, CX3CL1, CCL19, CXCL2 and LIGHT have significant coefficients. In this case, there is a better correspondence between NChPV and the prediction of the model (Figure 33), the residuals are lower and comparable to a gaussian distribution. The CVs are in the acceptable range, but VIF are high.



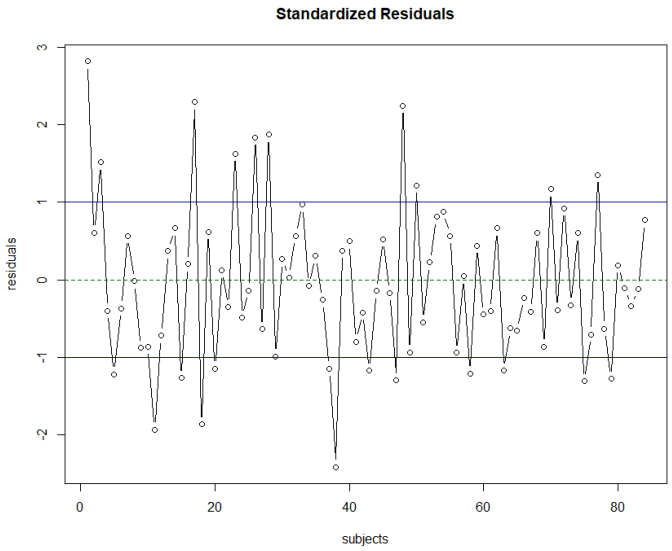


Figure 33 Upper Level – Left: Scatter Plot of Normalized Choroid Plexus Volume (NChPV) vs the prediction of the Stepwise regression model applied on 55 variables of 84 subjects; BT and BR are equally distributed. Right: Q-Q plot of NChPV compared to model prediction. Lower Level: standardized residuals of the model.

	Estimate	**Std. Error**	**P-value**	**$CV_{estimate}$ (%)**	**VIF**
TNF	0,66	0,14	1.36e-05***	21.16	
CXCL12	0,44	0,21	0.040925*	47.9	6.43
CX3CL1	-0,74	0,21	0.001027**	29.02	7.62
CCL21	-0,39	0,22	0.083733.	56.89	7.94
ProteineTotaliCSF	-0,27	0,18	0.142216	67.27	5.38
Pentraxin_3	1,32	0,32	0.000145***	24.69	17.60
sIL_6Alpfa	-0,29	0,17	0.102689	60.37	4.75
IL16	-0,41	0,26	0.118327	63.14	11.29
Chitinase3_like1	0,32	0,16	0.051311.	50.32	4.38
CCL19	-0,53	0,18	0.005090**	34.43	5.57
CXCL1	-0,62	0,28	0.031572*	45.46	11.90
IL32	-0,40	0,18	0.025992*	43.84	5.18
I_TAC_CXCL11	1,07	0,24	3.13e-05***	22.26	9.34
IFNγ	0,49	0,24	0.048586*	49.71	9.59
MDC_CCL22_29_	-0,37	0,21	0.088579.	57.8	7.49
CCL7	-0,36	0,17	0.036785*	46.86	4.64
CXCL2	-0,48	0,18	0.008525**	36.8	5.32

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
EOTAXIN_CCL11	1,65	0,38	4.72e-05***	22.85	23.69
CCL13	-0,53	0,29	0.076298.	55.47	14.12
IL29_IFNlambda1	-0,39	0,18	0.035916*	46.63	5.03
IFNalfa2	0,73	0,30	0.018218*	41.23	14.11
LIGTH_TNFSF14	-0,56	0,17	0.001773**	30.61	4.82

Table 11 Stepwise regression coefficients. CVs are in the acceptable range (<100%), but VIF are high (>10). The most relevant contribution is given by TNF, Pentraxin_3, CXCL11 and CCL11; also, CX3CL1, CCL19, CXCL2 and LIGHT have significant coefficients.

After adding age as a predictor, the model changes as follows (Table 12). The importance of each molecule in the linear model is similar to the previous model.

Age has a high CV, while the VIF values are similar to the previous model. So, before analyzing, the performances of the model, some variables have been excluded (Table 13).

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
Age	0,09	0,11	0,412311	121.14	
TNF	0,59	0,15	0,000148***	24.71	3,29
CXCL12	0,40	0,20	0,051686.	50.39	7,36
CX3CL1	-0,72	0,21	0,000922***	28.71	7,64
CCL21	-0,37	0,22	0,094876.	58.94	8,43
ProteineTotaliCSF	-0,28	0,18	0,121647	63.70	5,48
Pentraxin_3	1,25	0,31	0,000197***	25.23	17,67
sIL_6Alpfa	-0,22	0,18	0,220921	80.86	4,97
IL16	-0,40	0,25	0,112130	62.04	11,30
Chitinase3_like1	0,26	0,17	0,122642	63.88	4,45
CCL19	-0,47	0,19	0,014536*	39.75	5,66
CXCL1	-0,57	0,27	0,040792*	47.85	13,23
IL32	-0,38	0,17	0,029770*	44.94	5,24
I_TAC_CXCL11	1,04	0,23	0,000029***	22.11	9,44
IFNγ	0,47	0,23	0,046953*	49.31	9,80
MDC_CCL22_29_	-0,35	0,20	0,095224.	59.01	7,49
CCL7	-0,34	0,16	0,039249*	47.46	4,70

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
Age	0,09	0,11	0,412311	121.14	
CXCL2	-0,41	0,18	0,028010*	44.43	5,32
EOTAXIN_CCL11	1,56	0,36	0,000067***	23.37	23,73
CCL13	-0,51	0,28	0,076741.	55.54	14,45
IL29_IFNlambda1	-0,35	0,18	0,051111*	50.26	5,41
IFNalfa2	0,60	0,31	0,057043.	51.55	15,00
LIGTH_TNFSF14	-0,49	0,18	0,007702**	36.28	4,87

Table 12 Summary of the mode with Stepwise regression coefficients and age. CVs are in the acceptable range (<100%) except for age, but VIF are high (>10). The most relevant contribution is given by TNF, Pentraxin_3, CXCL11 and CCL11, CX3CL1 and LIGHT.

Given the large number of predictors and the high values of VIFs, the predictors have been reduced to 12. Age has been included as a predictor as well, getting a variable count of 13 (Table 13). The selection is based not only on CVs and VIF, but also considering the correlation between the molecules and the contribution to the model. The Multiple becomes $R^2 = 0.54$, while the Adjusted $R^2 = 0.46$. IFNα2, CCL13, IFNλ1, GroAlphaCXCL1, sIL6Alpha, IL16, ProteineTotali, CCL19, and GroBetaCXCL2 have been excluded from the model. Among the predictors, EOTAXIN_CCL11 and Fractalkine_CX3CL1 are the most significant, followed by age, CCL21, Pentraxin3, and CXCL11. Looking at Figure 34, it can be seen that the prediction has discrete adherence to the NChPV; residuals are worse than the original Stepwise model (Figure 33), but still acceptable.

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
Age	0.30	0.09	0.002074**	31.28	
TNF	0.31	0.13	0.019279*	41.76	2.52
CXCL12	0.36	0.17	0.037837*	47.25	4.21
CX3CL1	-0.78	0.20	0.000154***	25.01	5.94
CCL21	-0.52	0.17	0.002388**	31.74	4.02
Pentraxin_3	0.86	0.25	0.001199**	29.63	9.99
Chitinase3_like1	0.18	0.12	0.142725	67.47	2.35
CCL22	-0.47	0.19	0.013814*	39.61	5.43
IL32	-0.37	0.14	0.013016*	-39.25	3.23

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
Age	0.30	0.09	0.002074**	31.28	
CXCL11	0.71	0.21	0.001153**	29.52	6.84
IFNγ	0.33	0.16	0.044374*	48.86	3.99
CCL7	-0.31	0.15	0.035786*	-46.73	3.31
CCL11	0.79	0.23	0.000857***	28.72	8.01

Table 13 Summary of the optimized model with selected Stepwise regression coefficients applied to 84 subjects without missing values.

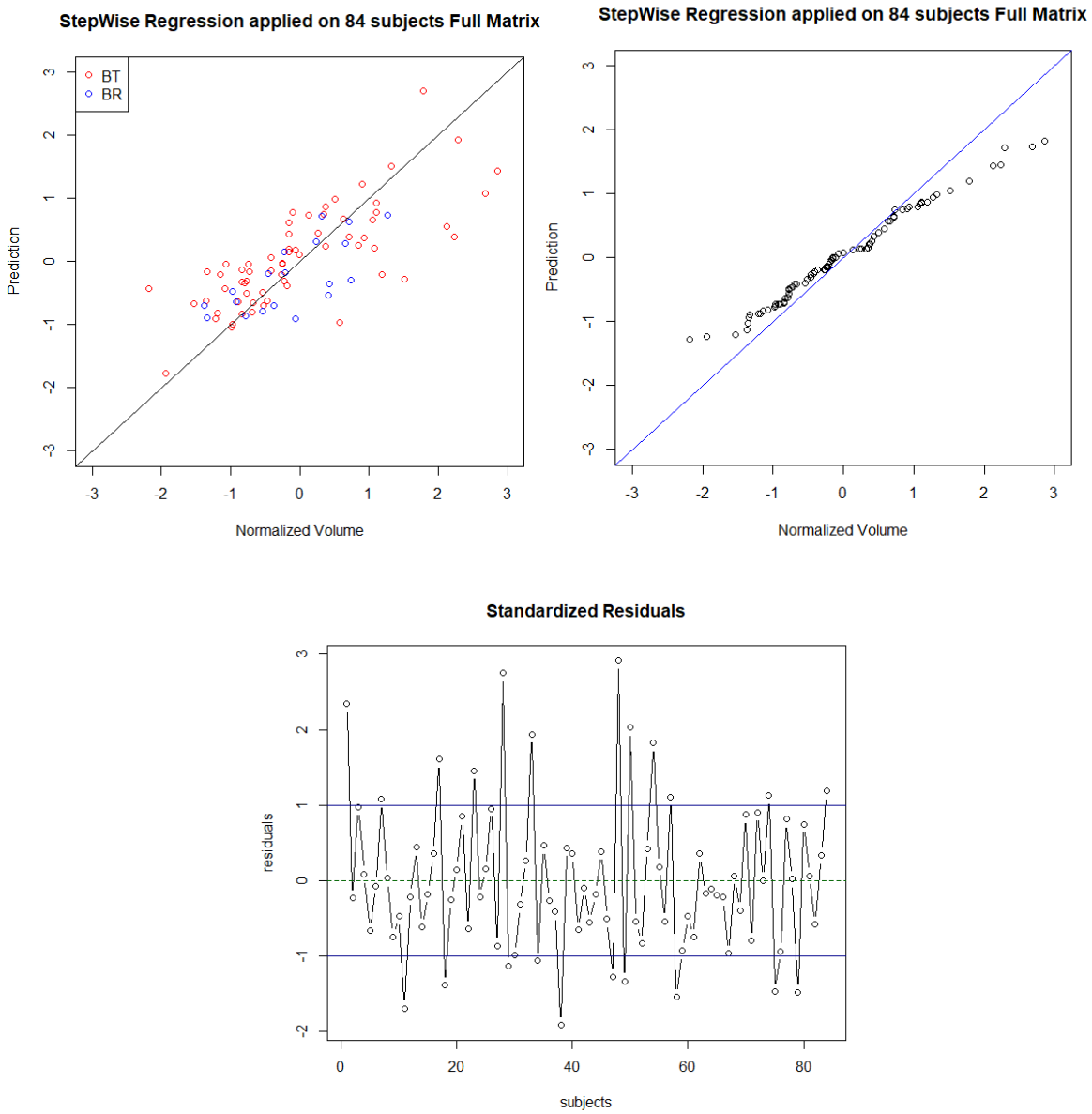


Figure 34 Illustrations of performances of the Stepwise regression optimized for all 84 subjects (R^2 = 0.357). Upper Level – Left: plot of normalized Choroid Plexus Volume (NChPV) vs prediction of the model on the same dataset, with a view of the two scanners (BT and BR). Right: Q-Q plot of the Stepwise regression model. Lower Level: standardized residuals.

The same model in Table 11 has been applied on the entire dataset of 167 subjects, with the addition of age as a predictor. It reaches a Multiple $R^2 = 0.440$ (Adjusted $R^2 = 0.281$) (Figure 35), but coefficients show high CVs (defined as CV ≥ 100) so before analyzing the performances the predictors number has been reduced.

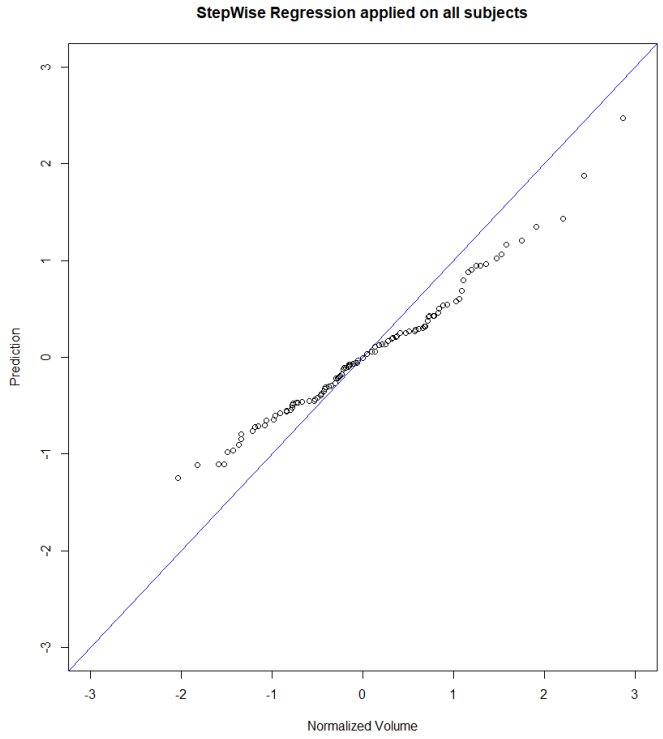


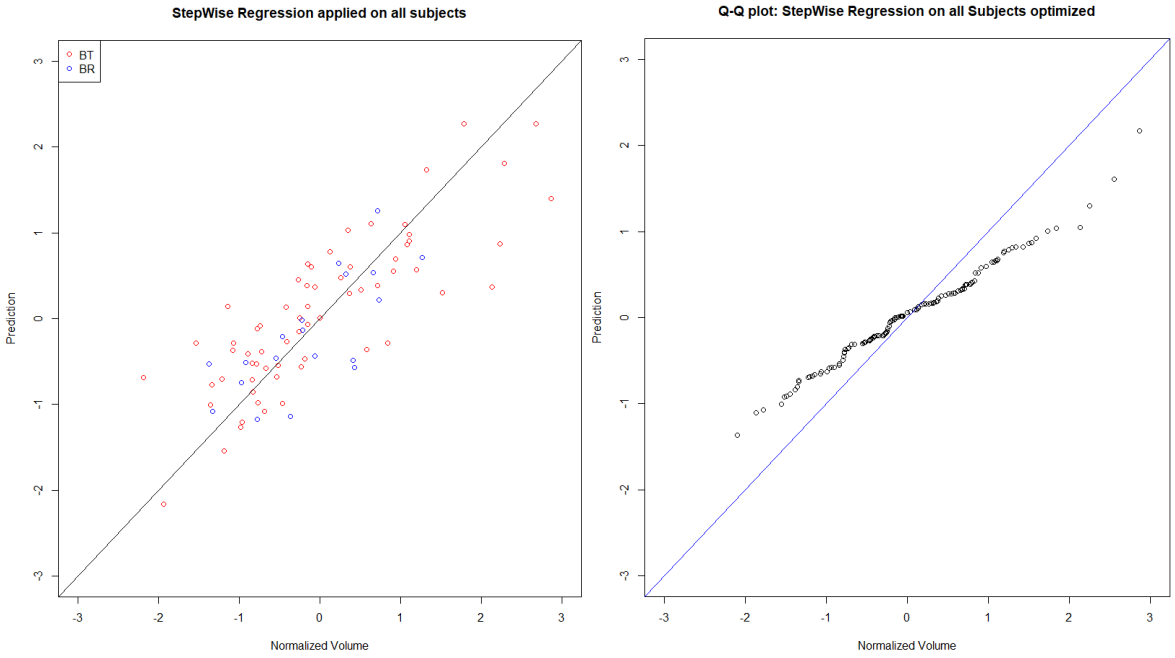
Figure 35 Q-Q plot of the Stepwise regression model applied to the entire dataset.

Based on the values of CVs and of VIF, the model predictors considering the entire cohort have been reduced to 10 (Table 14), returning a Multiple $R^2 = 0.357$ (Adjusted $R^2 = 0.295$). CVs and VIFs in this model have acceptable values (CV < 75%, VIF < 6), and the most relevant molecules in the model are Age, TNF, CCL21, CXCL11, CCL22, CCL7. The generalized model is still able to explain part of the NChPV variance even if it has lower performances in respect to when it is applied only on 84 subjects, and the standardized residuals are acceptable (Figure 36).

	Estimate	**Std. Error**	**P-value**	**$CV_{estimate}$ (%)**	**VIF**
Age	0,33	0,08	8.00e-05***	24.44	
TNF	0,37	0,12	0.00336**	33.39	2.12
CCL21	-0,46	0,17	0.00618**	35.85	4.51
Pentraxin_3	0,40	0,16	0.01427*	40.19	5.05

	Estimate	Std. Error	P-value	$CV_{estimate}$ (%)	VIF
Age	0,33	0,08	8.00e-05***	24.44	
IL16	-0,20	0,14	0.14667	68.43	3.81
IL32	-0,24	0,12	0.04933*	50.33	1.94
I_TAC_CXCL11	0,77	0,15	1.35e-06***	19.61	4.05
IFNγ	0,15	0,11	0.17088	72.57	2.17
MDC_CCL22_29_	-0,31	0,12	0.00866**	37.44	2.45
CCL7	-0,33	0,12	0.00602**	35.73	2.23
IL_29_IFN_lambda1	-0,20	0,11	0.07227.	55.12	2.21

Table 14 Summary of optimized Stepwise model for the entire cohort. The most relevant molecules in the model are Age, TNF, CCL21, CXCL11, CCL22, CCL7. VIF and CVs are in the acceptable range.



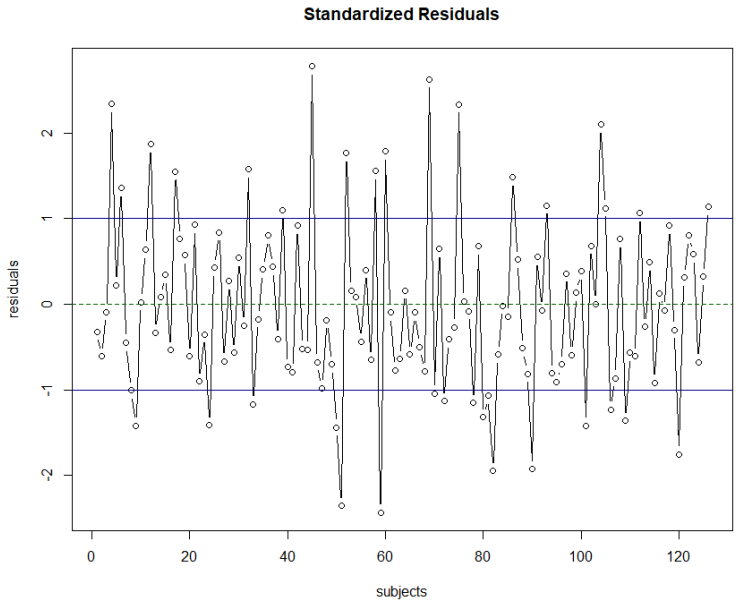


Figure 36 Illustrations of performances of the Stepwise regression model generalized and optimized for all 167 subjects (R^2 = 0.357). Upper Level - Left: plot of normalized Choroid Plexus Volume (NChPV) vs prediction of the model on the same dataset, with a view of the two scanners (BT and BR). Right: Q-Q plot of the Stepwise regression model. Lower Level: standardized residuals.

4.6 Logistic regression: IgG OCB

The detection of IgG_OCBs in CSF contributes to the diagnosis of MS (Section 1.2). Logistic regression is applied here to study the correlation between NChPV and the presence of IgG_OCB, and to test the power of the NChPV in predicting the IgG OCB value. IgG = 1 indicates the presence of the OCBs and therefore it is considered the positive group. The variable has 91 samples, of which 68 are positive and 23 are negative and the two groups don't show a significant difference in NChPV mean (t = 0.68677, df = 38.587, p-value = 0.4963). IgG_OCB of 52 subjects is correctly predicted (10 negatives and 42 positive).

The accuracy of the logistic regression model is 0.57, the sensitivity is 0.61 and the specificity is 0.43. The Confusion Matrix (Table 15) shows the results of predictions based on the logistic regression model:

		Reference	
		0	**1**
Prediction	**0**	10	26
	1	13	42

Table 15 Confusion matrix of the logistic regression model that predicts IgG OCB presence from Choroid Plexus Volume. IgG_OCB of 52 subjects is correctly predicted.

Figure 37 shows the AUC in the ROC curve, which illustrates the low performance of the model (AUC = 0.526), close to a random classification.

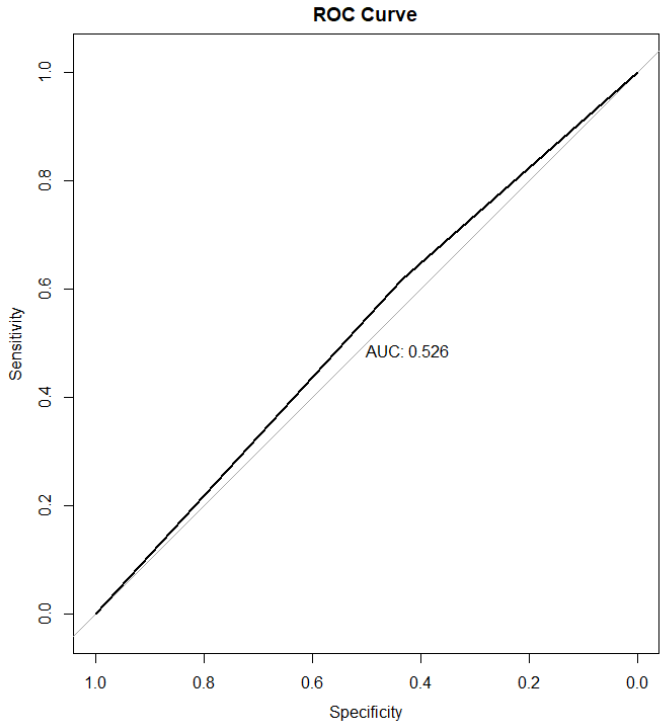


Figure 37 ROC curve of the logistic regression model that tries to predict the presence or absence of IgG OCB in cerebrospinal fluid (CSF). AUC = 0.53

5. DISCUSSION

In this section there is a critical analysis of the results reported in the previous chapter. It starts with a discussion about the output of the automatic segmentation of ChP and the results of the segmentation processing, followed by comments about the outcomes of the study of the relationship between NChPV and the CSF derived biomarkers.

5.1 Results of the Choroid Plexus segmentation analysis

The results achieved in this first step regarding the ChP segmentation lead to the following arguments. The first main observation to do on segmentations performed by the DNN is that a part of the subject population showed an inclusion of the hippocampus in the segmentation. This was noticed at first with a visual inspection of the ChP volumes superimposed on MRI images, both studying the mosaic images and then deepening with a more accurate analysis with ITK-SNAP three-dimensional view.

This pattern was highlighted then by the volume computation, because the ChPV was in general higher in patients with inclusion of hippocampus in the segmentation in the slices under the choroid plexus found through visual inspection. The confirmation has come with the transformation into the MNI space that allowed a comparison between ChP positions of the entire dataset, and in particular allowed to analyze in which position of the axial plane the first slice containing a voxel with ChP segmentation was. This analysis showed a clear bimodal pattern of the index of the inferior slice in the axial plane, where the first mode was formed by subjects whose segmentation included part of the hippocampus under the ChP, and the second was the part of subjects with correct ChP identification. This bimodal profile reflects the inclusion of the hippocampus in approximately half of the dataset. The bimodal distribution was encountered also in the range of the slices in which the segmentation was present, but not in the superior slices containing ChP, which were more uniform across the subjects and characterized by a low variability. Hence comes the decision to apply a threshold and “cut” the segmentation to exclude the voxels of MNI slices belonging to the hippocampus.

The inclusion of the hippocampus has been observed in almost all subjects from the dataset BR+ALT and a minor part of the BT dataset, resulting in a systematic bias that also influences the NChPV. The cause has been sought in previous studies of the DNN ensemble, in particular in the training set of the algorithm, and the same bias in MNI inferior slices has been observed in ground truth segmentation performed by the two different trained radiologists. This is the

cause of the different performance of the ensemble in the two new datasets, and confirms the difficulties in ChP segmentation also from trained personnel and the need of a robust automatic segmentation method.

Other minor issues regarding the ChP segmentations are the presence of some voxels belonging to ChP not included in the segmentation, or vice versa, some voxels out of ChP included in the segmentation. They are considered not affecting the ChPV in a significant way and therefore no correction has been made. In general, apart from the inclusion of the hippocampus and one outlier, the algorithm could recognize and segment correctly the ChP.
After the threshold the segmentations were more anatomically correct, but the difference between the two scanners was still statistically significant, especially when dividing for TIV and so considering NChPVs; therefore, the NChPV has been harmonized through NeuroCombat using BT Dataset as a reference for removing differences due to multiple-site acquisitions. In this way, subjects with less reliable information were processed to have a similar distribution of the reference site. Volumes derived from the harmonization are used to study relationships between NChPV and CSF biomarkers.

5.2 Results of the statistical analysis on the relationship between Cerebrospinal Fluid derived molecules

The first aim of this study regarding the CSF dataset was to have a clear overview of the statistical characteristics of the variables. For each variable, the number of MVs and therefore of the remaining available samples has been calculated, and the molecules with less than 30 samples were excluded from following analyses. Among the 84 variables left, there were also 30 molecules, included in the so-called Subset1, that in previous studies have been found connected with some other features of MS like the number of lesions or cortical thickness. Besides them, also Parvalbumin, Fibrinogen, NFL and CSF Albumin are related to the mechanisms of MS, but they had a too small number of samples to perform a statistical and linear regression analysis and therefore they have been studied separately. Only a small contribution of NFL has been found when trying to predict NChPV, the profile of the 84 variables have been studied to decide whether to consider them normally distributed or to apply some processing. No molecule results normally distributed using ShapiroWilk and Lillietest, so the base-2 logarithm has been implemented, which is common in studying biological variables. In this case, even if the tests for normality have returned only a few positive results, through a visual inspection it was decided to consider all of them suitable for the scaling through the Z-

score; age and NChPV result normally distributed. It was thus obtained a dataset with zero mean and unitary variance that allowed us to compare variables between each other and with NChPV.

Correlation between variables was computed using Pearson coefficient and adjusting p-values for multiple comparisons; it was found very high in some cases, like among TNF-related molecules, but no significant correlation was found between any molecule and NChPV except when considering Subset2 that showed negative correlations with some biomarkers.

5.3 Results on the regression analysis: relationship between Cerebrospinal Fluid derived molecules and Choroid Plexus volume

The principal objective of this thesis is to find some molecules from the large initial CSF dataset that are related to the NChPV. In particular, it's been investigated if molecules from Subset1, that are certainly related to MS disease mechanisms, are also selected from regression methods among all the other molecules. In other words, an attempt was made to derive the same molecules from literature, or rather hypothesis-driven variables, also extracting the most relevant variables within a large dataset using statistical shrinkage methods. The correspondence between hypothesis-driven selection and data-driven selection would imply a robust relationship between the extracted molecules and NChPV, meaning a connection between inflammation of the ChP and the physiological pattern of the aforementioned molecules.

The first shrinkage tool used to select the most relevant molecules for predicting the NChPV was the Lasso regression. In order to deal with the high number of MVs in some molecules, this method was applied only in the 80% of the subjects and only with 44 variables: this under-sizing of the dataset was performed in a way that the remaining data was a full-matrix without MVs and with the highest number of variables taken into account.

This new smaller dataset has been enlarged with 10 molecules that were excluded from the selection but included in Subset1 and therefore considered of interest from previous literature. TNF and sCD163 are the only Subset1 molecules that are present both before and after the addition of the 10 molecules. CX3CL1, CCL21, IL35, CCL7, IL29(IFNλ1) and CXCL13 have non-zero coefficients in the second case. The same method was applied on Subset1 to see which variables were highlighted among the interesting variables. The biggest coefficients belong again to TNF and CCL21, followed by CXCL13, CCL15, CX3CL1 and IFNγ.

The resulting molecules from Lasso regression were used as predictors in linear regression, together with age, to predict the NChPV. The models have been designed in a way that estimates CVs and VIF values, that account for collinearity between molecules, would be in an accessible range (respectively, <100% and <10). In general, performances of these linear models are low; despite this, some molecules give a significant contribution to predicting NChPV. Considering the predictors from Subset2, age, IL6Beta, Chitinase3, CCL7 and TNF give the major contribution. The selection from Subset1 variables has a worse performance, and the major contribution belongs to IFNγ, CCL15 and CXCL13. In this case, only a few samples from BR have been used, due to MVs.

Stepwise regression is the second method used to select the most meaningful variables to assess a link to NChPV. It is applied on the 84 subjects from the Subset2 pool of molecules, and in general it produces better results than Lasso Regression. The initial selection comprises 22 variables (of which 11 belong to Subset1) and reaches an R^2=0.65 but has high VIF values and so high collinearity between variables. When adjusting VIF and CVs and adding age, the model includes 13 predictors (of which 6 belong to Subset1) and reaches an R^2=0.54. The major contributions belong to Age, CX3CL1, CCL21, Pentraxin3, CXCL11 (a chemokine that links microbial stimuli to intestinal inflammation [57]) and CCL11; again, TNF and IFNγ give significant contributions.

The Stepwise model has been generalized also for the whole dataset. Reducing the predictors to get VIFs and CVs in the acceptable range, the model reaches an R^2 = 0.36 using 10 molecules, including the most significant TNF, CXCL11, CCL21, CCL22 (a macrophage derived chemokine [58]) and CCL7.

The last model has been used on IgG OCBs, a biomarker that contributes to the diagnosis of MS. Initially, investigating significant differences between positive and negative OCBs, no results have been found. Performing logistic regression for predicting the presence or the absence of OCBs through NChPV values, accuracy is low (0.57), as well as sensitivity (0.61) and specificity (0.43).

It can therefore be concluded that even if we couldn't find a specific subset of variables able to predict with a good accuracy the ChPV, StepWise and Lasso Regression identified a relationship between NChPV and some molecules that are connected to the inflammatory MS state. Taking into account the most recurring variables, we can observe for example that TNF has well-established roles in innate and adaptive immunity and in physiological functions of immune cells [35]; CCL21 as well is involved in inflammatory and immunoregulatory processes [59] ; CX3CL1 is a chemokine that participates in vascular inflammation [60];

CCL11 is related in some diseases to neurodegeneration and lowered neurogenesis [61]; CCL7 recruits leukocytes to infected tissues for immune response mediation [62]; Pentraxin3 is an innate immunity pattern recognition molecule [38]. The similarity between hypothesis-driven and data-driven molecules brings to the conclusion that the ChPV is related to the inflammatory processes of MS, and this makes it a good candidate as a MS biomarker.

6. CONCLUSION

Multiple sclerosis is a chronic inflammatory disease that affects more than two million of people of all ages. It manifests with a broad range of symptoms that make the diagnosis time-consuming and that imply a need for new disease biomarkers.
Patients affected by multiple sclerosis present not only lesions in MRI scans, but also altered levels of cytokines and chemokines in the cerebrospinal fluid that are related to proinflammatory patterns or to immune response.
Choroid Plexus, a vascular tissue inside ventricles, is the main producer of the cerebrospinal fluid, and recently it is increasing interest as a new potential biomarker for multiple sclerosis diagnosis and progression. This is due to its role in immune cell trafficking between blood and cerebrospinal fluid, and the capability to provide a proxy to possibly quantify inflammatory alterations non-invasively computing its volume..

The aim of this thesis was to study the relation between the choroid plexus volume and a large dataset of molecules extracted from liquor samples in patients from two different acquisition sites. To do so, the first step was to validate ASCHOPLEX, a new approach that obtains Choroid Plexus segmentation through a deep learning ensemble; the second objective was to analyze the dataset of molecules from the cerebrospinal fluid, to understand the correlations between each other and process data to make them suitable for the subsequent analysis. Finally, it has been investigated how to select the most relevant molecules in predicting the Choroid Plexus volume.

The first issue to be considered is the necessity to validate deep learning methods for Choroid Plexus segmentation for every scanner type. In this case, segmentation of Choroid Plexus underwent a quality check to evaluate its performances that highlighted a systematic bias between the two scanners due to the partial inclusion of the hippocampus near the temporal horn of the ventricles, located in the inferior portion of the segmentation. The distribution of this error has been studied also in the standard MNI space, and it has been solved by applying a threshold to remove the voxels that didn’t belong to the Choroid Plexus. This operation corrected the segmentations but didn’t solve the bias between scanners, which has been fixed using a harmonization method on the volume.

The main outcome of this study is the presence of a connection between the Choroid Plexus volume and several molecules in cerebrospinal fluid involved in the multiple sclerosis proinflammatory patterns. Lasso and Stepwise regression have been used to select the most relevant variables and to avoid the inclusion of highly correlated molecules in linear models. Even if there is no relevant correlation between molecules and Choroid Plexus volume, and even if models couldn't explain it with elevated performances, several variables used in different combinations have a significant predictive power. Moreover, variables extracted from the aforementioned selected methods include molecules already found in literature to be linked to multiple sclerosis mechanisms and are all connected with inflammatory processes or to the recruitment of immune cells. These molecules extracted among the entire dataset through a statistical analysis is consistent with previous studies of multiple sclerosis potential biomarkers, and it is consistent also with the hypothesis of involvement of the Choroid Plexus in the inflammatory process.

To conclude, the selection of proinflammatory molecules by the models might reflect the physiological role of Choroid plexus in multiple sclerosis, and therefore it suggests that Choroid Plexus volume, thanks to its involvement in inflammation process should be further investigated as a potential multiple sclerosis biomarker.

6.1 Future Directions

Choroid Plexus volume has been obtained using a deep neural network ensemble, but the segmentations needed a subsequent processing to be used in the analysis due to the bias between the scanners. This matter raises some questions on why deep learning methods present these difficulties in generalizing for different scanners and how these differences between scanners can be solved.

Statistical and regression analysis in this study have been performed on a real dataset with a large number of missing values that have been easily excluded from the models. Methods of imputation can be applied in future to handle missing values and to better exploit all the information from data.
Moreover, linear models are the most simplistic and it might be too reductive to represent the complex relationship between molecules and their possible connection to the increasing of

choroid plexus volume. Further studies are necessary to investigate more complex formulations that could describe more precisely the inflammation pattern related to multiple sclerosis; bioinformatic models could be a potential tool for creating networks of molecules and understanding the biological mechanisms underlying choroid plexus enlargement its relationship with and increased concentrations of some molecules in cerebrospinal fluid.

LIST OF ACRONYMS

AI	Artificial Intelligence
ANTs	Advanced Normalization Tools
BBB	Blood Brain Barrier
BCSFB	Blood Cerebrospinal Fluid Barrier
ChP	Choroid Plexus
ChPV	Choroid Plexus Volume
CIS	Clinically Isolated Syndrome
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
DNN	Deep Neural Network
GM	Gray Matter
HC	Healthy Control
IFN	Interferon
IL	Interleukin
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSE	Mean Square Error
NChPV	Normalized Choroid Plexus Volume
NFL	Neurofilament Light
OCB	Oligoclonal Bands
RSS	Residual Sum of Squares
TIV	Total Intracranial Volume
TNF	Tumor Necrosis Factor
VIF	Variance Inflation Factor
WM	White Matter

BIBLIOGRAPHY

[1] J. Yang, M. Hamade, Q. Wu, Q. Wang, R. Axtell, S. Giri and Y. Mao-Draayer, "Current and Future Biomarkers in Multiple Sclerosis," *Int J Mol Sci,* p. 23(11):5877, 2022 May.

[2] D. Taft, M. Ehsan and K. Xixis, Multiple Sclerosis, Treasure Island (FL): StatPearls, 2022.

[3] B. Lo Sasso, L. Agnello, G. Bivona, C. Bellia and M. Ciaccio, "Cerebrospinal Fluid Analysis in Multiple Sclerosis Diagnosis: An Update.," *Medicina (Kaunas),* p. 4;55(6):245, 2019 Jun.

[4] H. Lassman, "Pathogenic mechanisms associated with different clinical courses of multiple sclerosis," *Frontiers in Immunology,* p. 9(3116), 2019.

[5] A. Thompson, B. Banwell, F. Barkhof, W. Carroll, T. Coetzee and Cohen, "Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria," *Lancet Neurol,* pp. 17(2):162-173, 2018.

[6] F. Deisenhammer, H. Zetterberg, B. Fitzner and U. Zettl, "The Cerebrospinal Fluid in Multiple Sclerosis," *Front Immunol,* p. 12;10:726, 2019 Apr.

[7] D. Smith, C. Johanson and R. Keep, "Peptide and peptide analog transport systems at the blood-CSF barrier," *Adv Drug Deliv Rev.,* pp. 14;56(12):1765-91, 2004 Oct.

[8] V. Ricigliano and B. Stankoff, "Choroid plexuses at the interface of peripheral immunity and tissue repair in multiple sclerosis," *Curr Opin Neurol,* pp. 36(3):214-221, 2023 Jun.

[9] R. Spector, R. Keep, S. Robert Snodgrass and Smith QR, "A balanced view of choroid plexus structure and function: Focus on adult umans.," *Exp Neurol.,* pp. 267:78-86, 2015 May.

[10] S. Carloni, A. Bertocchi, S. Mancinelli, M. Bellini, M. Erreni, M. Borreca, D. Braga, S. Giugliano, A. Mozzarelli, D. Manganaro, D. Fernandez Perez, F. Colombo, A. Di Sabatino, D. Pasini, G. Penna, M. Matteoli, S. Lodato and M. Rescigno, "Identification of a choroid plexus vascular barrier closing during intestinal inflammation.," *Science,* vol. 374(6566), p. 439–448, 2021.

[11] H. Damkier, P. Brown and J. Praetorius, "Cerebrospinal fluid secretion by the choroid plexus," *Physiological Reviews,* vol. 93(4), pp. 1847-1892, 2013.

[12] p. Lizano, O. Lutz, G. Ling, A. Lee, S. Eum, J. Bishop, S. Kelly, O. Pasternak, B. Clementz, G. Pearlson, J. Sweeney, E. Gershon, C. Tamminga and M. Keshavan, "Association of Choroid Plexus Enlargement With Cognitive, Inflammatory, and Structural Phenotypes Across the Psychosis Spectrum.," *American Journal of Psychiatry,* vol. 176(7), pp. 564-572, 2019.

[13] S. Balasu, S. Brkic, C. Libert and Vandenbroucke, "The choroid plexus-cerebrospinal fluid interface in Alzheimer's disease: More than just a barrier.," *Neural Regeneration Research,* vol. 11, pp. 534-537, 2016.

[14] A. Reboldi, C. Coisne and D. Baumjohann, "C-C chemokine receptor 6-regulated entry of TH-17 cells into the CNS through the choroid plexus is required for the initiation of EAE," *Nat Immunol,* pp. 10(5):514-523, 2009.

[15] B. Engelhardt, K. Wolburg-Buchholz and H. Wolburg, "Involvement of the Choroid Plexus in Central Nervous System Inflammation," *Microspocy Research and Technique,* vol. 52(1), p. 112–129, 2001.

[16] V. Ricigliano, E. Morena, A. Colombi and et al., "Choroid Plexus Enlargement in Inflammatory Multiple Sclerosis: 3.0-T MRI and Traslocator Protein PET Evaluation," *Radiology,* pp. 301(1):166-177, 2021.

[17] V. Fleischer , G. Gonzalez-Escamilla, D. Ciolac and et al., "Traslational value of choroid plexus imaging for tracking neuroinflammation in mice and humans.," *Proc Natl Acad Sci U S A,* p. 118(36), 2021.

[18] M. Vercellino, B. Votta, C. Condello and et al., "Involvement of the choroid plexus in multiple sclerosis autoimmune inflammation: a neuropathological study.," *J Neuroimmunol,* pp. 199(1-2):133-141, 2008.

[19] P. Yushkevich, J. Piven and H. Hazlett, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage,* vol. 31, pp. 1116-1128, 2006.

[20] B. Fischl, "FreeSurfer," *NeuroImage,* vol. 62, pp. 774-781, 2012.

[21] V. Ricigliano, C. Louapre, E. Poirion, A. Colombi, A. Yazdan Panah and B. Stankoff, "Imaging Characteristics of Choroid Plexuses in Presymptomatic Multiple Sclerosis: A Retrospective Study.," *Neurol Neuroimmunol Neuroinflamm.,* 2022 Oct.

[22] J. Müller, T. Sinnecker, M. Wendebourg and et al., "Choroid Plexus Volume in Multiple Sclerosis vs Neuromyelitis Optica Spectrum Disorder," *Neurology - Neuroimmunollogy Neuroinflammation,* p. 9(3):e1147, 2022.

[23] N. Bergsland, M. Dwyer, D. Jakimovski, E. Tavazzi, R. Benedict, B. Weinstock-Guttman and R. Zivadinov, "Association of Choroid Plexus Inflammation on MRI Clinical Disability Progression Over 5 Years in Patients With Multiple Sclerosis," *Neurology,* pp. 100 (9) e911-e920, 2023.

[24] E. Tadayon, B. Moret and G. Sprugnoli, "Improving choroid plexus segmentation in the healthy and diseased brain: relevance for tau-PET imaging in dementia.," *J Alzheimers Dis.,* vol. 74(4), pp. 1057-1068, 2020.

[25] S. Smith, Y. Zhang and M. Jenkinson, "Accurate, robust, and automated longitudinal and cross-sectional brain change analysis," *Neuroimage,* vol. 17(1), pp. 479-489, 2002.

[26] S. Klistorner, M. Barnett, J. Parratt, C. Yiannikas, S. Graham and A. Klistorner, "Choroid plexus volume in multiple sclerosis predicts the espansion of chronic lesions and brain atrophy," *Ann Clin Transl Neurol,* pp. 9(10):1528-1537, 2022.

[27] M. Schmidt-Mengin, V. Ricigliano, B. Bodini and et al., "Axial multi-layer perceptron architecture for automatic segmentation of choroid plexus in multiple sclerosis," *Published Online,* 2021 Sept.

[28] L. Storelli, E. Pagani, M. Rubin, M. Maragoni, M. Filippi and M. Rocca, "A Fully Automatic Method to Segment Choroid Plexuses in Multiple Sclerosis Using Conventional MRI Sequences.," *J Magn Reson Imaging,* 2023.

[29] L. Zhao, X. Feng, C. Meyer and D. Alsop, "Choroid Plexus Segmentation Using Optimized 3D U-Net.," *Proceedings - International Symposium on Biomedical Imaging,,* p. 381–384, 2020.

[30] A. Yazdan-Panah, Schmidt-Mengin M. and V. Ricigliano, "Automatic segmentation of the choroid plexuses: method and validation in controls and patients with multiple sclerosis," *Neuroimage Clin.,* p. 38:103368, 2023.

[31] J. Jarrod, Eisma, D. Colin, McKnight and H. Kilian, "Deep learning segmentation of the choroid plexus from structural magnetic resonance imaging (MRI): validation and normative ranges across the adult lifespan," *PREPRINT (Version 1) available at Research Square,* 13 September 2023.

[32] N. Manouchehri and O. Stüve, "Choroid plexus volumetrics and brain inflammation in multiple sclerosis," *Proceedings of the National Academy of Sciences,* vol. 118(40), 2021.

[33] R. Magliozzi, O. Howell, R. Nicholas, C. Cruciani, M. Castellaro, C. Romualdi, S. Rossi, M. Pitteri, M. Benedetti, A. Gajofatto and F. Pizzini, "Inflammatory intrathecal profiles and cortical damage in multiple sclerosis.," *Ann Neurol.,* pp. 739-755, 2018.

[34] R. Janssens, S. Struyf and P. Proost, "The unique structural and functional features of CXCL12," *Cellular & Molecular Immunology,* vol. 15, pp. 299-311, 2018.

[35] J. Sethi and G. Hotamisligil , "Metabolic Messengers: tumour necrosis factor," *Nature Metabolism,* vol. 3, p. 1302–1312, 2021.

[36] L. Ivashkiv, "IFNγ: signalling, epigenetics and roles in immunity, metabolism, disease and cancer immunotherapy," *Nature Reviews Immunology ,* vol. 18, p. 545–558, 2018.

[37] H. Wang, K. Wang, X. Zhong, W. Qiu, Y. Dai, A. Wu and X. Hu, "Cerebrospinal fluid BAFF and APRIL levels in neuromyelitis optica and multiple sclerosis patients during relapse," *J Clin Immunol.,* vol. 32(5), pp. 1007-11, 2012.

[38] C. Tan, C. Burrus Lane and C. Eroglu, "Chapter Nine - Role of astrocytes in synapse formation and maturation," *Current Topics in Developmental Biology,* vol. 142, pp. 371-407, 2021.

[39] R. Magliozzi, F. Pezzini, M. Pucci, S. Rossi and F. Facchiano, "Changes in Cerebrospinal Fluid Balance of TNF and TNF Receptors in Naïve Multiple Sclerosis Patients: Early Involvement in Compartmentalised Intrathecal Inflammation.," *Cells,* p. 6;10(7):1712, 2021.

[40] R. Magliozzi, S. Hametner, F. Facchiano, D. Marastoni and M. Castellaro, "Iron homeostasis, complement, and coagulation cascade as CSF signature of cortical lesions in early multiple sclerosis," *Ann Clin Transl Neurol.,* p. 6(11): 2150–2163, 2019.

[41] T. Zhao, Z. Su, Y. Li, X. Zhang and Q. You, "Chitinase-3 like-protein-1 function and its role in diseases," *Signal Transduction and Targeted Therapy,* vol. 5, p. 201, 2020.

[42] MONAI, "MONAI: Medical Open Network for AI," Consortium MONAI. , 2022. [Online]. Available: https://monai.io/.

[43] "AUTO3DSEG," [Online]. Available: https://github.com/Project-MONAI/tutorials/tree/main/auto3dseg.

[44] Ronneberger O., P. Fischer and T. Brox, " U-Net: Convolutional Networks for Biomedical Image Segmentation.," p. https://arxiv.org/abs/1505.04597, 2015.

[45] F. Isensee, P. Jaeger, Kohl and J. Petersen, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation.," *Nature Methods,* vol. 18(2), p. 203–211, 2021.

[46] Hatamizadeh, A., Tang, Y., V. Nath, D. Yang, Myronenko, A and Landman, B, "UNETR: Transformers for 3D Medical Image Segmentation.," *Transformers for 3D Medical Image Segmentation.,* 2021.

[47] Ilya Loshchilov & Frank Hutter, "DECOUPLED WEIGHT DECAY REGULARIZATION," no. arXiv:1711.05101v3 [cs.LG], 2019.

[48] V. Fonov, A. Evans, K. Botteron, C. Almli, R. McKinstry, D. Collins and B. D. C. Group., "Unbiased average age-appropriate atlases for pediatric studies," *Neuroimage,* pp. 54(1):313-27, 2011.

[49] J. E. Iglesias, C. Liu, P. Thompson and Z. Tu, "Robust brain extraction across datasets and comparison with publicly available methods.," *IEEE Transactions on Medical Imaging,* vol. 30(9):, p. 1617–1634, 2011;.

[50] J. Kleesiek, G. Urban and D. Schwarz , "A. Deep MRI brain extraction: a 3D convolutional neural network for skull stripping," *NeuroImage,* pp. 460-469, 2016.

[51] J. Fortin, D. Tunc, D. Parker and T. Watanabe, "Harmonization of multi-site diffusion tensor imaging data," *Neuroimage,* pp. 1;161:149-170, 2017.

[52] J. Fortin, N. Cullen, Y. Sheline, W. Taylor and I. Aselcioglu , "Harmonization of cortical thickness measurements across scanners and sites.," *Neuroimage,* pp. 15;167:104-120, 2018.

[53] Y. Benjiamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing.," *Journal of the Royal Statistical Society Series B,* vol. 57, p. 289–300, 1995.

[54] T. L. Chap, Introductory Biostatistics, John Wiley & Sons, Inc., 2003.

[55] J. Kim, "Multicollinearity and misleading statistical results.," *Korean J Anesthesiol.,* pp. 72(6):558-569, 2019.

[56] Hastie and J. Trevor, The Elements of Statistical Learning : Data Mining, Inference, and Prediction., Springer, 2001.

[57] Z. Liu, X. Chen, X. Wang, X. Chen, C. Song, Y. Du, J. Su, S. Yaseen and P. Yang, "Chemokine CXCL11 links microbial stimuli to intestinal inflammation," *Clin Exp Immunol. ,* vol. 164(3), pp. 396-406, 2011.

[58] U. Yamashita and E. Kuroda, "Regulation of macrophage-derived chemokine (MDC, CCL22) production," *Crit Rev Immunol,* vol. 22(2), pp. 105-14, 2002.

[59] I. Comerford, Y. Harata-Lee, M. Bunting, C. Gregor, E. Kara and S. McColl, "A myriad of functions and complex regulation of the CCR7/CCL19/CCL21 chemokine axis in the adaptive immune system," *Cytokine & Growth Factor Reviews,* vol. 24(3), pp. 269-283, 2013.

[60] B. Jones, M. Beamer and S. Ahmed, "Fractalkine/CX3CL1: A Potential New Target for Inflammatory Diseases," *Mol Interv.,* vol. 10(5), p. 263–270, 2010.

[61] M. Ivanovska, Z. Abdi, M. Murdjeva, D. Macedo, A. Maes and M. Maes, "CCL-11 or Eotaxin-1: An Immune Marker for Ageing and Accelerated Ageing in Neuro-Psychiatric Disorders," *Pharmaceuticals (Basel),* vol. 13(9), p. 230, 2020.

[62] P. Menten, A. Wuyts and J. Damme, "Monocyte chemotactic protein-3," *Eur Cytokine Netw,* vol. 12(4), pp. 554-60, 2001.

[63] J. Mazziotta, A. W. Toga, A. Evans, P. Fox and J. Lancas, "A Probablistic Atlas of the Human Brain: Theory and Rationale for Its Development," *NeuroImage,* pp. 2:89-101, 1995.

[64] A. Weideman , C. Barbour , M. Tapia-Maltos, T. Tran, K. Jackson and P. Kosa, "New Multiple Sclerosis Disease Severity Scale Predicts Future Accumulation of Disability.," *Front Neurol.,* p. 10;8:598, 2017.