



UNIVERSIDADE ESTADUAL DE CAMPINAS  
Faculdade de Engenharia Elétrica e de Computação

Livia Maria de Aguiar Rodrigues

**Advancements in Hypothalamus Segmentation:  
Methodologies, Datasets, and Neurological  
Applications**

**Avanços na segmentação do hipotálamo:  
metodologias, conjuntos de dados e aplicações  
neurológicas**

**Campinas**

**2024**



UNIVERSIDADE ESTADUAL DE CAMPINAS  
Faculdade de Engenharia Elétrica e de Computação

Livia Maria de Aguiar Rodrigues

**Advancements in Hypothalamus Segmentation: Methodologies,  
Datasets, and Neurological Applications**

**Avanços na segmentação do hipotálamo: metodologias,  
conjuntos de dados e aplicações neurológicas**

Thesis presented to the Faculty of Electrical and Computer Engineering at the State University of Campinas as part of the requirements for obtaining the title of Doctor in Electrical Engineering, in the area of Computer Engineering.

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutora em Engenharia Elétrica, na Área de Engenharia da Computação.

Orientador: Profa. Dra. Leticia Rittner

Este exemplar corresponde à versão final da tese defendida pela aluna Livia Maria de Aguiar Rodrigues, e orientada pela Profa. Dra. Leticia Rittner

---

Campinas

2024

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca da Área de Engenharia e Arquitetura  
Rose Meire da Silva - CRB 8/5974

R618a Rodrigues, Livia Maria de Aguiar, 1990-  
Advancements in hypothalamus segmentation: methodologies, datasets,  
and neurological applications / Livia Maria de Aguiar Rodrigues. – Campinas,  
SP : [s.n.], 2024.

Orientador: Letícia Rittner.  
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de  
Engenharia Elétrica e de Computação.

1. Hipotálamo. 2. Segmentação de imagens. 3. Inteligência artificial -  
Aplicações médicas. 4. Doenças neurodegenerativas. 5. Ressonância  
magnética. 6. Aprendizado profundo. I. Rittner, Letícia, 1972-. II. Universidade  
Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação.  
III. Título.

Informações Complementares

**Título em outro idioma:** Avanços na segmentação do hipotálamo: metodologias, conjuntos  
de dados e aplicações neurológicas

**Palavras-chave em inglês:**

Hypothalamus

Image segmentation

Artificial intelligence - Medical applications

Neurodegenerative diseases

Magnetic resonance image

Deep learning

**Área de concentração:** Engenharia de Computação

**Titulação:** Doutora em Engenharia Elétrica

**Banca examinadora:**

Letícia Rittner [Orientador]

Hélio Pedrini

Gabriela Castellano

Nina Sumiko Tomita Hirata

Ricardo José Ferrari

**Data de defesa:** 06-06-2024

**Programa de Pós-Graduação:** Engenharia Elétrica

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-3476-6640>

- Currículo Lattes do autor: <http://lattes.cnpq.br/7821643415950245>

## COMISSÃO JULGADORA - TESE DE DOUTORADO

**Candidata:** Livia Maria de Aguiar Rodrigues      RA: 180545

**Data da Defesa:** 06 de junho de 2024

**Título da Tese:** Advancements in hypothalamus segmentation: methodologies, datasets, and neurological applications

Profa. Dra. Letícia Rittner

Prof. Dr. Hélio Pedrini

Profa. Dra. Gabriela Castellano

Profa. Dra. Nina Sumiko Tomita Hirata

Prof. Dr. Ricardo José Ferrari

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de PósGraduação da Faculdade de Engenharia Elétrica e de Computação.

*Dedico esta tese a minha família*

# Agradecimentos

O desenvolvimento desse trabalho, assim como minha jornada no doutorado não teria sido possível sem a ajuda e colaboração de diversas pessoas.

Inicialmente, agradeço à minha orientadora, prof. Letícia Rittner, por ter sido a melhor orientadora que eu poderia ter. Muito obrigada por ter sempre me apoiado (mesmo nas ideias mirabolantes), me guiado e ter mostrado o quanto eu sou capaz. Obrigada por ter sido paciente, por ter confiado em mim e por ter me acolhido e me entendido incontáveis vezes durante essa jornada dupla de pesquisa e maternidade. Sou imensamente grata a você. Se hoje sou uma pesquisadora mais madura e mais confiante, é porque você esteve me guiando da melhor forma possível durante todos esses anos.

Agradeço a meu marido, Alexandre, por ser sempre meu maior apoiador e por acreditar em mim quando nem eu acreditei. Obrigada por compartilhar comigo a vida, os dias bons e ruins, e por estar sempre disposto a me ajudar em absolutamente tudo (até mesmo em encontrar *bugs* nos códigos). Sem você eu jamais teria chegado até aqui. Te amo demais!

A meus pirralhinhos das bagunças unidas - Tutu, Dan, Liliz e Ga-, obrigada por trazerem tanta leveza, alegria e amor à minha vida. Em especial agradeço a meus filhos, Arthur e Daniel, por serem minha inspiração, por me fazerem sempre querer dar o melhor de mim e por serem a minha maior (e melhor) torcida. Vocês são minha luz.

Agradeço a meus pais, Cecília e Frederico, por tanto amor e apoio incondicional. Por compartilharem comigo a alegria de cada conquista e por estarem sempre dispostos a tudo pelo meu bem e meu sucesso (incluindo pegar um avião de última hora para que eu conseguisse terminar de escrever essa tese). Obrigada por serem tão entusiastas em relação a meu trabalho, por ouvirem cada detalhe com atenção e orgulho (mesmo quando não entendiam muito bem) e por serem

minha fonte de força. Se cheguei até aqui, é porque vocês trilharam boa parte do caminho para mim.

A meu irmão e irmã, agradeço todo amor, carinho, companheirismo. À Nana, especialmente, agradeço por ter me acolhido tantas vezes durante esses anos de pesquisa, ter vibrado comigo nas minhas conquistas e me acalentado quando as coisas não iam exatamente como o esperado. Sendo a professora-irmã maravilhosa que é, agradeço por ter me guiado emocionalmente e ser sempre uma fonte de inspiração para mim.

Às minhas cunhadas, Nandinha e Mone, obrigada por serem irmãs que a vida me deu. Agradeço a vocês, à minha sogra, Stela, e a meu sogro, Cesar, pelo carinho, pelo amor, presença e por sempre torcerem por mim.

Agradeço aos pesquisadores da FCM, prof. Marcondes França, profa. Simone Appenzeller e Thiago Rezende pela colaboração essencial nos trabalhos.

Agradeço a meus colegas de laboratório por serem tão solícitos e por terem compartilhado esses anos comigo. Em especial, agradeço à Gustavo e Diedre, que foram os melhores companheiros de doutorado que eu poderia ter.

Agradeço a colegas e professores que me ajudaram de diversas formas nestes anos, seja dando ideias de projetos, dando orientações sobre processos seletivos, ou me colocando em contato com outros pesquisadores. Em especial, agradeço aos profs. Roberto Lotufo e Richard Frayne e aos amigos Suellen, Irene, Mariana, Roberto e Oeslle.

Finally, I am grateful to Eugenio, my advisor during the time I spent at the Martinos Center, for his reception, attention, and for helping me grow so much both technically and professionally. I also thank all my colleagues at Martinos who, in addition to helping me so much technically, made my days in Boston so wonderful.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

# Resumo

O hipotálamo é uma estrutura cerebral composta por vários subnúcleos que contêm os corpos celulares de múltiplos subtipos de neurônios. Apesar de suas pequenas dimensões, o hipotálamo desempenha um papel significativo no controle do sono, da temperatura corporal, do apetite e das emoções, entre outras funções. Sua disfunção tem sido implicada em vários distúrbios neurológicos, como demência frontotemporal comportamental, esclerose lateral amiotrófica e doença de Alzheimer.

Embora a ressonância magnética (RM) seja amplamente utilizada para estudar o hipotálamo, sua segmentação manual é suscetível a alta variabilidade inter e intra-observador devido ao seu pequeno tamanho e baixo contraste com os tecidos vizinhos. No início deste projeto, não havia nenhum método automatizado para segmentação do hipotálamo disponível na literatura. Mesmo após a proposição de novos métodos, os métodos automatizados existentes eram limitados a sequências e resoluções específicas de RM, restringindo o potencial de pesquisa.

Este projeto visa resolver essas limitações fornecendo dois conjuntos de dados públicos: MiLI (MICLab-LNI Initiative), que foca na estrutura completa em imagens T1 ponderadas com 1mm isotrópico, e HELM (Mapas de Rótulos *ex vivo* do Hipotálamo - do inglês "*Hypothalamus ex vivo label maps*"), que contém a segmentação manual do hipotálamo e suas sub-regiões. Além disso, desenvolvemos dois métodos automatizados de segmentação: HypAST, focado em imagens T1 ponderadas, e H-SynEx, um método capaz de acomodar diversas sequências e resoluções de RM, facilitando assim uma pesquisa mais ampla sobre o hipotálamo. Por fim, aplicamos o H-SynEx em pacientes com três diferentes tipos de doenças (ataxias, lúpus e esclerose lateral amiotrófica) e encontramos diferenças significativas na maioria dos grupos quando comparados com imagens de controle.

**Palavras-chaves:** Hipotálamo, Método de Segmentação, Dataset, Doenças neurodegenerativas, Ressonância Magnética (RM), Aprendizado Profundo



# Abstract

The hypothalamus is a brain structure composed of various subnuclei containing the cell bodies of multiple neuron subtypes. Despite its small dimensions, the hypothalamus plays a significant role in controlling sleep, body temperature, appetite, and emotions, among other functions. Its dysfunction has been implicated in various neurological disorders such as behavioral frontotemporal dementia, amyotrophic lateral sclerosis, and Alzheimer’s disease.

Although magnetic resonance (MR) imaging is widely used to study the hypothalamus, its manual segmentation is prone to high inter- and intra-rater variability due to its small size and low contrast with neighboring tissues. At the start of this project, no automated method for hypothalamus segmentation was available in the literature. Even after new methods were proposed, existing automated methods were limited to specific MRI sequences and resolutions, restricting research potential.

This project aims to address these limitations by providing two public datasets: MiLI (MICLab-LNI Initiative), which focuses on the whole structure in T1-weighted 1mm isotropic images, and HELM (Hypothalamus Ex Vivo Label Maps), which contains manual segmentation of the hypothalamus and its subregions. Additionally, we developed two automated segmentation methods: HypAST, focused on T1-weighted images, and H-SynEx, a method capable of accommodating diverse MRI sequences and resolutions, thereby facilitating broader research on the hypothalamus. Finally, we applied H-SynEx to patients with three different diseases (ataxias, lupus, and amyotrophic lateral sclerosis) and found significant differences in most of the groups when compared to control images.

**Keywords:** Hypothalamus, Segmentation Method, Dataset, Neurodegenerative Diseases, Magnetic Resonance Imaging (MRI), Deep Learning

# List of Figures

Figure 1.1 – Hypothalamus on the adult brain on a T1w MR image. Highlighted on the left, we have the region where the hypothalamus is located. On the right, we can see more clearly the structure through the manual segmentation . . . . .	25
Figure 3.1 – On the left, an example of manual segmentation containing only 5 pixels is shown. In the center and right, examples of disrupted and pixelated manual segmentation are depicted. Highlighted in red is the region where the hypothalamus is located. . . . .	35
Figure 3.2 – Example of partial volume effect: On the left, a brain slice is depicted. Highlighted on the right is a border of the hypothalamus exhibiting varying shades of gray, posing difficulty in distinguishing between hypothalamic voxels and adjacent tissues. .	35
Figure 3.3 – Fluxogram illustrating schematic for experiments based on challenges and raised hypotheses . . . . .	37
Figure 7.1 – Different annotations used on model training and testing: axial 2D view (first row) and volumetric rendering (second row) . . .	45
Figure 7.2 – Samples from datasets used on model training/testing depicting the high variability between them: axial slices highlighting the hypothalamus location (first row); Zoom of the highlighted region (second row). From left to right: OASIS, IXI, MiLI and CC359 samples . . . . .	47
Figure 8.1 – EfficientNet-B4 architecture, based on an image from Agarwal (AGARWAL, ). Despite being originally a classification architecture, it was altered for segmentation purposes in this project. In red, blocks are used for EfficientSeg. In blue, the blocks used for Modified Eff-UNet. . . . .	49
Figure 8.2 – EfficientSeg architecture . . . . .	49

Figure 8.3 – Modified Eff-UNet architecture: it uses an EfficientNet-B4 as encoder of a U-Net . . . . .	50
Figure 8.4 – Correct2Seg scheme proposed composed of two modules, Segmentation and Correction, trained together. The final prediction is the output of the segmentation module. . . . .	51
Figure 8.5 – Quantitative results of our best Correct2Seg model (Model IV), reported by dataset and comparative with Billot method. . . . .	58
Figure 8.6 – Automated <i>versus</i> Manual segmentation. Qualitative results from our best-trained model (Model IV) . . . . .	59
Figure 8.7 – 3D and 2D (axial) views from automated and manual specialist annotations. The automated segmentation can properly divide the right and left hypothalamus, while manual segmentation may join both sides. This may happen once the images do not have enough contrast for the human eye while the CNN model can better recognize the nuances. . . . .	60
Figure 8.8 – Real VS predicted volumes on OASIS dataset. Volumes are measured in $\text{mm}^3$ . Left: Correct2Seg (Model IV). Right: Billot <i>et al.</i> (2020) . . . . .	60
Figure 9.1 – Top Row: HypAST, the Hypothalamus automated segmentation tool. Bottom row: The tool offers volumetric analysis and returns histogram texture attributes of the hypothalamus . . . . .	65
Figure 12.1–Examples of original <i>ex vivo</i> images (up) and images after the pre-processing steps (down). First, we re-orient the images to positive RAS standard and remove non-cerebral elements from the background. Then, we resample the images voxels to $0.3 \times 0.3 \times 0.3\text{mm}$ isotropic and perform bias field correction. . . . .	72
Figure 12.2–Recess of the hypothalamus used for the delineation of superior boundary. Tub_sup = Tuberal Superior subregion; Tub_inf = Tuberal inferior subregion . . . . .	74

Figure 12.3–Segmentation protocol: (a) Subregions delineation: Anterior superior (pink), anterior inferior (red), tuberal superior (yellow), tuberal inferior (blue) and posterior (green) (b) Sagittal landmarks for subregion delineation (AC = Anterior Commissure and MB=mamillary body) (c) Coronal View of superior and inferior tuberal subregions (d) Coronal landmark for superior/inferior division . . . . .	76
Figure 12.4–Examples of three different label maps derived from the same image. From left to right: $k = 4$ , $k = 6$ , $k = 8$ . . . . .	77
Figure 12.5–(a) Whole brain segmentation: example with $k=4$ (b) Manual segmentation simply overlapping the whole brain segmentation. We can see that there are a few inconsistent voxels, that should be labeled either as hypothalamus or background that have different labels. (c) To fix these inconsistencies, a mathematical morphology-based algorithm is applied. . . . .	77
Figure 12.6–Label map creation: Following the segmentation step, a half-brain label map is generated (a). However, given the hypothalamus’s central location within the brain, mirroring is essential to provide contextual information. For the mirroring process, translation and rotation are applied to the RAS coordinates. This involves moving the brain in close proximity to the $x=0$ axis from the negative side, without surpassing into the positive side. Essentially, a final cost function is computed, penalizing positive values of $x$ (b) and high negative values. Finally, we prevent the overlap between brain hemispheres (d) and also prevent them from ending up at unnaturally distant positions (e). . . . .	80
Figure 13.1–Example of different modalities (FSM dataset) . . . . .	83
Figure 13.2–Generation of synthetic images: The synthetic images $S$ are generated using the label maps from the <i>ex vivo</i> images. . . . .	85

Figure 13.3–Examples of coronal slices from 3D synthetic images used as input. The images shown here came from the label maps cropped around the hypothalamus. The use of aggressive data augmentation along random contrast values on the generative model results in large variability in the appearance of the input images.	85
Figure 13.4–Training Flowchart: There are two training blocks, one focused on the entire hypothalamus and another specialized in subregion segmentation. The training of the two blocks is done subsequently. We first trained the whole structure segmentation model( $M_{hyp}$ ), and later, the model for the subregions segmentation( $M_{sub}$ ). However, the output of $M_{hyp}$ is used to assist the input creation of $M_{sub}$ during training.	86
Figure 13.5–Inference flowchart. The inference image $I_{inf}$ goes through a pre-processing step to find the input array $A_{inf}$ . $A_{inf}$ is then applied to the whole structure segmentation model( $M_{hyp}$ ). Finally, using VDC (see text), $A_{inf}$ and the output of $M_{hyp}$ ( $O_{hyp\_inf}$ ), we create the input for the subregion segmentation model ( $M_{sub}$ ) and find the final subregions segmentation.	88
Figure 13.6–Top row: IXI dataset, which only presents the segmentation of the whole structure (excluding the mammillary bodies). Middle row: FSM dataset, that contains the segmentation of the hypothalamus and its subregions. Bottom row: H-SynEx and manual segmentation (target) volumes for FSM dataset	93
Figure 13.7–Qualitative results in different datasets, sequences, and resolutions for H-SynEx. Other methods, when applied to sequences different from T1w, return no results	94
Figure 13.8–Normalized volume correlation for FLAIRs <i>vs</i> T1w (ADNI Dataset) using H-SynEx segmentation. Up: Control subjects; Down: AD patients. We can see that besides the posterior subregion, we can find a positive correlation between FLAIR and T1w normalized volumes.	96

Figure 16.1–Correlation between SARA scores (top) and disease duration (bottom) and hypothalamic imaging (subregion volumes and texture attributes). Interpretation of Pearson’s r coefficient applied to medicine - analysis for absolute values (Akoglu, 2018): 0:None; 0.1-0.29:poor; 0.3-0.59:Fair; 0.6-0.79: Moderate; 0.8-0.99: Very Strong; 1: Perfect. Gray blocks mean no significant correlation. . . . .	109
Figure 18.1–Subregions where we could find volume and/or texture differences between patients and controls. Left: Juvenile Lupus; Right: Adult Lupus. Pink: anterior superior, red: tuberal superior, yellow: tuberal inferior, and green: posterior . . . . .	114

# List of Tables

Table 1.1 – Divergence of manual segmentation protocols in the literature . . .	25
Table 1.2 – Divergence of the hypothalamus volume in the literature . . . . .	26
Table 2.1 – Hypothalamus fully automated segmentation methods . . . . .	31
Table 7.1 – Datasets and respective types of label used on the model devel- opment and testing . . . . .	44
Table 8.1 – Summary of all conducted experiments on Correct2Seg. While developing the model, five models were trained using different data/annotations. For assessing model reliability, three experi- ments were conducted. . . . .	54
Table 8.2 – Results for Model Development and Generalization Experiments. DC = Dice Coefficient; VS = Volume Similarity; AVD = Average Hausdorff Distance. Average values and standard deviations were measured among the metrics of three different runs. Results of Billot <i>et al.</i> are from their inference on OASIS dataset. . . . .	57
Table 8.3 – Inter-Rater Metrics computed for 5 images from test set . . . . .	58
Table 8.4 – Test-Retest metrics on Kirby21 dataset. The values correspond to the mean and standard deviation among the 21 pairs of acqui- sition. For Correct2Seg method, we used one of the three models trained on Model V experiments. . . . .	60
Table 12.1–Landmarks used for hypothalamus division into subregions . . . .	75
Table 12.2–Intra-rater metrics (median) for 5 subjects. . . . .	79
Table 13.1–Datasets used for model validation and testing; WS: Whole Struc- ture, SR: Subregion . . . . .	83
Table 13.2–Summary of conducted experiments . . . . .	89
Table 13.3–Inter-rater metrics (median) for 10 subjects from FSM . . . . .	91

Table 13.4–AVD and DC (median) for H-SynEx and Billot <i>et al.</i> on different subregions for FSM dataset. † indicates statistical significance on a two-sided Wilcoxon rank-sum test using Bonferroni correction for $p < 0.05$ . . . . .	92
Table 13.5–AVD and DC (median) for H-SynEx, ScLimbic (GREVE <i>et al.</i> , 2021) and Billot <i>et al.</i> (BILLOT <i>et al.</i> , 2020) on different datasets (MiLI, IXI, OASIS, and FSM) for the entire hypothalamus (except MB). The symbols indicate statistical significance on a two-sided Wilcoxon rank-sum test using Bonferroni correction for $p < 0.05$ : (*) Billot <i>vs</i> H-SynEx; (†) ScLimbic <i>vs</i> H-SynEx; (‡) Billot <i>vs</i> ScLimbic. Since ScLimbic was trained using the FSM dataset, we did not consider these results. Similarly, since HypAST was trained using data from MiLI, IXI and the same segmentation protocol as OASIS, we did not consider these results . . . . .	95
Table 13.6–AUROC Values for patients vs. controls (hypothalamus volume) for H-SynEx and Billot methods in ADNI and NIFD datasets. For ADNI dataset, we also analyze our method when applied to FLAIR images with spacing of 5mm. Stars indicate the level of statistical significance (two-sided Wilcoxon rank-sum test) between both cohorts (* $p < 0.05$ , ** $p < 0.01$ ). † indicates statistical significance on the DeLong test ( $p < 0.05$ ) between H-SynEx and Billot methods. ‡ indicates statistical significance on the DeLong test ( $p < 0.05$ ) between H-SynEx applied on T1w and H-SynEx applied on FLAIRS. . . . .	97
Table 16.1–SCA patients and controls demopgraphics . SARA: scale for assessment and rating of ataxia . . . . .	107
Table 16.2–Subregions that presented significant difference when comparing patient and control groups (Mann-Whitney test with confidence level of 95%) . . . . .	108
Table 18.1–Adult and Juvenile lupus patients and control demopgraphics . . .	113



# List of Acronyms and Notations

AD	Alzheimer’s Disease
ADNI	Alzheimer’s Disease Neuroimaging Initiative
ALS	Amyotrophic Lateral Sclerosis
AUROC	Receiving Operating Characteristic Curve
AVD	Average Hausdorff Distance
bvFTD	Behavioral Variant Frontotemporal Dementia
CNN	Convolutional Neural Networks
DC	Dice Coefficient
DL	Dice Loss
DTI	Diffusion Tensor Imaging
FA	Fractional Anisotropy
FLAIR	Fluid-Attenuated Inversion Recovery
FSM	FreeSurfer Maintenance Dataset
HD	Hausdorff Distance
HELM	Hypothalamus <i>ex vivo</i> Label Maps
HypAST	Hypothalamus Automated Segmentation Tool
ICC	Intraclass Correlation
LNI	Neuroimaging Laboratory
MB	Mammillary Body
MICLab	Medical Image and Computing Laboratory
MiLI	MICLab-LNI Initiative initiative
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
NIFD	Neuroimaging in Frontotemporal Dementia dataset
PD	Proton Density
qT1	Quantitative T1
SARA	Scale for Assessment and Rating of Ataxia
SCA	Spinocerebellar Ataxias
SLE	Systemic Lupus Erythematosus
STAPLE	Simultaneous Truth and Performance Level Estimation
TIV	Total Intracranial Volume
VDC	Ventral Diencephalon
VS	Volume Similarity

# Contents

<b>I</b>	<b>Introduction</b>	<b>23</b>
<b>1</b>	<b>An introduction to the Hypothalamus</b>	<b>24</b>
<b>2</b>	<b>Literature Review</b>	<b>27</b>
2.1	Deep Learning for Medical Imaging Segmentation	27
2.2	Deep Learning Architectures	28
2.3	Hypothalamus Segmentation	29
<b>3</b>	<b>Motivations, Objectives, Challenges, and Hypotheses</b>	<b>32</b>
3.1	Motivations	32
3.2	Objectives	33
3.3	Challenges	34
3.4	Hypotheses	36
<b>4</b>	<b>Main Contributions</b>	<b>38</b>
4.1	Datasets	38
4.2	Methods	38
4.3	Applications	39
<b>5</b>	<b>Thesis Structure</b>	<b>40</b>
<b>II</b>	<b>Hypothalamus Segmentation on T1w Images</b>	<b>41</b>
<b>6</b>	<b>Introduction</b>	<b>42</b>
<b>7</b>	<b>MiLI: the MICLab-LNI Initiative</b>	<b>43</b>
7.1	Datasets	43
7.2	Labels	44
7.3	Ethics Committee Approval	46
7.4	Manual Segmentation Protocol	47
7.5	Dataset Availability	47
<b>8</b>	<b>Hypothalamus Automated Segmentation Method</b>	<b>48</b>
8.1	Method	48

8.1.1	EfficientSeg . . . . .	48
8.1.2	Modified Eff-UNet . . . . .	49
8.1.3	Correct2Seg . . . . .	50
8.2	Evaluation Metrics . . . . .	51
8.3	Experiments . . . . .	53
8.3.1	Choosing Correct2Seg modules . . . . .	53
8.3.2	Correct2Seg Model development . . . . .	54
8.3.2.1	Pre-training . . . . .	55
8.3.2.2	Training . . . . .	55
8.3.2.3	One-Step training . . . . .	56
8.3.3	Model Reliability . . . . .	56
8.3.3.1	Inter-Rater Metrics . . . . .	56
8.3.3.2	Test-retest . . . . .	56
8.3.3.3	Generalization Test . . . . .	56
8.4	Results . . . . .	57
8.4.1	Model Development . . . . .	57
8.4.2	Model Reliability . . . . .	57
8.4.2.1	Inter-Rater Metrics . . . . .	57
8.4.2.2	Test-retest . . . . .	58
8.4.2.3	Generalization Test . . . . .	59
8.5	Discussion . . . . .	61
<b>9</b>	<b>Graphic User Interface Tool . . . . .</b>	<b>64</b>
<b>10</b>	<b>Considerations . . . . .</b>	<b>66</b>

<b>III Hypothalamus Segmentation across different sequences and resolution</b>	<b>67</b>
<b>11 Introduction . . . . .</b>	<b>68</b>
<b>12 HELM - Hypothalamus <i>ex vivo</i> Label Maps . . . . .</b>	<b>70</b>
12.1 Materials . . . . .	70
12.2 Methods . . . . .	71
12.2.1 Pre-processing of <i>ex vivo</i> scans . . . . .	71

12.2.2	Segmentation . . . . .	73
12.2.3	Hemisphere mirroring . . . . .	76
12.3	Manual segmentation quality assessment . . . . .	78
12.4	Data Usage and Availability . . . . .	79
<b>13</b>	<b>H-SynEx . . . . .</b>	<b>81</b>
13.1	Materials . . . . .	81
13.1.1	Training Data . . . . .	81
13.1.2	Test Data . . . . .	81
13.2	Methods . . . . .	83
13.2.1	Data Preprocessing . . . . .	83
13.2.2	Training . . . . .	84
13.2.3	Inference and Post processing . . . . .	87
13.2.4	Statistical Analysis . . . . .	87
13.3	Experiments . . . . .	88
13.3.1	Inter-rater metrics . . . . .	89
13.3.2	Direct comparison with manual segmentation on different sequences . . . . .	90
13.3.3	Comparing against other state-of-the-art methods . . . . .	90
13.3.4	Application to group studies . . . . .	90
13.3.5	Resilience to large slice spacing . . . . .	91
13.4	Results . . . . .	91
13.4.1	Inter-rater metrics . . . . .	91
13.4.2	Direct comparison with manual segmentation on different sequences . . . . .	92
13.4.3	Comparing against other state-of-the-art methods . . . . .	92
13.4.4	Application to group studies . . . . .	95
13.4.5	Resilience to large slice spacing . . . . .	95
13.5	Discussion . . . . .	96
<b>14</b>	<b>Considerations . . . . .</b>	<b>101</b>

<b>IV Applications</b>	<b>102</b>
<b>15 Applying H-SynEx in Group Studies</b>	<b>103</b>
15.1 Materials	103
15.2 Methods	103
15.2.1 Image Segmentation	104
15.2.2 Image Analysis	104
15.2.3 Statistical Analysis	105
<b>16 Spinocerebellar Ataxias</b>	<b>106</b>
16.1 Description	106
16.2 Materials	106
16.3 Results and Discussion	107
<b>17 Amyotrophic Lateral Sclerosis</b>	<b>110</b>
17.1 Description	110
17.2 Materials	110
17.3 Results and Discussion	111
<b>18 Systemic Lupus Erythematosus</b>	<b>112</b>
18.1 Description	112
18.2 Materials	112
18.3 Results and Discussion	113
<b>V Conclusion</b>	<b>115</b>
<b>19 Conclusions</b>	<b>116</b>
<b>20 Next Studies</b>	<b>118</b>
<b>21 Publications, Collaborations, Awards</b>	<b>119</b>
21.1 Publications on the Hypothalamus	119
21.2 Publications on collaborative projects	119
21.3 Colaborations	120
21.4 Awards	121
<b>Bibliography</b>	<b>122</b>

<b>ANNEX A Ethical Approvals . . . . .</b>	<b>134</b>
--	------------

## Part I

### Introduction

# 1 An introduction to the Hypothalamus

The hypothalamus is a gray matter brain structure composed of several subnuclei, being part of the limbic system (WOLFF *et al.*, 2018). It works analogously to a signal transmitter, connecting the central nervous system to its periphery, and has important functions in controlling sleep, and regulating body temperature, appetite, and emotions (VERCRUYSSSE *et al.*, 2018). Although it is believed that its average volume in adults was approximately  $4\text{cm}^3$ , more recent studies have already demonstrated that it does not reach half this value (SCHÖNKNECHT *et al.*, 2013). In the literature, several studies establish a connection between the whole hypothalamus and neurodegenerative diseases such as Alzheimer’s disease (PIYUSH; RAMAKRISHNAN, 2014), Huntington’s disease (GABERY *et al.*, 2015; BARTLETT *et al.*, 2019), Behavioral Variant Frontotemporal Dementia (bvFTD) (BOCCHETTA *et al.*, 2015; PIGUET *et al.*, 2011), Amyotrophic Lateral Sclerosis (ALS) (GORGES *et al.*, 2017; AHMED *et al.*, 2021a), among others (SEONG *et al.*, 2019; MODI *et al.*, 2019; WOLFE *et al.*, 2015; GUTIERREZ *et al.*, 1998), with some studies suggesting a differential involvement of the hypothalamic subregions across conditions (BOCCHETTA *et al.*, 2015). To better understand the role of the hypothalamus, several studies use different magnetic resonance images (MRI) sequences (GORGES *et al.*, 2017; SEONG *et al.*, 2019; WOLFE *et al.*, 2015; PIYUSH; RAMAKRISHNAN, 2014; SCHUR *et al.*, 2015). However, these studies are limited to select sites and require specialists with neuroanatomical knowledge to perform manual annotation.

Despite many studies in the literature use hypothalamus segmentation to analyze the volume variation, in many cases, the segmentation is still done manually and susceptible to human mistakes and different protocol approaches, which may include or exclude other structures (Table 1.1).

Besides, for being a small region and hard to visualize in MRI (Figure 1.1), it is difficult to determine the hypothalamus’ morphological landmarks



Table 1.1 – Divergence of manual segmentation protocols in the literature

	<b>Mammillary Bodies</b>	<b>Fornix</b>
(SCHINDLER <i>et al.</i> , 2013)	included	excluded
(TOGNIN <i>et al.</i> , 2012)	excluded	included
(TERLEVIC <i>et al.</i> , 2013)	excluded	included
(MAKRIS <i>et al.</i> , 2013)	included	included

(SCHÖNKNECHT *et al.*, 2013; GABERY *et al.*, 2015). For example, two different studies on schizophrenia have been published reporting contradictory results: while Goldstein *et al.* (GOLDSTEIN; SEIDMAN, 2007) found increased volume in the hypothalamus of patients, Klomp *et al.* (KLOMP *et al.*, 2012), on the other hand, reported preserved volumes. In the case of Huntington’s disease, Gabery *et al.* (GABERY *et al.*, 2015) found no difference between the volume of the hypothalamus of patients and control subjects, whereas Bartlett *et al.* (BARTLETT *et al.*, 2019) found a reduction in the gray matter of the hypothalamus in patients.

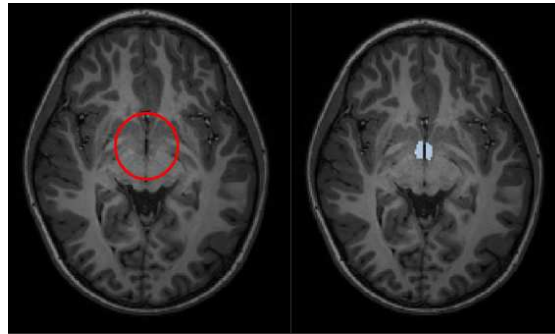


Figure 1.1 – Hypothalamus on the adult brain on a T1w MR image. Highlighted on the left, we have the region where the hypothalamus is located. On the right, we can see more clearly the structure through the manual segmentation

Analyzing some studies applied on different databases (Table 1.2), we can see that the hypothalamic volume for the control group varies about 57% from the smaller (910mm<sup>3</sup>, including mammillary bodies, excluding fornix) to the greater(1455mm<sup>3</sup>, counting both sides, including both mammillary bodies and fornix).

Table 1.2 – Divergence of the hypothalamus volume in the literature

Author	Database	Volume(mean/std)
<b>Tognin <i>et al.</i> (TOGNIN <i>et al.</i>, 2012)</b>	26 healthy controls 26 with schizophrenia	$910 \pm 10 \text{ mm}^3$ for controls
<b>Schindler <i>et al.</i> (SCHINDLER <i>et al.</i>, 2013)</b>	10 healthy subjects	$1130.64 \pm 103.48 \text{ mm}^3$
<b>Boccheta <i>et al.</i> (BOCCHETTA <i>et al.</i>, 2015)</b>	18 patients with bvFTD 18 controls	$944 \pm 73 \text{ mm}^3$ for control group.
<b>Wolff <i>et al.</i> (WOLFF <i>et al.</i>, 2018)</b>	23 healthy controls 20 with bipolar depression 41 with major depression	$703 \pm 53$ to $732 \pm 63 \text{ mm}^3$ for the right hypothalamus $719 \pm 75$ to $752 \pm 54 \text{ mm}^3$ for the left hypothalamus for control group

Also, hypothalamus manual segmentation is time-consuming. According to Schindler *et al.* (SCHINDLER *et al.*, 2013), each volume takes up to 3 hours to be segmented, which may become a bottleneck for researchers using large datasets. To reduce this time, many authors use semi-automated methods (WOLFF *et al.*, 2018; MAKRIS *et al.*, 2013; SCHINDLER *et al.*, 2013). Yet Wolff *et al.* (WOLFF *et al.*, 2018), for instance, developed a semi-automated method for which each evaluator needed to be trained for three to four weeks, five to six hours per day. Each hypothalamus segmentation took around 20-40 minutes to be generated.

## 2 Literature Review

In this chapter, we will discuss some concepts of automated segmentation found in the literature, as well as the deep learning architectures used throughout this work. We will also cover the literature on automated segmentation with a focus on the hypothalamus.

### 2.1 Deep Learning for Medical Imaging Segmentation

To reduce the time and subjectivity of medical segmentations in general, consequently improving reliability, many researchers are using automatic segmentation methods, based on deep convolutional neural networks (CNNs) (RONNEBERGER *et al.*, 2015; WACHINGER *et al.*, 2018; CHEN *et al.*, 2018). In literature, we can find methods of automatic brain segmentation developed either to target specific structures, such as the hippocampus (ATALOGLOU *et al.*, 2019) and corpus callosum (PARK *et al.*, 2018), or studies aiming at several structures at the same time, such as the work of Wachinger *et al.* (WACHINGER *et al.*, 2018).

While large-scale brain imaging studies using MRI have great potential to enhance our understanding of the human brain in both health and disease conditions, these studies are often limited by the need for manual annotation. Automated segmentation methods have been developed to circumvent this problem and allow for a greater quantity of data to be utilized. Existing methods include classical atlas-based approaches (IGLESIAS; SABUNCU, 2015; CABEZAS *et al.*, 2011) and more recent deep learning networks (WANG *et al.*, 2022; ROMÁN *et al.*, 2020).

Also, the use of semi-supervised models on medical images enhances the generalization of networks without necessarily increasing the quantity of annotated data (FAYJIE *et al.*, 2022; BORTSOVA *et al.*, 2019). However, most of these

models work only in one type of MRI sequence and usually need retraining to adapt to different sequences. Currently, the use of synthetic images has also been a constant subject of study in the field since it allows the construction of training datasets and flawless ground truths (THAMBAWITA *et al.*, 2022; BILLOT *et al.*, 2023) and the development of methods capable of generalizing in across different MRI sequences (IGLESIAS, 2023; BILLOT *et al.*, 2023). Besides the synthetic approach, the usage of ultra-high resolution *ex vivo* MRI has proven to be beneficial in the segmentation of small structures such as the hippocampus, amygdala, and thalamus (IGLESIAS *et al.*, 2015; SAYGIN *et al.*, 2017; IGLESIAS *et al.*, 2018), as it permits a better visualization of their anatomical boundaries, leading to more accurate manual annotation.

## 2.2 Deep Learning Architectures

During the development of this project, we mostly explored two CNN architectures: U-Net (RONNEBERGER *et al.*, 2015) and EfficientNet (MINGXING; QUOC, 2019).

The U-Net (RONNEBERGER *et al.*, 2015) is a segmentation architecture developed for medical imaging segmentation. For being simple and effective, it is widely explored in the literature (DO *et al.*, 2019; JUAREZ A. GARCIA, 2019; CARMO; SILVA, 2021). The U-Net has two paths: a contracting path, to capture context, and an extension path, for precise location. The main idea of U-Net is to replace pooling operators with upsampling operators on extension paths, which will increase output resolution. For location purposes, upsampled output is combined with high-resolution features from the contracting path.

The EfficientNet (MINGXING; QUOC, 2019), on the other hand, is the state of the art for natural image classification. The authors proposed a compound scaling method in order to scale the network depth, width, and resolution in a fixed ratio. Their baseline, EfficientNet-B0 has 5.3M parameters and is scaled up until EfficientNet-B7, with 66M parameters. Despite the primary goal of the network differs from ours, we can see in the literature that the use of EfficientNet as an

encoder has been explored for medical imaging segmentation (MATHEWS *et al.*, 2020; NGUYEN *et al.*, 2020). In this project, we use the EfficientNet-B4.

## 2.3 Hypothalamus Segmentation

At the time this project started, one could not find any automated segmentation method for the hypothalamus on the literature. There were, though, a few semi-automated methods and manual segmentation protocols published. For instance, in 2018, (WOLFF *et al.*, 2018) proposed a semi-automatic segmentation method by overlapping MRI images and obtained intraclass correlation (ICC) of 0.78 and 0.82 between three raters. However, each rater had to be trained for three to four weeks, five to six hours a day, and each volumetry took about 20-40min to be generated.

The first automated method for hypothalamus segmentation was developed by the candidate (RODRIGUES *et al.*, 2020). The proposed method used three modified 2D U-Nets, creating a consensus from the axial, sagittal, and coronal views. It reported a Dice coefficient of 0.77. The dataset was composed of 205 images, divided into 70%/15%/15% images for training, validation, and testing, resulting in approximately 30 images for the test set.

At the beginning of this work, there were no fully automated segmentation methods of the hypothalamus. Even today, besides the aforementioned method (RODRIGUES *et al.*, 2020), we can only find five options for hypothalamus automated segmentation in the literature, being two of them developed by the candidate and better explained throughout this thesis.

- (RODRIGUES *et al.*, 2022): This study, developed by the candidate, uses a teacher-student-based architecture containing two blocks: segmentation and correction, where the second corrects the imperfections of the first block. The method used over 1300 images for training/validation. It is focused on 1mm isotropic T1w images and it is openly available.

- (BILLOT *et al.*, 2020): The authors used a 3D U-Net-based architecture to segment the hypothalamus and its subunits. The authors used aggressive data augmentation and 37 volumes for training. Finally, their test set had a Dice coefficient of 0.84. However, the authors predicted 675 images from the public ADNI dataset to better analyze their method. A specialist qualitatively evaluated the segmentations and assessed whether they could be part of a neuroimaging study. Only six images were rejected.
- (GREVE *et al.*, 2021): The authors trained a U-Net model to automatically segment multiple subcortical limbic structures, including the hypothalamus. The training utilized 39 manually labeled T1w MRI datasets, employing spatial, intensity, contrast, and noise augmentation techniques. The method is focused on 1mm isotropic T1w images.
- (ESTRADA *et al.*, 2023): Different from the previous studies, HypVINN is capable to work on T1w and T2w images with resolutions from 0.8 to 1mm isotropic. The authors integrated VINN (HENSCHER *et al.*, 2022)) into their network design, based on FastSurferCNN (HENSCHER *et al.*, 2020). Both methods utilize a 2.5D approach with three 2D F-CNNs per anatomical view. They used images from Rhineland Study (RS)(BRETELER *et al.*, 2014) and the UK Biobank (UKB) (ALFARO-ALMAGRO *et al.*, 2018).
- (RODRIGUES *et al.*, 2024): H-SynEx (better explained on Part III) is an automated segmentation approach for hypothalamic subregions, capable of generalizing across various MRI sequences and resolutions without necessitating retraining. It was trained using synthetic images generated from label maps constructed from ultra-high-resolution *ex vivo* MRI scans, enabling more detailed manual segmentation compared to standard 1 mm isotropic *in vivo* images.

Table 2.1 – Hypothalamus fully automated segmentation methods

<b>Author</b>	<b>Segmentation Target</b>	<b>MRI Sequence</b>	<b>Voxel resolution</b>
(RODRIGUES <i>et al.</i> , 2020)	whole structure	T1w	1mm isotropic
(BILLOT <i>et al.</i> , 2020)	whole structure and subregions	T1w	1mm isotropic
(RODRIGUES <i>et al.</i> , 2020)	whole structure	T1w	1mm isotropic
(GREVE <i>et al.</i> , 2021)	whole structure	T1w	1mm isotropic
(ESTRADA <i>et al.</i> , 2023)	whole structure and subregions	T1w and T2w	0.8mm isotropic 1mm isotropic
(RODRIGUES <i>et al.</i> , 2024)	whole structure and subregions	Multiple sequences	Multiple resolution

## 3 Motivations, Objectives, Challenges, and Hypotheses

In this section, we will explore the primary motivations behind the work and outline its main and secondary objectives. Furthermore, we will identify the major challenges faced during the execution of the project and present the hypotheses developed to overcome them.

### 3.1 Motivations

When analyzing patients with bvFTD and ALS, the literature points to atrophy of the hypothalamus (AHMED *et al.*, 2021a; PIGUET *et al.*, 2011) and for ALS, some studies have found hypothalamic atrophy even in pre-symptomatic gene carriers compared with a control group (AHMED *et al.*, 2021a). In Huntington’s disease, some studies have indicated the volumetric alteration of the hypothalamus in patients (BARTLETT *et al.*, 2019) while others showed no volume difference between patients and the control group (GABERY *et al.*, 2015). However, even the latter indicate that variations can occur in specific nuclei of the hypothalamus, being lost when analyzing the structure as a whole (GABERY *et al.*, 2015).

Currently, the diagnosis time for rare neurodegenerative diseases is long, delaying treatment in the early phase of the symptoms. For example, in ALS, HD, and bvFTD, the diagnosis is based on clinical signs, and in the exclusion of other motor dysfunctions (HARDIMAN *et al.*, 2011; JOHNEN; BERTOUX, 2019). For bvFTD, patients are often mistaken for having Alzheimer’s disease (AD) once there are no biological markers for diagnosis (JOHNEN; BERTOUX, 2019; BEBER; CHAVES, 2013). Patients with ALS usually take 9–15 months to be correctly diagnosed after the onset of symptoms (HARDIMAN *et al.*, 2011). Early diagnosis of these diseases could lead to more research on possible treatments



and earlier treatment, improving patient quality of life (CHO; SHUKLA, 2020). For instance, moderate exercise may prolong the early stages of ALS (OLIVEIRA; PEREIRA, 2009).

In this scenario, it is evident that proper morphological evaluation of the hypothalamus *in vivo* images might be relevant to understanding normal human physiology and multiple disease states. MRI of the brain, mainly using high-resolution acquisition, may assist with this task. However, this approach is challenging. As highlighted before, identifying morphological landmarks of the hypothalamus is often difficult, making segmentation and volumetric analyses challenging. The reproducibility of the results and the available means to study the hypothalamus remains a critical issue.

## 3.2 Objectives

To improve the study of the hypothalamus, it is essential to reduce the time and subjectivity of medical manual segmentations, consequently improving reliability. Additionally, there is a need for the development of fully automated methods capable of generalizing across various MRI sequences and resolutions. However, to foster ongoing investigation into this structure, it's equally vital to furnish resources, namely, a publicly accessible dataset for use by other researchers. With this in mind, our primary objective is :

**Help to expand the study of the hypothalamus by developing segmentation methods and distributing datasets specifically tailored to this structure and its subregions**

From the main objective we can derive the following specific objectives:

- Create and publish a dataset focused on hypothalamus segmentation on T1w images;
- Develop a method for feasibility study, focused on generalizing on different T1w datasets;

- Create a second dataset, based on ultra-high-resolution *ex vivo* images, which may be generalized to different structures;
- Taking advantage of the expanded dataset and ultra-high-resolution data, develop a new version of the hypothalamus subnuclei segmentation method, capable of working across MR sequences and resolutions;
- Develop a pipeline for volume and texture analysis of hypothalamus subnuclei on MR images of different conditions and neurodegenerative diseases.

### 3.3 Challenges

During the development of the project, we dealt with six main challenges:

1. Small size of the structure: The hypothalamus is a small structure, with a total volume of less than  $4 \text{ cm}^3$ . In terms of pixels, this means that some slices of the MR volume may contain only 10 pixels or fewer representing the structure (Figure 3.1).
2. Low contrast with neighboring tissues: The hypothalamus lacks clear, distinct borders. Instead, its edges are often indistinguishable from surrounding tissues.
3. Protocol variations: Manual segmentations were conducted by different specialists following distinct protocols. As a result, labels may include or exclude the fornix.
4. Noisy labels: Due to the aforementioned challenges, manual segmentation is prone to errors. Additionally, it lacks smoothness and continuity, appearing pixelated and disrupted, particularly along borders (Figure 3.1).
5. Partial volume effect: Partial volume occurs when information from two or more types of tissues is represented in the same voxel, making it difficult to discern the boundaries of the structure (BILLOT *et al.*, 2023) (Figure 3.2).

6. Lack of publicly available data: At the beginning of this project, there were no openly accessible datasets specifically focused on the hypothalamus.

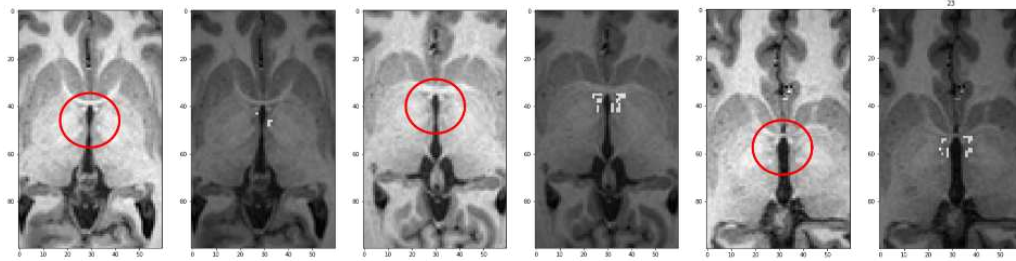


Figure 3.1 – On the left, an example of manual segmentation containing only 5 pixels is shown. In the center and right, examples of disrupted and pixelated manual segmentation are depicted. Highlighted in red is the region where the hypothalamus is located.

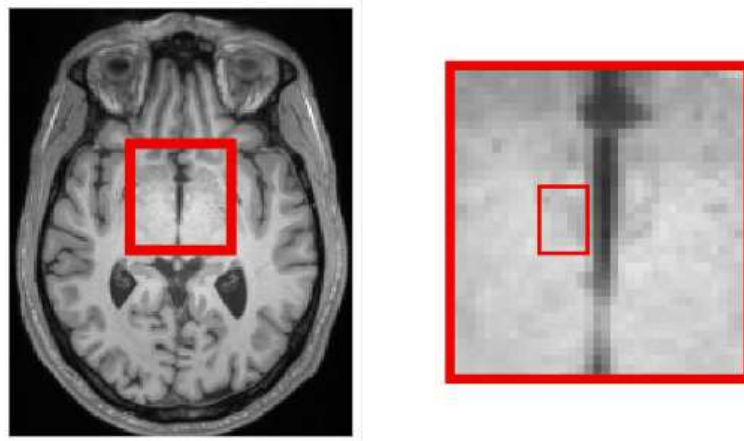


Figure 3.2 – Example of partial volume effect: On the left, a brain slice is depicted. Highlighted on the right is a border of the hypothalamus exhibiting varying shades of gray, posing difficulty in distinguishing between hypothalamic voxels and adjacent tissues.

## 3.4 Hypotheses

The aforementioned challenges gave rise to eight hypotheses regarding the training of a deep learning model to achieve the segmentation task.

- a. When segmenting small structures, it is better to use an architecture with fewer downsampling layers.
- b. Annotation noise often accumulates at the borders of the structure.
- c. Training with a combination of automated and manual segmentation, rather than solely relying on manual segmentation, could decrease noise and alleviate the impact of varying protocols.
- d. To minimize noise, it is essential to utilize more data, even if the labels are not manually annotated.
- e. Designing an architecture capable of learning from its own segmentation can effectively reduce noise.
- f. Ultra-high resolution images can be leveraged to precisely delineate the morphological borders of small structures.
- g. Training with synthetic images allows for training a network with minimal original images, taking advantage of label augmentation.
- h. Models trained with synthetic images can generalize across different sequences and resolutions without the need for retraining.

Finally, we conducted two main steps, to validate our hypotheses:

- I. Hypothalamus Segmentation on T1w Images: In this step, we focused on the deep learning architecture and on developing a dataset focused on T1w images that could be used for training.

- II. Hypothalamus Segmentation across different sequences and resolution: In this second step, we leverage the information acquired in the first step, but now we focus more on data and how can we use it to address our challenges.

The relation between challenges, experiments and the hypotheses are summarized in Figure3.3.

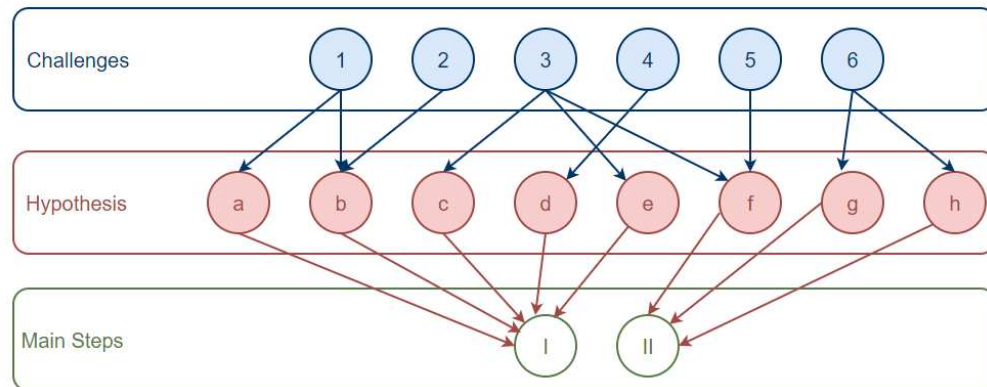


Figure 3.3 – Fluxogram illustrating schematic for experiments based on challenges and raised hypotheses

## 4 Main Contributions

At the beginning of this project, we could not find any automated method or public dataset focused on hypothalamus segmentation. Therefore, the main contribution of this work was to help filling this gap in the literature and provide tools for hypothalamus study in different neurodegenerative diseases and other conditions.

### 4.1 Datasets

Focusing on expanding the study on hypothalamus segmentation, we published two openly available datasets:

- **MiLI:** The MICLab-LNI Initiative is a dataset focused on the segmentation of the whole hypothalamus, with no subdivision, in T1w 1mm isotropic images. It has over 1300 images from 4 different datasets and the labels come from manual and automated segmentations.
- **HELM:** The Hypothalamus *ex vivo* Label Maps contains label maps derived from 10 *ex vivo* images focusing on the hypothalamus and its subdivisions. This dataset is intended to be used on models trained with synthetic images, that are generated using the label maps. The label maps have also the potential to be used in other structures.

### 4.2 Methods

The first automated segmentation method for hypothalamus segmentation was developed and published by the author of this thesis (RODRIGUES *et al.*, 2020). However, for being a feasibility study, it was trained on images from a sin-

gle dataset and was not capable of generalizing in images from different datasets. Given that, two other methods were developed:

- **HypAST:** The Hypothalamus Automated Segmentation Tool is a segmentation method trained using MiLI and capable of generalizing on T1w images from different scanners. When we compare HypAST with other methods from the literature using datasets that were not involved in the training of any method, HypAST delivers superior results across all metrics.
- **H-SynEx:** Finally, aiming to fill the gap in the hypothalamus segmentation in different sequences, we developed H-SynEx, a method trained with synthetic images derived from HELM. H-SynEx is capable of working across different sequences and resolutions, including  $5mm$  spacing FLAIR images.

### 4.3 Applications

Finally, we use H-SynEx to run a study group on three different conditions: Ataxias, Lupus and ALS. The study groups analyzed volume and texture attributes of the hypothalamus subregions and could find significant difference in most of the cases.

## 5 Thesis Structure

This thesis is divided into five parts:

- **Part I:** An **Introduction** to the hypothalamus, literature review, objectives, motivations, challenges, and main contributions;
- **Part II:** Explains the first part of the project, focused on **Hypothalamus Segmentation on T1w Images**. In this part, we introduce the hypothalamus automated segmentation tool (HypAST) and the MICLab-LNI Initiative (MiLI) dataset, both focused on T1w 1mm isotropic images;
- **Part III:** This part focus on the **Hypothalamus Segmentation across different sequences and resolution**. In this part, we introduce the hypothalamus segmentation method based on synthetic images from *ex vivo* images (H-SynEx) and the Hypothalamus *ex vivo* label maps (HELM) dataset, both for images from different MRI sequences and resolutions;
- **Part IV:** In this part, we focus on the method **Applications**, applying H-SynEx in different diseases, such as ataxias, lupus, and ALS. We conduct volumetric and texture analysis comparing control and patient groups.
- **Part V:** In the **Conclusions**, we explore our considerations, papers published during the author's PhD, and future works.



## Part II

# Hypothalamus Segmentation on T1w Images

## 6 Introduction

Aiming to support further development of hypothalamus segmentation models, we present here the first public hypothalamus segmentation dataset along with an automated segmentation method. The dataset consists of diverse T1w MRI datasets comprising 1381 subjects from IXI <sup>1</sup>, CC359 (SOUZA *et al.*, 2018), OASIS (LAMONTAGNE *et al.*, 2019), and MiLI (the latter created specifically for this benchmark). All data are provided with automatically generated hypothalamus masks and a subset containing manually annotated masks. For the method, we present a teacher-student-based model for fully automated segmentation of the hypothalamus on T1w MRI, and a comprehensive framework for reproducing the method and supporting further development of hypothalamus segmentation models.

Given that, this chapter is composed of three main parts:

- **MiLI: the MICLab-LNI Initiative:** Given that at the beginning of this project we could not find any openly available dataset for hypothalamus segmentation, our first step was to create this dataset. For that, we used a combination of openly available MRI data and in-house acquisitions, made at the Unicamp Hospital. This stage aids in addressing Challenge 1 and serves to validate Hypotheses *c* and *d*.
- **Model development:** Using the created dataset, we developed the first version of the segmentation method, focusing only on T1w images and the segmentation of the whole structure. The method was developed based on Hypothesis *a* and *e*.
- **Tool development:** Aiming to facilitate the use of our method, we developed a graphic user interface tool, in which the user could analyze the volumetry and texture attributes of the hypothalamus.

---

<sup>1</sup> <<https://brain-development.org/ixi-dataset/>>

## 7 MiLI: the MICLab-LNI Initiative

At the beginning of this project, no dataset specifically focused on hypothalamus segmentation was available. Therefore, the first step of this project is to create a publicly accessible dataset to support the development of models with strong generalization capabilities on T1w images.

### 7.1 Datasets

For the development of MiLI, we used images from four datasets in this project:

- **MiLI**: The MICLab-LNI Initiative comprises 452 T1w MR image subjects, including 317 controls and 135 patients with inherited ataxias and motor neuron diseases that affect hypothalamus morphology. The average age of the subjects was 36.14 years, 212 male and 240 female. Images were acquired at the Hospital of the University of Campinas. All subjects underwent an MR imaging scan on a 3T Philips Achieva scanner (Philips, Best, The Netherlands) using standard 8-channel head coils. To segment the hypothalamus, we acquired 3D high-resolution T1 volumetric images of the brain in sagittal orientation. Voxel matrix  $240 \times 240 \times 180$ , voxel size  $1 \times 1 \times 1 \text{ mm}^3$ , TR/TE 7/3.201 ms, and flip angle  $8^\circ$ . MiLI was specifically created for this project.
- **CC359** (SOUZA *et al.*, 2018): The Calgary-Campinas dataset is a public dataset of 359 T1w MR healthy brain images from three different vendors (i.e., GE, Philips, and Siemens) and at two magnetic field strengths (i.e., 1.5 T and 3 T).
- **IXI**<sup>1</sup>: Public dataset of 581 T1w MR healthy brain images.

---

<sup>1</sup> <<https://brain-development.org/ixi-dataset/>>

- **OASIS** (LAMONTAGNE *et al.*, 2019): Public dataset of T1w MR images. We used 21 images from OASIS only for testing. Unlike other datasets, this dataset was not used for training.
- **Kirby 21** (LANDMAN *et al.*, 2011): Composed of 42 acquisitions from 21 different subjects. This dataset was used only for test-retest analysis. Therefore, manual segmentation was not required.

## 7.2 Labels

According to Karimi *et al.* (KARIMI *et al.*, 2020), labels can contain three main sources of noise: Inter-rater observability, annotation error, and error in computed-generated labels. In the case of segmentation tasks, especially in challenging structures such as the hypothalamus, “annotation error” means that some pixels of the image can be wrongly segmented by the specialist. The same logic is applied to “computed-generated errors”, however here the mistake is done by an automated tool. The annotation of the data varied according to the dataset (Table 7.1).

Table 7.1 – Datasets and respective types of label used on the model development and testing

Dataset	Total Used	Inexperienced Annotation	Expert Annotation	Consensus	STAPLE
<b>MiLI</b>	452	407	45	×	407
<b>CC359</b>	359	×	18	341	×
<b>IXI<sup>a</sup></b>	560	×	22	538	×
<b>OASIS</b>	10	×	10	×	×
<b>Total</b>	1381	407	95	879	407

<sup>a</sup>Some images from IXI were excluded from the experiments since the automated methods were not able to segment the hypothalamus

Because manual annotation is time-consuming, most of our training data were segmented using only automated methods. To deal with different types of segmentation and reduce label noise, we used label-cleaning strategies such as

STAPLE and majority voting. Souza *et al.* (SOUZA *et al.*, 2018) defined this as the silver standard labels (Figure 7.1) We chose to not visually inspect the automated segmentations. Instead, we used the mentioned techniques to minimize label noise and segmentation errors of the automated methods, because we wanted to analyze whether the network could learn from them (Hypothesis *d*).

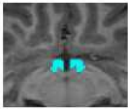
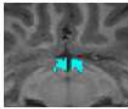
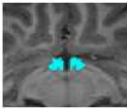
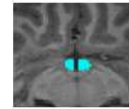
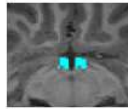
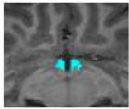

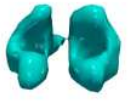



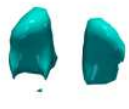
	Expert Segm.	Inexp. Segm.	Billot <i>et al</i>	Rodrigues <i>et al</i>	STAPLE	Consensus
2D View						
3D View						

Figure 7.1 – Different annotations used on model training and testing: axial 2D view (first row) and volumetric rendering (second row)

- **Specialist manual annotation:** Segmentation was performed by a specialist, following the segmentation protocol established for the study, as explained in Section 7.4.
- **Inexperienced rater manual annotation:** Segmentation is performed by an inexperienced rater. This segmentation followed the same protocol described in Section 7.4, but it was performed by a rater with no expertise in MR images and neuroanatomy. However, this increases the amount of data available for the training.
- **Automated annotation:** Segmentation is performed using available automated tools (BILLOT *et al.*, 2020; RODRIGUES *et al.*, 2020), and the

method developed by Billot *et al.* (BILLOT *et al.*, 2020) segments the sub-units of the hypothalamus, including mammillary bodies (MB). Once our manual protocol did not include MBs, we excluded the labels referring to them.

Aiming to reduce label noise and improve the generalization of the network, we combined the previously mentioned annotations:

- **STAPLE:** Simultaneous truth and performance level estimation (WARFIELD *et al.*, 2004). STAPLE is an expectation-maximization algorithm that analyzes different segmentations and calculates a probabilistic estimation of the real segmentation. Here, it was applied to cases where we had three types of segmentations: experienced rater manual annotation and two automated annotations (RODRIGUES *et al.*, 2020; BILLOT *et al.*, 2020).
- **Consensus:** In cases where manual segmentation was not performed, the consensus was created using the intersection of the two automated segmentation methods.

A specialist manually annotated the images from MiLI, IXI, CC59, and OASIS, leading to a total of 95 manually segmented volumes: 18 from CC359, 22 from IXI, 45 from MiLI (30 controls, and 15 patients), and 21 from OASIS (Figure 7.2).

### 7.3 Ethics Committee Approval

All data from MiLI used in preparation for this study were approved by the local ethical committee and fully anonymized (CEP/Conep, number 3435027). All participants were duly informed and signed a consent form agreeing to participate in the study. The use of public datasets (OASIS, IXI, CC359, and Kirby 21) is based on their respective licenses.

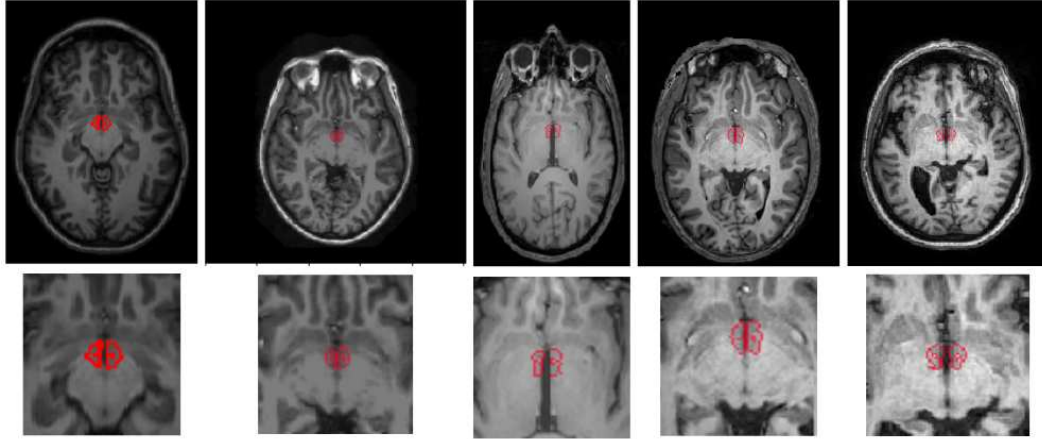


Figure 7.2 – Samples from datasets used on model training/testing depicting the high variability between them: axial slices highlighting the hypothalamus location (first row); Zoom of the highlighted region (second row). From left to right: OASIS, IXI, MiLI and CC359 samples

## 7.4 Manual Segmentation Protocol

The segmentation of the hypothalamus follows the anatomical landmarks proposed by pathological studies and has been validated for MRI-based studies (GABERY *et al.*, 2015; GORGES *et al.*, 2017; CARDOZO-HERNÁNDEZ *et al.*, 2020). In summary, the hypothalamus surrounds the third ventricle and is laterally delimited by the optic tract and hypothalamic sulcus. The longitudinal axis was delimited by the anterior commissure in more cranial slices and mammillary bodies in more caudal slices. Fornix, mammillary bodies, and optic tracts were excluded from the study area.

## 7.5 Dataset Availability

The MiLI (MICLab-LNI Initiative) dataset and the manual and automated annotations used in this project for IXI, OASIS, CC359, and MiLI are publicly available<sup>2</sup>.

<sup>2</sup> <<https://sites.google.com/view/calgary-campinas-dataset/hypothalamus-benchmarking>>

## 8 Hypothalamus Automated Segmentation Method

By using MiLI, we developed an automated segmentation method for the hypothalamus focused on T1w 1mm isotropic images.

### 8.1 Method

Notations: Defining an MRI volume as  $\mathbf{V}[\mathbf{D}_v \times \mathbf{H}_v \times \mathbf{W}_v]$ , the input patch  $\mathbf{I}[3 \times \mathbf{H} \times \mathbf{W}]$  is formed by three consecutive slices of  $\mathbf{V}$ , and we are interested in the segmentation of the central slice. The manual segmentation of  $\mathbf{V}$  is  $\mathbf{M}[\mathbf{D}_v \times \mathbf{H}_v \times \mathbf{W}_v]$ , and the label used for minimizing the network is  $\mathbf{L}[1 \times \mathbf{H} \times \mathbf{W}]$ , referring to the central slice of  $\mathbf{I}$ . The output of the network can be finally defined as  $\mathbf{O}[1 \times \mathbf{H} \times \mathbf{W}]$ . The final automated segmentation of the volume  $\mathbf{V}$  is  $\mathbf{A}[\mathbf{D}_v \times \mathbf{H}_v \times \mathbf{W}_v]$ .

Three architectures were used during the development of the final method, as listed as follows.

#### 8.1.1 EfficientSeg

In this work, we decided to use the EfficientNet-B4 (Figure 8.1), since it has the best balance between a number of parameters and computational cost (MATH-EWS *et al.*, 2020). It was developed to deal with input images  $\mathbf{I}[3 \times 224 \times 224]$  or superior, ending with a feature map of shape  $[1792 \times 7 \times 7]$  before the dense layer. To adapt it for segmentation, we removed the stem block composed of one 2D convolution of stride 2 and replaced it with a 2D convolution of stride 1. Also, we removed the final dense layers from the original EfficientNet-B4 and added four transposed convolutions (Figure 8.2). The final architecture has input  $\mathbf{I}[3 \times 112 \times 112]$  and output  $\mathbf{O}[1 \times 112 \times 112]$ .



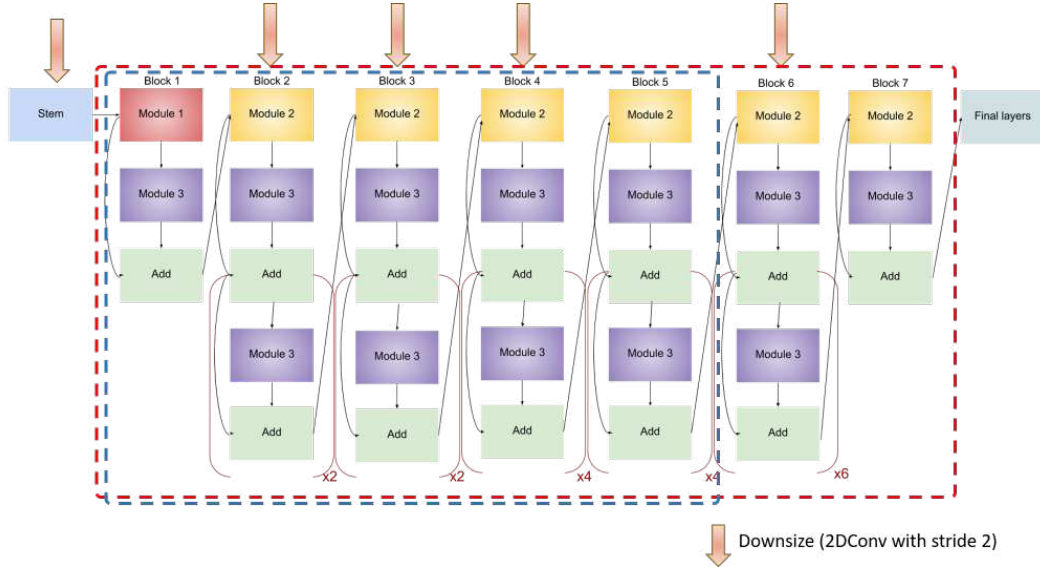


Figure 8.1 – EfficientNet-B4 architecture, based on an image from Agarwal (AGARWAL, ). Despite being originally a classification architecture, it was altered for segmentation purposes in this project. In red, blocks are used for EfficientSeg. In blue, the blocks used for Modified Eff-UNet.

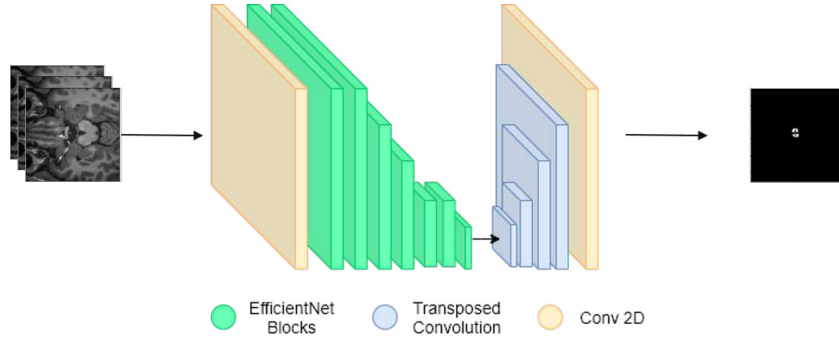


Figure 8.2 – EfficientSeg architecture

### 8.1.2 Modified Eff-UNet

Based on Baheti *et al.* (BAHETI *et al.*, 2020), we implemented a U-Net-based architecture with EfficientNet-B4 as the encoder (Figure 8.3). However, once we dealt with a small structure, we used fewer down-sampling layers (Hypothesis *a*). The architecture starts with input  $I[3 \times 112 \times 112]$  and has output

$\mathcal{O}[1 \times 112 \times 112]$ . At the end of the contraction path, the feature map has shape  $[160 \times 14 \times 14]$ . This is the base of our final architecture.

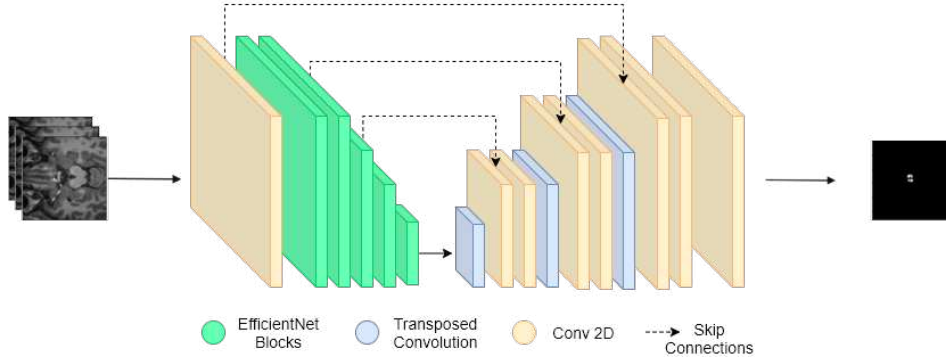


Figure 8.3 – Modified Eff-UNet architecture: it uses an EfficientNet-B4 as encoder of a U-Net

### 8.1.3 Correct2Seg

Using the Modified Eff-UNet, we implemented Correct2Seg, a Teacher-Student-based model (Figure 8.4), inspired by Yu *et al.* (YU *et al.*, 2019). We used two modules: segmentation and correction. Both modules were composed of Modified Eff-UNet architectures. The segmentation module has an input  $\mathbf{I}_s[3 \times 112 \times 112]$  and output  $\mathcal{O}[1 \times 112 \times 112]$ , whereas the Correction module has input  $\mathbf{I}_c[2 \times 112 \times 112]$  and output  $\mathcal{O}[1 \times 112 \times 112]$ .  $\mathbf{I}_c$  is composed of the concatenation of  $\mathcal{O}$  and the central slice of  $\mathbf{I}_s$ . The intuition behind the method is that the correction module will improve the output of the segmentation module based on the original image ( $\mathbf{I}_s$ ). The minimization of both modules was performed simultaneously. Once the correction module is used to improve the segmentation module (Hypothesis *e*), its output is only used during training. During inference, it is ignored, that is, only the output of the segmentation module is considered.

$$loss = WCE(S(\mathbf{I}_s), \mathbf{L}) + BCE(C(\mathbf{I}_c), \mathbf{L}) \quad (8.1)$$

being  $\mathbf{I}_s$  the input of the Segmentation module  $S$ ,  $\mathbf{I}_c$  the input of the Correction module  $C$  and  $\mathbf{L}$ , the label.  $WCE$  represents the Weighted Cross Entropy, used

on the segmentation module, and *BCE* the Binary Cross Entropy, used on the Correction module. We have run experiments using Dice Loss, however, the results using Cross Entropy were better. Therefore, all experiments in Section 8.3 were performed using the loss function presented in Equation (8.1). For the prediction, we use the output of the Segmentation module and ignore the Correction module.

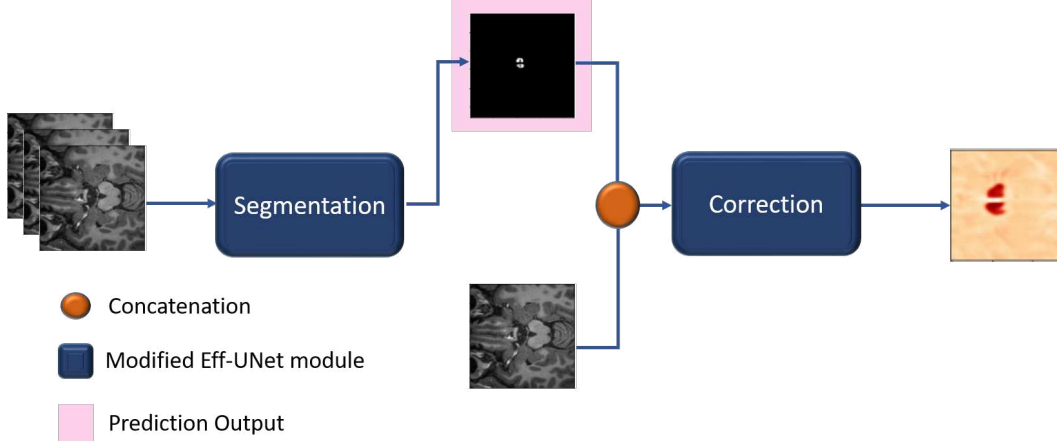


Figure 8.4 – Correct2Seg scheme proposed composed of two modules, Segmentation and Correction, trained together. The final prediction is the output of the segmentation module.

## 8.2 Evaluation Metrics

During the development of the methods, we used four different metrics to qualitatively analyze the results: the Dice coefficient (DC), average Hausdorff distance (AVD), Hausdorff distance (HD), and volume similarity (VS). Our choices were based on the hypothalamic characteristics (small structure with low contrast) and literature usage of the metrics.

- **Dice Coefficient:**

The DC is an overlap measure defined as follows:

$$DC = \frac{2 * |M \cap A|}{|M| + |A|} \quad (8.2)$$

With  $M$  being the manual annotation, and  $A$  the automated segmentation. The DC is sensitive to small segmentation and does not identify boundary errors. However, it can be used as a measure of reproducibility and is widely used for medical imaging segmentation analysis, being the most used metric in the medical imaging segmentation field (TAHA; HANBURY, 2015). DC results must be in the  $[0,1]$  range, where 1 is a perfect DC.

- **Hausdorff Distance:**

The HD measures the distance between two sets of points.

$$HD(A, M) = \max(h(A, M), h(M, A)) \quad (8.3)$$

where:

$$h(A, M) = \max_{a \in A} \min_{m \in M} \|a - m\| \quad (8.4)$$

With  $\|a - m\|$  being a norm such as Euclidean distance. Unlike DC, it can find boundary errors, which can occur owing to low contrast, and it is more robust for small structures, such as the hypothalamus. However, it is sensitive to outliers (TAHA; HANBURY, 2015). The smaller the HD between manual and automated segmentation, the better the automated segmentation.

- **Average Hausdorff Distance:**

As the name suggests, AVD is the averaged Hausdorff Distance over all points.

$$AVD(A, M) = \max(d(A, M), d(M, A)) \quad (8.5)$$

where:

$$d(A, M) = \frac{1}{n} \sum_{a \in A} \min_{m \in M} \|a - m\| \quad (8.6)$$

Similar to HD, it is also a spatial distance metric that is robust to small structures. However, on average, it is less sensitive to outliers (TAHA; HANBURY, 2015). The smaller the AVD between manual and automated segmentation, the better the automated segmentation.

- **Volume Similarity:**

Finally, the VS calculates the similarity between the two samples.

$$VS = 1 - \frac{||A| - |M||}{|A| + |M|} \quad (8.7)$$

With  $|X|$  being the module of  $X$ . Although it ignores borders and overlap, VS is a good metric for analyzing the segmentation volume when determining the volume of the structure is the main goal (TAHA; HANBURY, 2015). VS results must be in a  $[0,1]$  range, where 1 is a perfect VS.

## 8.3 Experiments

In our experiments, we addressed Hypotheses *a,c,d* and *e*. At first, we compared EfficientSeg and Modified Eff-UNet, since the first has more pooling layers than the latter. Then, we analyzed Correct2Seg, studying the impact of noisy labels on our results. For so, we checked the quantitative metrics from the same dataset we have trained on and then we used different datasets to assess the generalization. Therefore, we sometimes used STAPLE, which decreases noise, and other times, manual annotation from an inexperienced rater. Finally, we conducted a two-step training experiment, where we pretrained the network using noisy labels and trained it using manually labeled data. In total, we performed five different experiments using Correct2Seg (Table 8.1).

### 8.3.1 Choosing Correct2Seg modules

As explained in Section 8.1.3, the Correct2Seg architecture consists of two Eff-UNet modules. We conducted an experiment to determine the best architecture for these modules. Hypothesis *a* addresses the size of the last feature map in the contraction path. Given the small size of the hypothalamus, networks with more pooling layers might impair border detection. Consequently, we employed the Modified Eff-UNet (Section 8.1.2), a U-Net-based architecture with fewer down-samplings than EfficientSeg. The final dice score on the test set was  $0.739 \pm 0.06$  for Eff-UNet, compared to  $0.729 \pm 0.06$  for EfficientSeg.

Table 8.1 – Summary of all conducted experiments on Correct2Seg. While developing the model, five models were trained using different data/annotations. For assessing model reliability, three experiments were conducted.

	Experiment	Objective	Training Set		
			Data	Quantity	Label
Model Development	Model I	Pre-training using noisy labels	IXI/CC359	879 <sup>a</sup>	Consensus
			MiLI	407	Manual <sup>b</sup>
	Model II	Pre-training dealing with noisy labels	IXI, CC359	879 <sup>a</sup>	Consensus
			MiLI	407	STAPLE
	Model III	Train using Model I weights	IXI, CC359	27	Specialist Manual Annotation
			MiLI	31	
	Model IV	Train using Model II weights	IXI, CC359	27	
			MiLI	31	
	Model V	One Step Training	IXI, CC359	27	
			MiLI	31	
	Experiment	Objective	Test Set		
			Data	Quantity	
Model Reliability	Test-Retest	Analyze model imprecision	Kirby 21	42	
	Inter-Raters Metrics	Compare manual segmentation performed by two different specialists	MiLI (patients)	5	
	Generalization Test	Assess metrics on a different dataset, not used on training	OASIS	10	

<sup>a</sup>Some images from IXI were excluded from the experiments because the automated methods could not segment the hypothalamus returning only a few voxels or even a blank image. This happened on images with the poorest contrast and resolution. <sup>b</sup>Inexperienced rater manual annotation

Our results showed improvement over EfficientSeg. However, the comparison is not entirely direct since we also made additional modifications to the network, including adding convolutional layers and skip connections in the decoder.

### 8.3.2 Correct2Seg Model development

During the model development, we used *Correct2Seg*, presented in Section 8.1.3. To increase data variability during training, we used random crop, random rotation (from  $-10^\circ$  to  $10^\circ$ ), and elastic transformations as data augmentation in a batch size of 128. Our network was optimized using *Adam* and a learning

rate of  $5 \times 10^{-3}$ . For all experiments, we used the same hyperparameters, and as pre-processing, input images were normalized on a range  $[-1,1]$ . Images were not registered in a standard space. As a pre-processing step, we reoriented all images to the LPS (Left-Posterior-Superior) orientation to ensure that all inputs were in the same view. We also used a post-processing step to remove the prediction's noise. After segmenting all slices of volume  $\mathbf{V}$ , the post-processing step selects the largest 3D connected component (CC).

### 8.3.2.1 Pre-training

According to hypothesis *c*, by increasing the amount of training data, part of the noise present dissipates, even if the label is not manual. Therefore, the goal of the first experiment was to analyze the ability of the network to learn using a large amount of data annotated with noisy labels. We trained two models (I and II) using different labels (Table 8.1):

- Manual on MiLI: MiLI labels are manual annotation by inexperienced raters, therefore we are using the noise labels;
- Staple on MiLI: MiLI labels are STAPLE's, created according to Section 7.2, therefore we are dealing with the noise labels, by trying to reduce the noise.

In both cases, IXI and CC359 labels are consensus from automated segmentations, according to Section 7.2.

### 8.3.2.2 Training

This experiment analyzes the general improvement that two-step training brings when using different labels. We used the pre-trained models from previous experiments (Section 8.3.2.1) as the starting point. From each model, we trained two different models (Models III and IV) using specialist manual annotation.

### 8.3.2.3 One-Step training

In this experiment, we performed one-step training to obtain Model V and analyzed the impact of using fewer images annotated with high-quality labels. We used the same dataset used in Experiment 8.3.2.2, but without the network pre-training.

## 8.3.3 Model Reliability

### 8.3.3.1 Inter-Rater Metrics

In addition to being time-consuming, manual hypothalamus segmentation may have high variability among raters for small structures that are difficult to see on MRI. This experiment compared the variability between two manual annotations from different specialists and our automated segmentation.

### 8.3.3.2 Test-retest

We performed a test-retest to assess the model’s reproducibility. Reproducibility can be defined as the agreement of model results when it is applied uniformly and repeatedly to invariant objects (MENG, 2020). We used the Kirby 21 dataset, which contained 21 healthy subjects scanned twice on the same day. When the acquisition interval is short, the segmentation results for the same subjects should be similar. Therefore, a great difference may imply method imprecision.

### 8.3.3.3 Generalization Test

As our method is openly available, it is essential to check its ability to generalize to different datasets not seen by the network during training. In this experiment, we used 21 images from the OASIS dataset, manually segmented by a specialist following the protocol explained in Section 7.4.



## 8.4 Results

### 8.4.1 Model Development

We quantitatively analyzed our models using the metrics presented in Section 8.2. Because the training process is non-deterministic, results presented here are an average of three runs for each model. This analysis was performed to improve the confidence level of the results (Figures 8.5 and 8.6, Table 8.2).

Table 8.2 – Results for Model Development and Generalization Experiments. DC = Dice Coefficient; VS = Volume Similarity; AVD = Average Hausdorff Distance. Average values and standard deviations were measured among the metrics of three different runs. Results of Billot *et al.* are from their inference on OASIS dataset.

Experiment	Dataset	DC	VS	AVD	Hausdorff
Model I	IXI/CC359/MiLI	$0.63 \pm 0.02$	$0.76 \pm 0.01$	$0.58 \pm 0.06$	$5.43 \pm 0.78$
	OASIS	$0.58 \pm 0.02$	$0.84 \pm 0.01$	$0.79 \pm 0.29$	$5.50 \pm 1.63$
	All	$0.61 \pm 0.02$	$0.80 \pm 0.02$	$0.67 \pm 0.12$	$5.55 \pm 0.36$
Model II	IXI/CC359/MiLI	$0.65 \pm 0.01$	$0.77 \pm 0.03$	$0.47 \pm 0.01$	$4.23 \pm 0.14$
	OASIS	$0.54 \pm 0.03$	$0.76 \pm 0.04$	$0.68 \pm 0.08$	$4.80 \pm 0.30$
	All	$0.60 \pm 0.01$	$0.80 \pm 0.03$	$0.56 \pm 0.04$	$4.47 \pm 0.22$
Model III	IXI/CC359/MiLI	$0.82 \pm 0.01$	$0.93 \pm 0.02$	$0.24 \pm 0.03$	$2.82 \pm 0.64$
	OASIS	$0.65 \pm 0.04$	$0.86 \pm 0.02$	$0.45 \pm 0.10$	$3.41 \pm 0.66$
	All	$0.74 \pm 0.02$	$0.91 \pm 0.01$	$0.32 \pm 0.04$	$3.03 \pm 0.4$
Model IV	IXI/CC359/MiLI	<b><math>0.830 \pm 0.001</math></b>	<b><math>0.950 \pm 0.001</math></b>	<b><math>0.190 \pm 0.001</math></b>	<b><math>2.23 \pm 0.08</math></b>
	OASIS	<b><math>0.74 \pm 0.01</math></b>	<b><math>0.91 \pm 0.01</math></b>	<b><math>0.32 \pm 0.05</math></b>	<b><math>3.31 \pm 1.07</math></b>
	All	<b><math>0.79 \pm 0.01</math></b>	<b><math>0.940 \pm 0.001</math></b>	<b><math>0.24 \pm 0.01</math></b>	<b><math>2.69 \pm 0.51</math></b>
Model V	IXI/CC359/MiLI	$0.81 \pm 0.01$	$0.94 \pm 0.01$	$0.30 \pm 0.08$	$4.10 \pm 1.78$
	OASIS	$0.63 \pm 0.02$	$0.71 \pm 0.08$	$0.81 \pm 0.37$	$4.90 \pm 1.33$
	All	$0.73 \pm 0.01$	$0.90 \pm 0.01$	$0.40 \pm 0.04$	$4.07 \pm 0.98$
Billot <i>et al.</i> (2020)	OASIS	0.63	<b>0.91</b>	0.50	4.19

### 8.4.2 Model Reliability

#### 8.4.2.1 Inter-Rater Metrics

To directly compare our best method with manual segmentation, we asked a second expert to manually segment the MiLI images. We used 14 images from our test set (5 patients and 9 controls). The results were quantitatively (Table 8.3) and qualitatively compared (Figure 8.7).

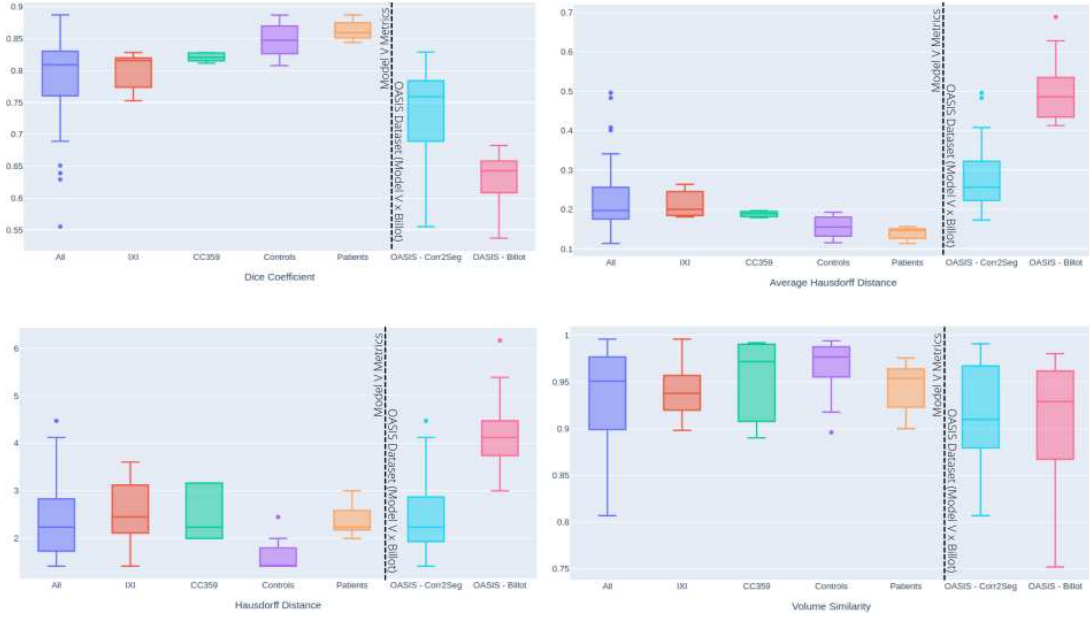


Figure 8.5 – Quantitative results of our best Correct2Seg model (Model IV), reported by dataset and comparative with Billot method.

Table 8.3 – Inter-Rater Metrics computed for 5 images from test set

Comparative	DC	VS	AVD	HD
<b>Rater1 VS Rater2</b>	0.71±0.04	<b>0.86±0.09</b>	0.31±0.04	2.93±0.51
<b>Automated Segmentation VS Rater 1</b>	<b>0.85±0.03</b>	0.95±0.03	<b>0.15±0.03</b>	<b>1.96±0.50</b>
<b>Automated Segmentation VS Rater 2</b>	0.73±0.04	0.86±0.08	0.27±0.04	2.67±0.66

#### 8.4.2.2 Test-retest

The Kirby 21 Dataset is composed of 42 images, 21 subjects scanned twice on the same day. Since each pair is composed by different acquisitions, the positioning of the subjects is different. To run the test-retest analysis, we first registered the images. The second acquisition of each patient was brought to the space of the first acquisition. Then, we ran Correc2Seg and Billot’s method on the images (Table 8.4). Both methods presented similar metrics with no statistical difference among them (p-value>0.05 for all metrics).

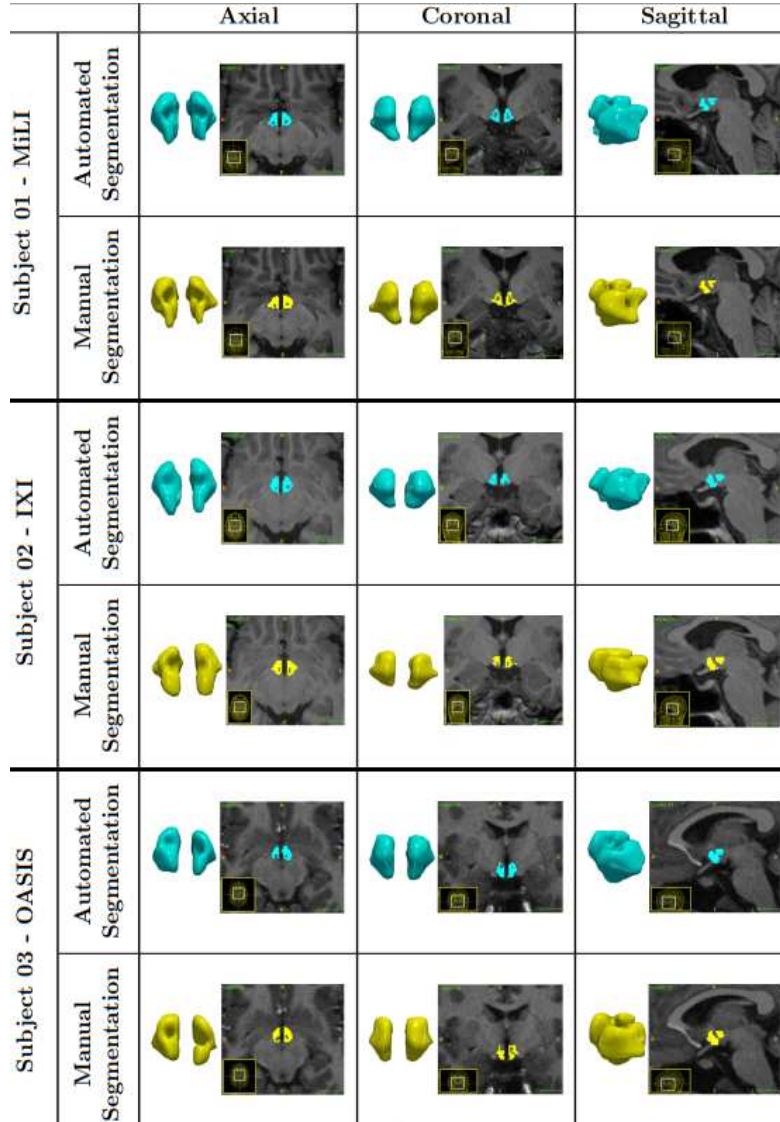


Figure 8.6 – Automated *versus* Manual segmentation. Qualitative results from our best-trained model (Model IV)

#### 8.4.2.3 Generalization Test

To analyze network generalization, we tested our models on the OASIS dataset and compared our generalization capability to other methods in the literature. The OASIS dataset was not used for training by either method. (Table 8.2 and Figure 8.8).

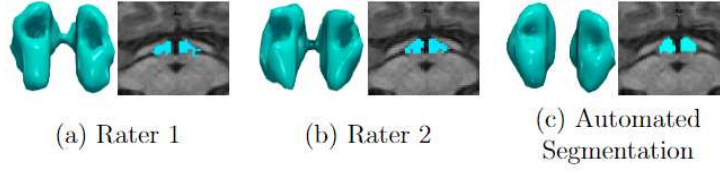


Figure 8.7 – 3D and 2D (axial) views from automated and manual specialist annotations. The automated segmentation can properly divide the right and left hypothalamus, while manual segmentation may join both sides. This may happen once the images do not have enough contrast for the human eye while the CNN model can better recognize the nuances.

Table 8.4 – Test-Retest metrics on Kirby21 dataset. The values correspond to the mean and standard deviation among the 21 pairs of acquisition. For Correct2Seg method, we used one of the three models trained on Model V experiments.

Method	DC	VS	AVD	Hausdorff
<b>Correct2Seg</b>	$0.88 \pm 0.06$	$0.97 \pm 0.03$	$0.13 \pm 0.07$	$1.47 \pm 0.61$
<b>Billot <i>et al.</i> (2020)</b>	$0.87 \pm 0.06$	$0.97 \pm 0.02$	$0.14 \pm 0.07$	$1.64 \pm 0.58$

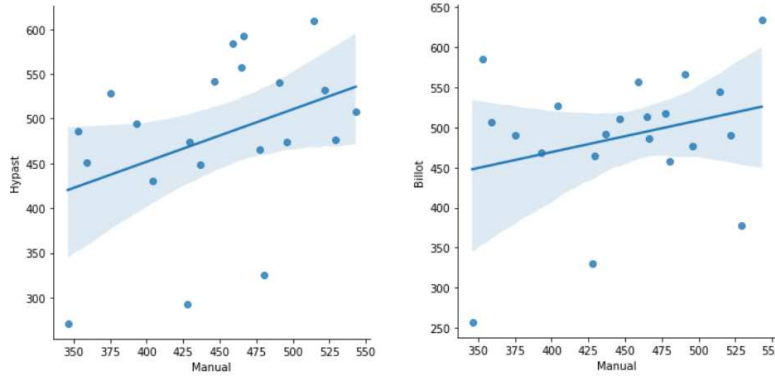


Figure 8.8 – Real VS predicted volumes on OASIS dataset. Volumes are measured in  $\text{mm}^3$ . Left: Correct2Seg (Model IV). Right: Billot *et al.* (2020)

## 8.5 Discussion

During our experiments, by comparing results from EfficientSeg and Modified Eff-UNet, we observed that using fewer downsampling layers yielded better results, supporting Hypothesis *a*. This improvement is likely due to the hypothalamus being a small structure. However, it is important to note that in this experiment, we also added skip connections between the encoder and decoder, which may have influenced the results.

In the model development experiments, we found that more data did not necessarily imply better metrics, not being able to confirm Hypothesis *c*. When we analyzed our Pre-training experiments (Table 8.2), both Models I and II could not achieve similar metrics as our one-step experiment (Model V), although the latter had almost  $21\times$  less elements in its training data compared to the datasets of the first two. The same can be verified using a generalization test. Model V generalized better on the OASIS, although it was trained with less data. This performance can be explained by the quality of the labels and protocols used for the training set annotation. In addition, in the pre-training experiments, we observed that STAPLE annotation improved the metrics. However, this occurred more prominently for the MiLI/IXI/CC359 datasets.

Analyzing the training experiments (Models III and IV), after a second step, we could improve our metrics compared to Model V. Here, we could use both plans in our favor: pre-training with a significant amount of data and training the network using good quality annotation. In addition, our best model (Model IV) was pretrained using STAPLE on the MiLI dataset, achieving the best results on the OASIS dataset. This reaffirms the importance of dealing with the noise presented on the labels. Once again, by analyzing the OASIS results, we compared our final metrics with other automated methods in the literature. Our model achieved a Dice coefficient of 0.83 on IXI/CC359/MiLI test images. Billot *et al.* (BILLOT *et al.*, 2020) reported a Dice coefficient of 0.84 on their test set. However, the comparison is not straightforward because their training and testing data are not the same as ours. When we predict images from OASIS,

a neutral dataset that is not used during the training of either of the two analyzed methods, both methods have a considerable performance loss. Our method reported a Dice coefficient of 0.74, and Billot *et al.* (BILLOT *et al.*, 2020) reported a Dice coefficient of 0.63. This performance difference was assessed with the use of the Wilcoxon paired test ( $p\text{-value} < 0.01$ ), which indicated statistical significance. Despite the loss in both methods, our network managed to maintain better generalization. Analysing predicted VS real volumes on OASIS dataset (Figure 8.8), we can see that both Correct2Seg (Model IV) and Billot *et al.* (2020) have similar behaviors. In both methods, the distribution returns a significant offset, that is, the predicted hypothalamus usually presents a larger volume than the manually segmented. This effect is possibly due to three factors: large use of data augmentation (also conducted by Billot *et al.* (BILLOT *et al.*, 2020)), data coming from different datasets in training, and label aggregation using STAPLE, which helps reduce noise.

Generalization loss may also occur because of OASIS’s lower resolution. We used 21 isometric samples from the first and second discs of OASIS, which are composed of 1.5T images (MARCUS *et al.*, 2007). Hence, with lower contrast, it is more challenging to identify the hypothalamus (Figure 7.2). This may also have influenced the final results of both methods. The boxplots (Figure 8.5) support this fact. We can verify on DC and AVD boxplots that MiLI, the only dataset solely composed of 3T images, presented the best results among the MiLI/CC359/IXI test set. On the other hand, half of IXI’s test set was composed of 1.5T images, making this dataset present the worst metrics among MiLI/CC359/IXI. By analyzing the boxplots, our method could segment patient and control cases for the MiLI dataset, maintaining a similar quality level: DC and AVD averages were the best among all groups, while Hausdorff distance and VS were comparable to controls in IXI and CC359.

Examining the qualitative results (Figure 8.6), we can observe that we need to evaluate our method using multiple metrics once the hypothalamus is a small structure. Subject 01 returned the best DC and AVD (0.89 and 0.12 respectively) of our best model (Model IV). In this case, the predicted volume was

593mm<sup>3</sup>, and the manually segmented volume was 577mm<sup>3</sup>, indicating an error of 2.8%. Subject 02 presented a DC and AVD of 0.82 and 0.19, respectively. Despite the worse DC, Subject 02 presented the best VS of our test set, with a predicted volume of 526mm<sup>3</sup> and a manually segmented volume of 530mm<sup>3</sup>, an error of 0.75%. We chose our best model during training based on the best DC for the validation set. However, it is interesting to analyze the final goal of the study to set the base metric. If the shape of the structure is essential (e.g., morphological studies), DC or AVD are applicable. However, if the final study focuses solely on volume variation, VS may be more interesting.

Besides generalization, we could assert our model's reliability by calculating inter-rater metrics and test-retest. First, for inter-rater metrics, we noticed a higher variability between the two specialists than when comparing our method and manual segmentation. For test-retest, our method achieved high DC and VS and low AVD which assesses the reproducibility of our method. Billot *et al.* achieved similar metrics on Kirby during test-retest and the Wilcoxon paired test showed no statistically significant difference among both methods' metrics.

One of the main challenges of hypothalamus segmentation is its low contrast when compared with surrounding tissues. The objective of this step is related to the generalization capability of the method and its use by physicians in neurological studies in 1.5T and 3T MR images, as they are, so far, the most widely used. However, the main limitation is that the method does not work across different MR sequences, requiring re-training and, consequently, more data. Also, so far, the method has focused on the whole structure, ignoring the segmentation of the hypothalamus subregions. Both limitations are addressed in the next step of the project.

## 9 Graphic User Interface Tool

Finally, to facilitate the use of the segmentation method, we developed a graphic user interface, HypAST<sup>1</sup>, the Hypothalamus Automated Segmentation Tool (Figure 9.1). Besides returning the automated segmentation of the hypothalamus, some of the tool functionalities are:

- **Volume Information:** A spreadsheet with the volume (in  $cm^3$ ) of all hypothalamus segmented.
- **Volume Analysis:** The user has access to a dispersion graph and boxplot, where possible outliers can be analyzed.
- **Texture Information:** A spreadsheet with histogram texture information of all hypothalamus segmented. The mean provides the average grayscale value of the image. Variance quantifies the deviation of gray pixel values from the mean. Skewness examines the symmetry of the histogram, determining whether there is a prevalence of light or dark pixels compared to the mean. A positive skewness implies an abundance of pixels with values below the mean, whereas a negative skewness indicates an abundance of pixels with values above the mean. Lastly, kurtosis measures the uniformity of pixel distribution relative to a normal distribution. Negative values denote a flat distribution, whereas positive values indicate a peaked distribution. Finally, the entropy measures the randomness of the gray level distribution (MILES *et al.*, 2013; PHARWAHA; SINGH, 2009).
- **Visualization Tool:** A window shows all original images masked with their respective segmentation.

The tool was developed with a focus on ease of use, catering to groups with diverse backgrounds such as computer scientists and physicians. No instal-

---

<sup>1</sup> <<https://github.com/MICLab-Unicamp/HypAST>>



lation is required. HypAST-GUI features three main buttons: "Open File," "Run Code," and "Save Segmentation."

- Click the "Open File" button to select .nii or .nii.gz files. A label will indicate the number of files opened.
- Click the "Run Code" button to generate the segmentation. Using a CPU (i5 8th gen), each volume will take approximately 6.5 seconds to process (the first segmentation may take slightly longer).
- Use the "Save Files" button to save the generated segmentation and attribute files.

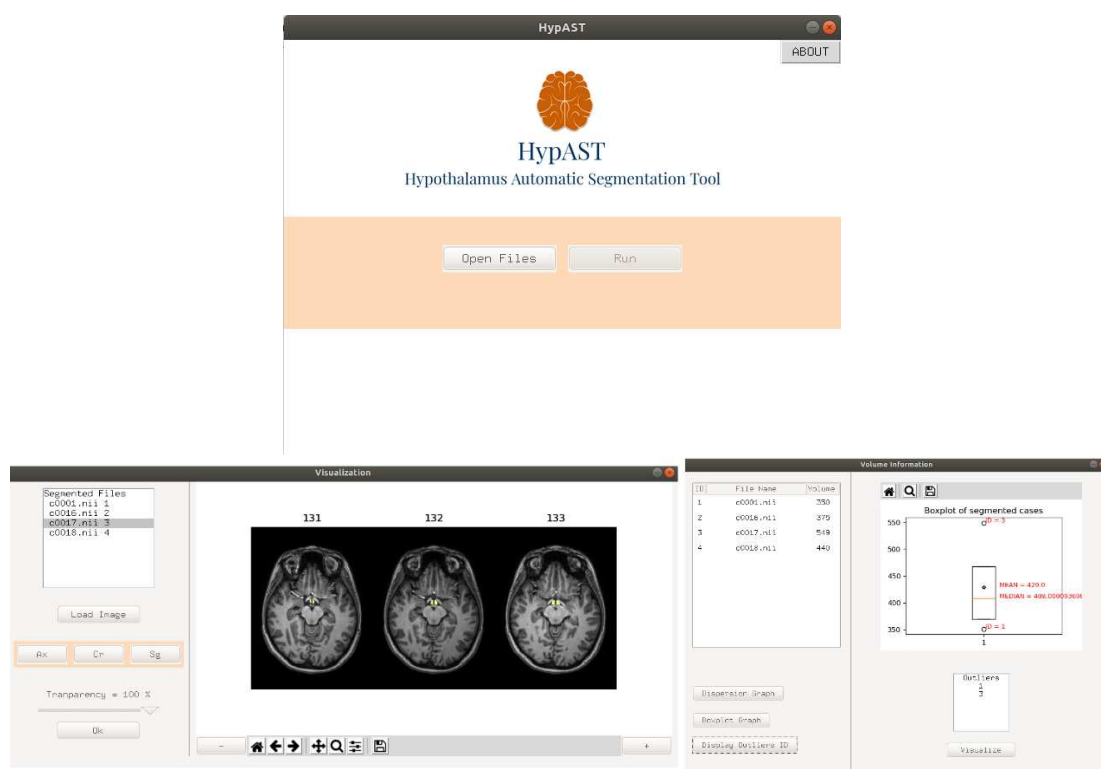


Figure 9.1 – Top Row: HypAST, the Hypothalamus automated segmentation tool. Bottom row: The tool offers volumetric analysis and returns histogram texture attributes of the hypothalamus

## 10 Considerations

The first stage of this project addressed Hypothesis *a*, *c*, *d* and *e*. Our results were able to confirm Hypothesis *a,c,e* and refute Hypothesis *d*.

During our experiments, we ended up with our final Correct2Seg model, which despite functioning only on T1w images, is capable of generalizing across datasets from different scanners.

In this stage, we also addressed Challenge 1, which concerns the lack of publicly available hypothalamus segmentation dataset. For that, we published MiLI, a dataset containing a total of 1381 hypothalamus segmentations (including both automated and manual) from four different databases (OASIS, IXI, CC359, MiLI), comprising isotropic and anisotropic images, as well as control and patient subjects.

Finally, we released HypAST, a tool that embedded Correct2Seg and that generates texture attribute values and volume of the hypothalamus, aiding studies across different groups.

Although this initial stage is crucial to assist in the study of hypothalamic images and understanding its role in neurodegenerative diseases, some gaps in the literature still exist. Initially, until the completion of this work, only one method was publicly available, and although it segmented the subregions of the hypothalamus, it was focused on only one type of MR sequence (T1w images). Therefore, the next stage of the project aims to develop a method that can be used across various MR sequences and resolutions without the need for retraining.

## Part III

Hypothalamus Segmentation across  
different sequences and resolution

# 11 Introduction

During the first stage of this project, we published the first publicly available dataset for hypothalamus segmentation (MiLI) and developed a method for T1w 1mm isotropic images, known as HypAST. However, upon reviewing the literature, we identified several gaps that still needed to be addressed. Initially, the public dataset encompassed the entire hypothalamus structure, but it is also crucial to study its sub-regions. Although T1w images provide good anatomical details of the structure, analyzing the hypothalamus in different sequences is also important. For instance, T2w images can be used to analyze gliosis in the structure, while Diffusion Tensor Imaging (DTI) and metrics such as Fractional Anisotropy (FA) can help study the microstructure of the hypothalamus and potential diffusion alterations of the tissue. By leveraging these different MRI sequences, researchers can gain comprehensive insights into the hypothalamus, ranging from macrostructural anatomy and volume to microstructural integrity and connectivity.

However, by the time we started this step of the project, there was no existing method capable of functioning with MRI sequences other than T1w and T2w. Therefore, the goal of this step was to fill the existing gap in hypothalamus segmentation in different sequences and MR resolutions. However, until then, the only publicly available dataset for hypothalamus segmentation was the MiLI, and due to the high cost of time and specialized labor, it would be unfeasible to create a dataset with manual segmentations in images of various sequences and resolutions. To solve this challenge, we developed the expansion part of the segmentation method, which consists of the development of:

- **HELM: Hypothalamus *ex vivo* Label Maps** is a dataset composed of label maps built from publicly available ultra-high resolution *ex vivo* MRI from 10 whole hemispheres, which can be used to develop segmentation methods using synthetic data. The label maps are obtained with a combination of manual labels for the hypothalamic regions and automated segmentations

for the rest of the brain, and mirrored to simulate entire brains (Hypothesis  $b$ ,  $f$  and  $g$ ).

- **H-SynEx:** A method for automated segmentation of hypothalamus subregions informed by ultra-high resolution *ex vivo* MRI, which generalizes across MRI sequences and resolutions without retraining (Hypothesis  $g$  and  $h$ ).

The details of dataset creation and method development are in the next sections.

## 12 HELM - Hypothalamus *ex vivo* Label Maps

In this section, we describe the creation of HELM - Hypothalamus *ex vivo* Label Maps, a dataset composed of label maps derived from *ex vivo* images, used to train H-SynEx.

### 12.1 Materials

HELM is derived from 10 *post mortem* ultra-high MRI acquisitions of brains provided by the DANDI Archive <sup>1</sup>. The *post mortem* images are openly available <sup>2</sup> and comply with all relevant ethical regulations (COSTANTINI *et al.*, 2023).

The MRI of the *ex vivo* brain hemispheres was obtained using multi-echo fast low-angle shot (ME FLASH or MEF) on a 7 T Siemens MR scanner with Repetition Time (TR) of 34 ms, time to echo (TE) of 5.65, 11.95, 18.25, and 24.55 ms, and field of view (FOV) of 192 mm by 81.3 mm. Before the MRI, the specimens were fixed in 10% formalin for a minimum of 90 days and packed in a 2% buffered paraformaldehyde solution. The images present an equal distribution of 5 male and 5 female control specimens who died of natural causes with no clinical diagnoses or neuropathology. As the images come from *post mortem* brains, the acquisition is free from motion artifacts and has a high resolution, ranging from 120 to 150 $\mu$ m isotropic, which improves the visualization of hypothalamic boundaries. Further details on the specimens and MRI acquisition can be found in the original publication (COSTANTINI *et al.*, 2023).

The age at the time of death ranges from 54 to 79 years, with an

---

<sup>1</sup> <<https://dandiarchive.org>>

<sup>2</sup> <<https://dandiarchive.org/dandiset/000026/draft/files?location=>>>

average of  $66.4 \pm 8.46$  years and they include only a single hemisphere of the brain. Although both halves of the hypothalamus exhibit symmetric tissue architecture (KISS *et al.*, 2020), the dataset includes four right-sided hemispheres and six left-sided hemispheres. Despite the limited size of the dataset, which is inherent to *ex vivo* images, this characteristic is compensated by the enhanced resolution of the data. For instance, it is possible to find in the literature other studies that use *ex vivo* MR images from only one hemisphere of the brain to generate segmentation of *in vivo* hippocampus and thalamus (IGLESIAS *et al.*, 2015; IGLESIAS *et al.*, 2018).

Using these MRI images as a starting point, we performed data pre-processing and generated automated whole-brain segmentation using unsupervised clustering. Unlike the hypothalamic labels, these automated segmentations are used only for image synthesis purposes and not as segmentation targets. Therefore, they can be noisy and not correspond directly to brain structures – so unsupervised clustering suffices. Our focus is on capturing context around the hypothalamus to produce synthetic intensities. Subsequently, we mirrored the hemispheres to create label maps that would serve as synthetic images for the input of the deep neural network without having to segment scans into hemispheres during testing.

## 12.2 Methods

### 12.2.1 Pre-processing of *ex vivo* scans

The first step of the method was the pre-processing of the images. The primary objective at this stage is to standardize the dataset and remove any background elements that could interfere with the subsequent steps (Figure 12.1).

- **Orientation:** Given that the primary objective of the dataset is to facilitate the development of automated segmentation methods, it was essential to ensure that all images were uniformly oriented. We decided to use the positive RAS (right-anterior-superior) standard. While our dataset includes images

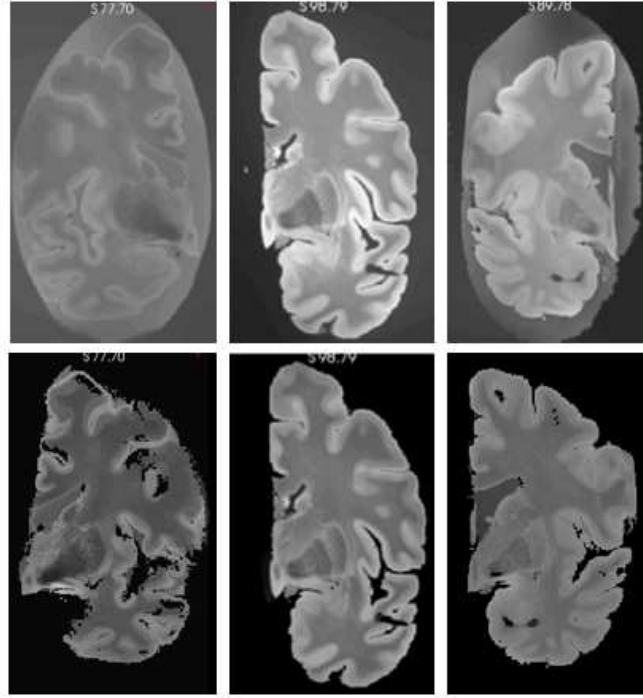


Figure 12.1 – Examples of original *ex vivo* images (up) and images after the pre-processing steps (down). First, we re-orient the images to positive RAS standard and remove non-cerebral elements from the background. Then, we resample the images voxels to  $0.3 \times 0.3 \times 0.3mm$  isotropic and perform bias field correction.

of both hemispheres, we flipped all right-sided hemispheres to the left side to ensure consistency and simplify subsequent preprocessing steps, resulting in a total of 10 left-sided hemispheres

- **Background segmentation:** The *ex vivo* brains were packaged in a bag for scanning, which is discernible in the images and undesirable in our model. To address this issue, we employed a Bayesian automated image segmentation approach adaptive to contrast (PUONTI *et al.*, 2016) to create a mask, which was then utilized to eliminate non-cerebral elements that were not related to brain structures.
- **Voxel resampling:** The original images have a voxel resolution ranging



from  $0.1$  to  $0.15\text{mm}^3$ . While this resolution assists in distinguishing structures during manual segmentation, it significantly prolongs image processing, particularly when the ultimate goal is to employ them in deep learning networks. Therefore, we adjusted the voxel resolution to a constant resolution equal to  $0.3 \times 0.3 \times 0.3\text{mm}$  isotropic, which provides a compromise between a high level of details and being storage- and processing-friendly

- **Bias Field Correction:** Finally, the last step in image pre-processing is the bias field correction (PUONTI *et al.*, 2016). This step is essential, as in the generation of the whole brain segmentation, we utilize an unsupervised clustering method that can be directly affected by the bias field.

### 12.2.2 Segmentation

In the second stage, we generated the label maps for the hypothalamus, its subregions, and the whole brain:

- **Hypothalamus manual segmentation:** All images were traced by one non-specialist. For manual segmentation, we relied on protocols focused on *in vivo* images as described in the literature. In particular, we followed the criteria described in Rodrigues *et al.* (RODRIGUES *et al.*, 2022) (whole hypothalamus) and Bocchetta *et al.* (BOCCHETTA *et al.*, 2015) (for the subregions). At  $0.3\text{mm}$  isotropic resolution, approximately 40-50 coronal slices include the hypothalamus.
  - **Whole structure:** The segmentation of the whole hypothalamus occurs on coronal view. To ensure the correct delineation of the landmarks, we also simultaneously inspected the sagittal and axial views. In *in vivo* images, the hypothalamus lies around the third ventricle. However, In the case of the *ex vivo* images, it is not always possible to distinguish the third ventricle, since we only have one hemisphere of the brain. Therefore, on the coronal view, we use the recess dorsal to the hypothalamus to define its most superior boundary (Figure 12.2).

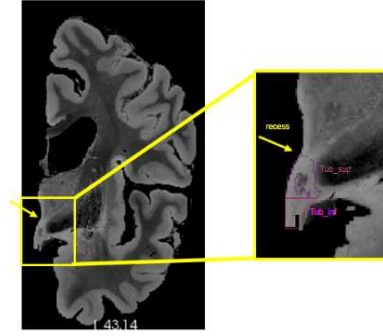


Figure 12.2 – Recess of the hypothalamus used for the delineation of superior boundary. Tub\_sup = Tuberal Superior subregion; Tub\_inf = Tuberal inferior subregion

Ventrally, the hypothalamus is defined by the optic tract and the hypothalamic sulcus in the most rostral slices. The most anterior coronal slice is defined as the one where the anterior commissure is visible, while the most caudal coronal slice is where the mammillary bodies (MB) are no longer visible. The mammillary bodies were included in the segmentation, while the fornix and optical tract were excluded from the segmentation.

- **Subregions:** We subdivided the hypothalamus into 5 subregions: Anterior inferior, Anterior Superior, Tuberal inferior, Tuberal Superior, and Posterior. While the manual segmentation of the whole hypothalamus was performed using the structure borders, the division of the subregions was done geometrically, using specific landmarks based on Bochetta *et al.* (BOCCHETTA *et al.*, 2015) to find the division between them (Table 12.1). An exception is the posterior subregion, which was segmented using the mammillary body morphological borders. As we only have one hemisphere for each brain, we could not segment both hypothalami for each subject. The most rostral slice of the anterior subregion coincides with the most rostral slice of the hypothalamus when looking at the coronal plane. On *in vivo* images, the anterior subregion is included from the most rostral coronal slice of the hypothalamus to

the most rostral part of the infundibulum. However, for *ex vivo* images we used the anterior commissure visible from the sagittal view as a landmark to delineate the most caudal part of the anterior regions (Figure 12.3(a)). The tuberal subregions begin posteriorly to the coronal slice where the anterior regions are visible (as defined by the anterior commissure sagittally) and extend to the most rostral slice where the MB are visible, which are included in the posterior subregion (Figure 12.3(b)). To delineate the superior and inferior portions of both the anterior and tuberal subregions, we drew a horizontal line on the coronal slice connecting the most medial to the most lateral point of the hypothalamus (Figure 12.3(c,d)).

Table 12.1 – Landmarks used for hypothalamus division into subregions

	<b>Anterior</b>	<b>Tuberal</b>
<b>Most rostral landmarks</b>	Most rostral slice of the hypothalamus on coronal view	Posterior to last coronal slice where anterior is visible
<b>Most caudal landmarks</b>	Defined by the anterior commissure on sagittal view (Figure 12.3(a,b))	Most rostral slice where mammillary bodies are visible
<b>Superior/ Inferior</b>	Horizontal line on the coronal slice connecting the most medial to the most lateral point of the hypothalamus (Figure 12.3(c,d)).	

- **Whole brain segmentation:** Unlike the hypothalamic subregions, semantic meaning is not necessary for the whole brain labels. That means we do not need to correctly delineate the morphological borders of the other structures. We just need to establish context around the hypothalamus to generate the synthetic images. We decided to use k-means, a non-supervised clustering method, to model the non-hypothalamic tissue based on grayscale levels of the pre-processed images. Seeking to increase data variability, we ran this step for values of k ranging from 4 to 9. As a result, for each of the 10 images, we have 6 different maps, totaling 60 label maps (Figure 12.4)
- **Segmentation merge:** Finally, it is necessary to merge both segmentations. In this stage, we needed to ensure that there would be no discrepancy during the overlay process. To achieve this, we employed mathematical morphol-

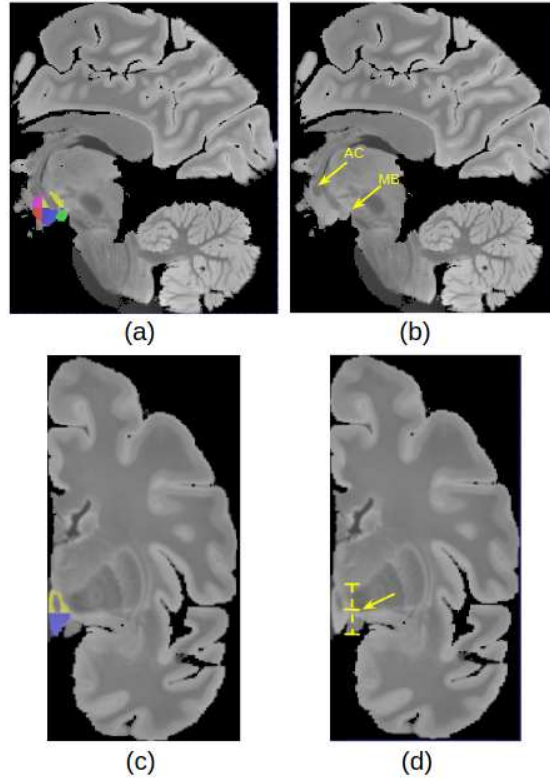


Figure 12.3 – Segmentation protocol: (a) Subregions delineation: Anterior superior (pink), anterior inferior (red), tuberal superior (yellow), tuberal inferior (blue) and posterior (green) (b) Sagittal landmarks for sub-region delineation (AC = Anterior Commissure and MB=mamillary body) (c) Coronal View of superior and inferior tuberal subregions (d) Coronal landmark for superior/inferior division

ogy closing to approximate delineation of the fornix and erosion to eliminate false positive voxels in the third ventricle area (Figure 12.5).

### 12.2.3 Hemisphere mirroring

The *ex vivo* images were acquired from single brain hemispheres. Given that the hypothalamus is situated in the most medial part of the brain, a model covering a single hemisphere lacks contextual information regarding its surroundings. To address this concern, we mirrored the images to generate a complete brain.

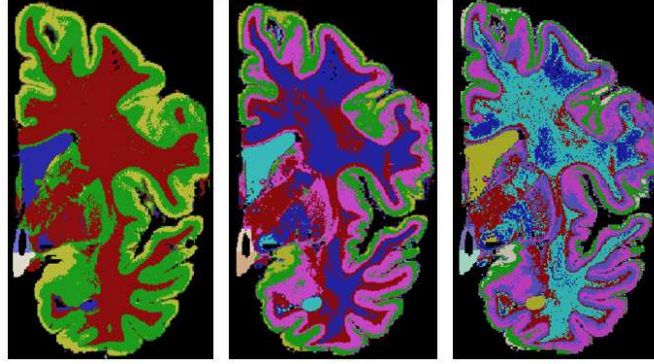


Figure 12.4 – Examples of three different label maps derived from the same image. From left to right:  $k = 4$ ,  $k = 6$ ,  $k = 8$

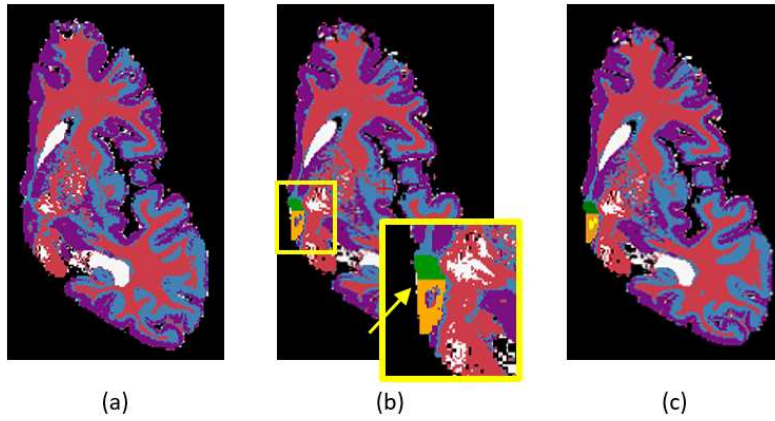


Figure 12.5 – (a) Whole brain segmentation: example with  $k=4$  (b) Manual segmentation simply overlapping the whole brain segmentation. We can see that there are a few inconsistent voxels, that should be labeled either as hypothalamus or background that have different labels. (c) To fix these inconsistencies, a mathematical morphology-based algorithm is applied.

In the mirroring process, two major concerns needed to be addressed: the original and mirrored hemispheres should not overlap, and they should be as close as possible to each other (Figure 12.6).

To tackle this issue, we encode the two constraints into softened versions and combine them into a weighted cost function that is optimized with respect to

a rigid transformation  $T$ :

$$\hat{T} = \underset{T}{\operatorname{argmin}} \sum_{i \in \Omega} \delta[x(v_i; T) > 0] x(v_i; T) - \alpha \sum_{i \in \Omega} x(v_i; T) \quad (12.1)$$

where  $\Omega$  is the brain mask,  $x(v_i; T)$  is the  $x$  real-world coordinate after rigid transformation with  $T$ ,  $\delta$  is Kronecker's delta, and  $\alpha$  a trade-off value that changes according to the image.

The rationale behind the cost function is that we want positive values of  $x$  to be penalized (Figure 12.6 (a)), so the images do not overlay (Figure 12.6 (d)). However, the image should not be too distant from the  $x=0$  axis (Figure 12.6 (b)) to not create an unrealistic gap between both hemispheres (Figure 12.6 (e)). Therefore, the larger the values of the absolute  $x$ , the more the function is penalized. For the latter, we added a trade-off value, which penalizes overlaps more heavily than gaps. For the last step, we mirror the transformed hemisphere and mesh both the right and left hemispheres into one image, ending up with 10 subregions, since we now have both right and left sides of the brain (Figure 12.6(f)).

## 12.3 Manual segmentation quality assessment

To test the intra-rater reliability of the manual segmentations, one non-specialist manually segmented 5 out of the 10 images twice, using the protocol described in Section 12.2.2. The images were randomly selected from the total sample and were re-segmented blindly, with the same software and computer settings, after four months from the first manual segmentation. For evaluation, we used the Dice Coefficient and Average Hausdorff Distance (Table 12.2).

The small size of the structure and the minimal contrast between adjacent tissues contribute to generally low intra-rater metrics for the hypothalamus. For example, Estrada *et al.* (ESTRADA *et al.*, 2023) found an intra-rater Dice coefficient of 0.82 for the entire structure based on in vivo images with a resolution of 0.8mm<sup>3</sup>. Similarly, Billot *et al.* (BILLOT *et al.*, 2020) reported a global Dice coefficient of 0.89 using in vivo images with a resolution of 1mm<sup>3</sup>. In both studies,

the authors highlighted the challenges associated with manual delineation due to the small size and low contrast of the hypothalamus.

Table 12.2 – Intra-rater metrics (median) for 5 subjects.

	Dice Coefficient	Average Hausdorff Distance
Whole	0.82	0.37
Anterior	0.73	0.63
Tuberal	0.78	0.37
Posterior	0.78	0.31

## 12.4 Data Usage and Availability

The label maps described here could be employed for training different networks dedicated to the hypothalamus or different brain structures. To achieve this, three steps need to be followed:

- **Manual segmentation:** Manually segment the target structure. Given that there are only 10 images, the manual segmentation will not demand as much effort as the ones typically used in supervised learning.
- **Merge with whole brain segmentation:** Merge the manual segmentation with the provided whole brain segmentation.
- **Mirroring:** Run the provided mirroring codes to generate the final label maps.

We also encourage the usage of this dataset in tasks other than segmentation. For instance, we can find in the literature the use of synthetic images applied to registration (IGLESIAS, 2023) and conversion of different MRIs into high-resolution T1w scans (IGLESIAS *et al.*, 2021). HELM is openly available for download <sup>3</sup>.

<sup>3</sup> <[https://www.nitrc.org/projects/hsynex\\_data/](https://www.nitrc.org/projects/hsynex_data/)>

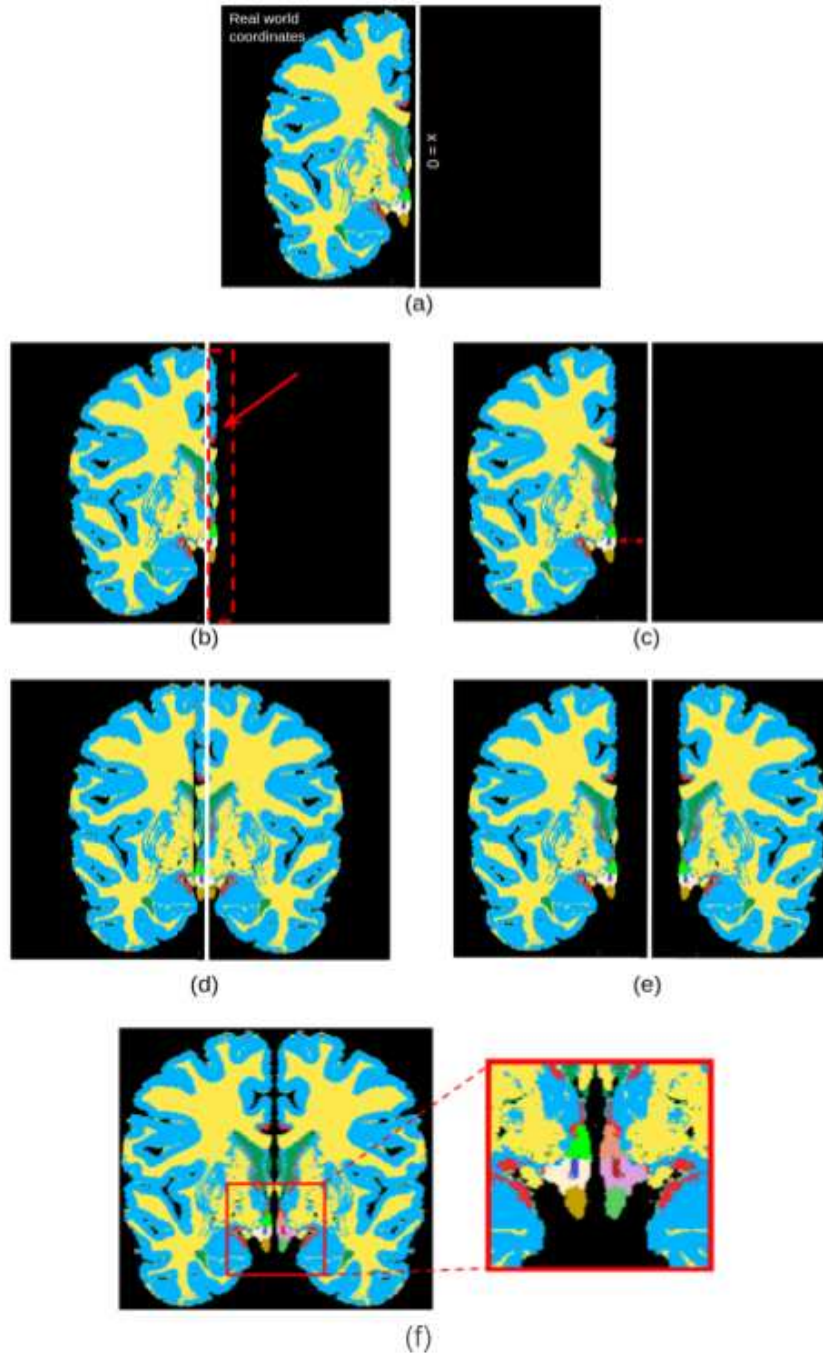


Figure 12.6 – Label map creation: Following the segmentation step, a half-brain label map is generated (a). However, given the hypothalamus’s central location within the brain, mirroring is essential to provide contextual information. For the mirroring process, translation and rotation are applied to the RAS coordinates. This involves moving the brain in close proximity to the  $x=0$  axis from the negative side, without surpassing into the positive side. Essentially, a final cost function is computed, penalizing positive values of  $x$  (b) and high negative values. Finally, we prevent the overlap between brain hemispheres (d) and also prevent them from ending up at unnaturally distant positions (e).



## 13 H-SynEx

Here, we describe the development of H-SynEx, a method for hypothalamus segmentation that is capable of working across different MR sequences and resolutions.

### 13.1 Materials

Here we describe the datasets used for the development of the method and those used for the experiments.

#### 13.1.1 Training Data

The data used for training H-SynEx comprises synthetic images derived from HELM. By using a label map derived from ultra-high resolution data, we intend to increase the final automated segmentation quality when comparing with *in vivo*-based models, especially on the structure borders, that suffer more with partial volume in *in vivo* images (Hypothesis *b* and *f*). Details on the synthetic image generation are detailed in Section 13.2.1.

#### 13.1.2 Test Data

The method evaluation relies on *in vivo* images from 6 different datasets (Table 13.1):

- **FreeSurfer Maintenance Dataset(FSM)** (GREVE *et al.*, 2021): Composed of 29 subjects from which 7 were used for validation and 22 for testing. For each subject, we have T1w, T2w, PD, FA, and qT1 acquisitions (Figure 13.1). FSM contains manual labeling for the whole hypothalamus and its subregions. This dataset was approved by the Massachusetts General Hospital Internal Review Board for the protection of human subjects and all

subjects gave written informed consent. All other datasets used for training and inference are openly available.

- **MiLI** (RODRIGUES *et al.*, 2022): Explained in Section 7, The MICLab-LNI Initiative comprises manual and automated segmentations of the entire hypothalamus conducted on T1w images. However, it lacks segmentations for hypothalamic subregions. It includes subjects from various open datasets such as MiLI, OASIS (LAMONTAGNE *et al.*, 2019), and IXI <sup>1</sup>. We only used the manually segmented images, totaling 55 from MiLI (30 controls and 25 ataxia patients), 23 from OASIS, and 19 from IXI. For the latter dataset, as it also encompasses T2w and proton density (PD) acquisitions, we incorporated these modalities in our experiments.
- **ADNI** (MUELLER *et al.*, 2005): We used a total of 572 controls (280 male and 292 female with average age of  $75.5 \pm 6.4$  and  $73.6 \pm 6.01$ , respectively) and 271 Alzheimer’s disease (AD) patients (143 male and 98 female with average age of  $75.34 \pm 7.6$  and  $73.8 \pm 7.6$ , respectively) for both T1w and fluid attenuated inversion recovery (FLAIR) modalities. The ADNI dataset does not have manual segmentation of the hypothalamus.
- **NIFD** <sup>2</sup>: From the Neuroimaging in Frontotemporal Dementia dataset, we used 111 controls (49 male and 62 female with average age of  $61.8 \pm 7.4$  and  $63.4 \pm 7.8$ , respectively) against 74 behavioral variant frontotemporal dementia (bvFTD) patients (51 male and 23 female with average age of  $61.16 \pm 5.8$  and  $62.4 \pm 7.7$ , respectively). The NIFD dataset does not have manual segmentation of the hypothalamus.

<sup>1</sup> <<https://brain-development.org/ixi-dataset/>>

<sup>2</sup> <<https://ida.loni.usc.edu/collaboration/access/appLicense.jsp>>

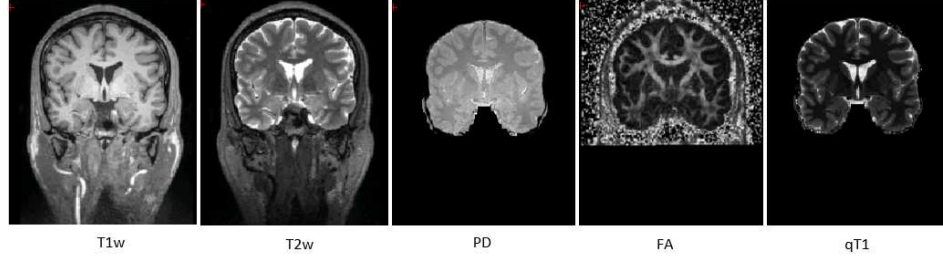


Figure 13.1 – Example of different modalities (FSM dataset)

Table 13.1 – Datasets used for model validation and testing; WS: Whole Structure, SR: Subregion

	Dataset Name	Sequence type	Number of Acquis.	Number of Subjects	Manual Segmentation Content	Segmentation Protocol
Validation	FSM	T1w, T2w, PD, FA, qT1	35	7 Controls	WS/SR	WS:ScLimbic (GREVE <i>et al.</i> , 2021) SR:Bocchetta <i>et al.</i> (BOCCHETTA <i>et al.</i> , 2015)
Testing	FSM	T1w, T2w, PD, FA, qT1	110	22 Controls	WS/SR	WS:ScLimbic (GREVE <i>et al.</i> , 2021) SR:Bocchetta <i>et al.</i> (BOCCHETTA <i>et al.</i> , 2015)
	MiLI	T1	55	30 Controls 25 Patients	WS	WS:Rodrigues <i>et al.</i> (RODRIGUES <i>et al.</i> , 2022)
	MiLI-OASIS	T1	23	23 Controls		
	MiLI-IXI	T1w, T2w, PD	57	19 Controls		
	ADNI	T1w, FLAIR	1686	572 Controls 271 AD Patients	No manual segmentation	—
	NIFD	T1	185	111 Controls 74 bvFTD patients		

## 13.2 Methods

### 13.2.1 Data Preprocessing

The training data comprises synthetic images generated from HELM. Despite the image synthesis being conducted during the model training, several steps are undertaken before training to create the final training images:

- **Creation of label maps:** The process of creating HELM, as described in Chapter 12.
- **Find MNI coordinates:** Using the label maps  $L[D \times H \times W]$ , from HELM, we generate a Gaussian image  $G[D \times H \times W]$  that simulates a T1w MRI. Subsequently, we register  $G$  to the MNI space using NiftiReg and obtain the MNI coordinates  $C[3 \times D \times H \times W]$  of the registered image.  $C$  will

be used during training to assist the network in locating the hypothalamus. It is important to emphasize that during training,  $L$  remains unregistered.

- **Crop:** We crop  $L$  and  $C$  around the hypothalamus, resulting in two standardized arrays,  $L_{\text{crop}} [200 \times 200 \times 200]$  and  $C_{\text{crop}} [3 \times 200 \times 200 \times 200]$ , which corresponds to a field of view of  $60 \times 60 \times 60 \text{mm}$ .
- **One-hot array:** We convert  $L_{\text{crop}}$  into a one-hot array  $L_{\text{one}} [V \times 200 \times 200 \times 200]$ , with  $V$  being the number of labels presented on  $L_{\text{crop}}$ .  $V$  varies according to the  $K$  labels employed at the whole brain segmentation.

### 13.2.2 Training

**Synthetic Images Generation** The synthetic image generation is performed on the fly, during training (Figure 13.2). At each iteration, one of the training label maps,  $L_{\text{one}}$ , is randomly selected. Then, we apply aggressive geometric augmentation that encompasses random crop, rotation, and elastic transformation on  $L_{\text{one}}$  and  $C_{\text{crop}}$ , ending up with  $L_{\text{trans}} [V \times 160 \times 160 \times 160]$  and  $C_{\text{trans}} [3 \times 160 \times 160 \times 160]$ , respectively. Next, we use the generative model proposed by SynthSeg (BILLOT *et al.*, 2023) based on Gaussian Mixture Models conditioned on  $L_{\text{trans}}$ , using randomized parameters for contrast and resolution to create the final synthetic images  $S [160 \times 160 \times 160]$  (Figure 13.3). The target  $T [V \times 160 \times 160 \times 160]$ , derived from  $L_{\text{trans}}$ , is also created during this step. To assist training, we use an Euclidean distance map,  $E [V + 1 \times 160 \times 160 \times 160]$  derived from  $T$ , which has been proven to help locate boundary features during segmentation tasks (LIU *et al.*, 2022b).  $E$  will be part of the loss function and is only employed during training, not being necessary during inference.

The final input of the network,  $I [4 \times 160 \times 160 \times 160]$ , is the concatenation of  $S$  and  $C_{\text{trans}}$ .

**Training architecture** Despite the good performance of Correct2Seg, we decided at this stage of the project to use a 3D architecture to better leverage synthetic data. Given that a 3D Modified Eff-UNet would be computationally

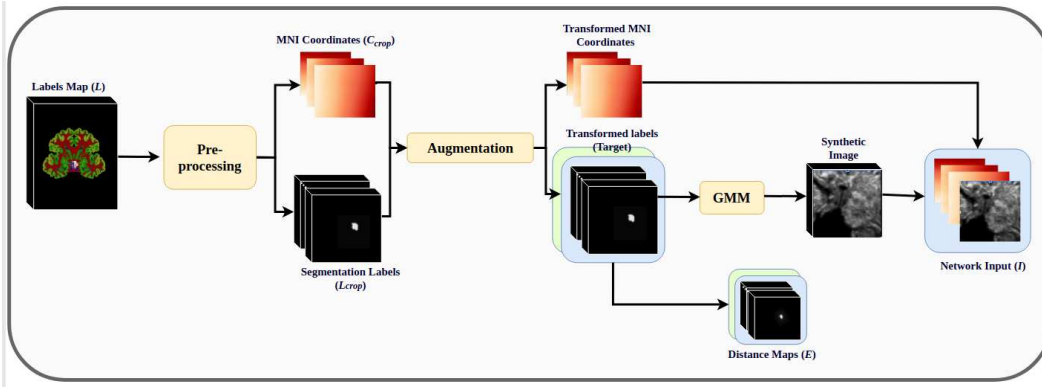


Figure 13.2 – Generation of synthetic images: The synthetic images  $S$  are generated using the label maps from the *ex vivo* images.

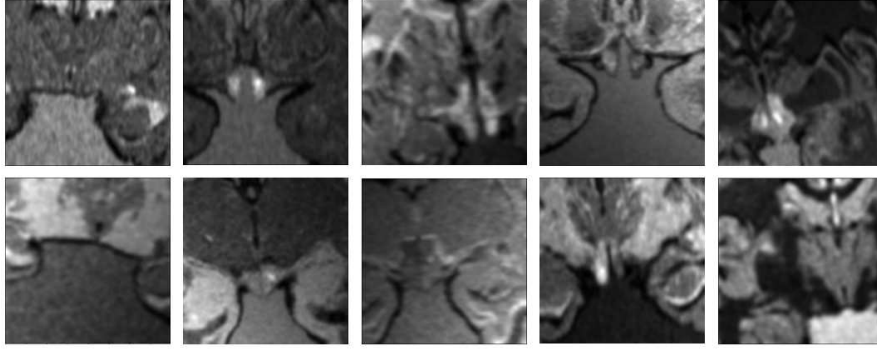


Figure 13.3 – Examples of coronal slices from 3D synthetic images used as input. The images shown here came from the label maps cropped around the hypothalamus. The use of aggressive data augmentation along random contrast values on the generative model results in large variability in the appearance of the input images.

costly, we opted to use a 3D U-Net as the module instead. Two distinct sub-models were trained separately, one for the entire hypothalamus ( $M_{hyp}$ ) and another specifically for its subregions ( $M_{sub}$ ) (Figure 13.4(b)).

Both  $M_{hyp}$  and  $M_{sub}$  are 3D-UNets (WOLNY *et al.*, 2020; ÇIÇEK *et al.*, 2016), however, in both cases, we added a skip connection between the input channels referring to  $C_{trans}$  and the final convolutional block to ensure that the original positional encoding is readily available at full-resolution also in the

decoder.  $M_{hyp}$  receives  $I$  and  $E$  as input and outputs  $O_{hyp}$  [ $2 \times 160 \times 160 \times 160$ ]. The input of  $M_{sub}$  is defined as  $I_{sub} = I * O_{sub}$ .

While  $O_{hyp}$  is a 2-channel array representing the hypothalamus and its background,  $O_{sub}$  [ $13 \times 160 \times 160 \times 160$ ], the output of  $M_{sub}$ , is a 13-channel array encompassing the subregions, right and left fornices and background.

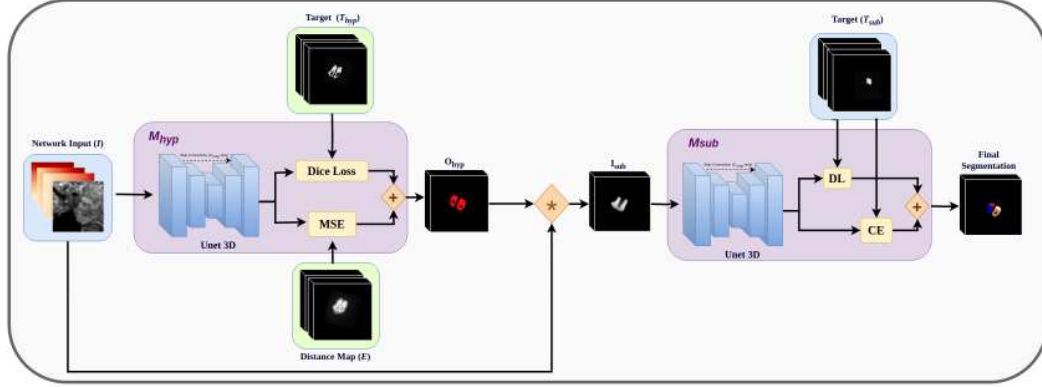


Figure 13.4 – Training Flowchart: There are two training blocks, one focused on the entire hypothalamus and another specialized in subregion segmentation. The training of the two blocks is done subsequently. We first trained the whole structure segmentation model( $M_{hyp}$ ), and later, the model for the subregions segmentation( $M_{sub}$ ). However, the output of  $M_{hyp}$  is used to assist the input creation of  $M_{sub}$  during training.

**Loss function and training details** The loss function applied to  $M_{hyp}$  (Eq. 13.1) is a combination of Dice Loss ( $DL$ ) and Mean Square Error ( $MSE$ ), while the loss function applied to  $M_{sub}$  (Eq. 13.2) combines  $DL$  and Cross Entropy ( $CE$ ):

$$L_{hyp} = \alpha * DL(T, T_{pred}) + (1 - \alpha) * MSE(E, E_{pred}) \quad (13.1)$$

$$L_{sub} = \alpha * DL(T, T_{pred}) + (1 - \alpha) * CE(T, T_{pred}) \quad (13.2)$$

with  $E$  being the Euclidian Distance map,  $T_{pred}$  being the network output and  $T$  the target. For both models, we used Adam optimizer with a learning rate of  $5 \times 10^{-5}$ , a batch size of 32, and value of  $\alpha$  as 0.3. As stop criteria, we simply trained

$M_{hyp}$  for approximately 40000 training steps and did not use any validation set. However, on  $M_{sub}$ , we used 35 images from FSM (5 acquisitions from different MRI sequences from 7 distinct subjects) as validation set (Table 13.1). We set an early stop criteria based on the DC of the validation set. For this, we defined the stopping criteria as  $\delta_{min} = 0.001$ . The network trained for approximately 28000 steps and stopped.

Both 3D U-Net modules are composed of an encoder of 5 levels with 24, 48, 96, 192, and 384 feature maps. Each convolutional block is composed of three layers: group normalization, convolution, and activation function (ReLU).

### 13.2.3 Inference and Post processing

The first step of the inference (Figure 13.5) is preprocessing, in which we find the MNI coordinates ( $C_{inf}$ ) of the input image, by using a fast deep learning algorithm, EasyReg (IGLESIAS, 2023). For  $M_{hyp}$ , the final input  $A_{inf}$  is the concatenation of the cropped inference image ( $CrI_{inf}$ ) and  $C_{inf}$ . The input for  $M_{sub}$ ,  $A_{inf\_sub}$ , is formed by the product of  $A_{inf}$ , the output of  $M_{hyp}$ ,  $O_{hyp\_inf}$ , and the ventral diencephalon (VDC) label, which is derived from the whole brain segmentation produced by EasyReg (IGLESIAS, 2023). The inclusion of the ventral-DC label is justified as it helps reduce false positives within the anterior subregion. The post-processing phase comprises two sequential steps: the rescaling of the final segmentation to match the voxel size of  $I_{inf}$ , and the exclusion of voxels that belong to the third ventricle by using the whole brain segmentation obtained from EasyReg (IGLESIAS, 2023).

### 13.2.4 Statistical Analysis

The statistical analysis was done using the AVD and DC combined with Wilcoxon signed-rank tests to assess the statistical significance of differences. We also compared the ability of H-SynEx and competing methods to find statistical differences in the volume of hypothalamus subregions of controls and patients (AD and bvFTD). For this, we used Wilcoxon rank-sum test to assess the significant

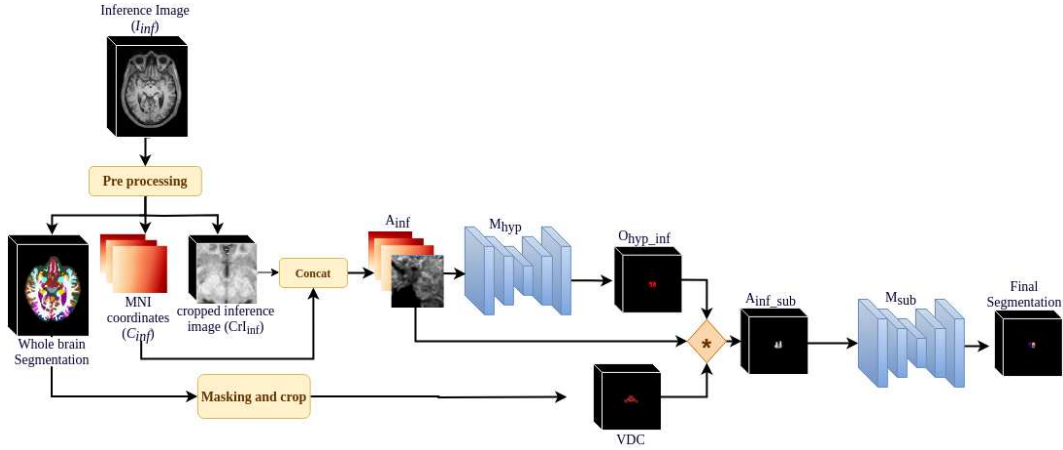


Figure 13.5 – Inference flowchart. The inference image  $I_{inf}$  goes through a pre-processing step to find the input array  $A_{inf}$ .  $A_{inf}$  is then applied to the whole structure segmentation model ( $M_{hyp}$ ). Finally, using VDC (see text),  $A_{inf}$  and the output of  $M_{hyp}$  ( $O_{hyp\_inf}$ ), we create the input for the subregion segmentation model ( $M_{sub}$ ) and find the final subregions segmentation.

difference in medians between groups and the area under the receiving operating characteristic curve (AUROC) as a non-parametric version of effect sizes between groups.

Finally, we used the DeLong test to compare AUROCs across methods operating on the same sample. All statistical tests were conducted with a confidence level of 95% ( $p - value < 0.05$ )

### 13.3 Experiments

As previously stated, our main hypothesis is that the use of synthetic images derived from ultra-high resolution *ex vivo* MRI would reduce label noise and lead us to a model capable of generalizing across various MRI sequences and resolutions (Hypothesis *b,f,g* and *h*). Consequently, our experiments were structured to assess the method's applicability under diverse conditions (Table 13.2). Initially, we conducted an inter-rater analysis to establish quantitative metric baselines.



Subsequently, we validated the method’s performance across different modalities using the FSM dataset (encompassing T1w, T2w, PD, FA, and qT1 sequences) and the IXI dataset (consisting of T1w, T2w, and PD sequences). Following this, we compared H-SynEx with other methods from the literature across multiple datasets (MiLI, MiLI-OASIS, MiLI-IXI, FSM), including one comprising both patient and control subjects (MiLI). We then conducted an application analysis in group studies involving patients with AD and bvFTD (using ADNI and NIFD datasets, respectively). Finally, after confirming the method’s usability with images of different sequences, we assessed its performance across various resolutions using FLAIR images from ADNI with a slicing of  $5mm$ .

Table 13.2 – Summary of conducted experiments

Experiment	Objective	Testing set		
		Dataset	Number of Acquisitions per MRI Sequence	MRI Sequence
Inter-Rater Metrics	To establish a baseline for evaluation metrics	FSM	10	T1
Direct comparison with manual segmentation on different sequences	To assess whether the method is capable to segment on different MRI sequences	FSM	22	T1w, T2w, PD, FA, qT1
		IXI	19	T1w,T2w,PD
Comparing against state-of-the-art methods	Comparing H-SynEx against other state-of-the-art available methods using only T1w images	MiLI	55	T1
		MiLI-OASIS	23	
		MiLI-IXI	19	
		FSM	22	
Application in Group Studies	Assess the method usability in group studies	ADNI	843	T1
		NIFD	185	
Resilience to large slice spacing	To assess usability on diverse MRI resolutions	ADNI	843	T1w, FLAIR

### 13.3.1 Inter-rater metrics

One of the primary challenges in analyzing the results of our experiment is that each dataset used in testing has a distinct manual segmentation protocol (Table 13.1), none of which aligns with the one employed in training H-SynEx due to the difference between *in vivo* and *ex vivo* brain morphology (Section 12.2.2). Therefore, our initial experiment aims to establish a baseline by comparing inter-rater metrics (AVD and DC) using distinct segmentation protocols performed on T1w images. We compare manual segmentations in 10 FSM images delineated by

two different raters: the first uses the FSM protocol (Table 13.1) while the second employs the protocol used during the label maps construction (Section 12.2.2).

### 13.3.2 Direct comparison with manual segmentation on different sequences

In this experiment, we aim to evaluate the ability of H-SynEx to properly segment the subregions of the hypothalamus in different MRI sequences. We employed five different sequences from FSM - T1w, T2w, PD, FA, and quantitative T1 (qT1)- and three from IXI -T1w, T2w, and PD. As the openly available methods exclusively operate on T1w images, a quantitative comparison of their metrics with H-SynEx was not possible in this experiment.

### 13.3.3 Comparing against other state-of-the-art methods

To compare H-SynEx with other state-of-the-art models (BILLOT *et al.*, 2020; RODRIGUES *et al.*, 2022; GREVE *et al.*, 2021), we used T1w images from MiLI and FSM datasets and analyzed the whole hypothalamus segmentation. It is worth noting that the MiLI segmentation protocol does not include the mammillary bodies. Therefore, for this dataset, we excluded the posterior subregion from the results before computing the metrics. Similarly, HypAST does not segment the posterior subregion, therefore we excluded it from FSM in this case, before running the metrics.

### 13.3.4 Application to group studies

In the literature, we can find some studies that point to hypothalamic atrophy in both AD and bvFTD patients (BOCCHETTA *et al.*, 2015; TAO *et al.*, 2021). Therefore, to evaluate the group studies, we compared the hypothalamic subregion volumes of patients and control groups from ADNI (AD subjects) and NIFD (bvFTD subjects). We normalized the volumes by dividing them by the total intracranial volume (TIV) provided by SynthSeg (BILLOT *et al.*, 2023). For

comparative purposes, we conducted the analysis using Billot *et al.* and compared with H-SynEx through DeLong test (DELONG *et al.*, 1988).

### 13.3.5 Resilience to large slice spacing

In this experiment, we applied H-SynEx on FLAIR images from the ADNI dataset acquired with a slice spacing (and thickness) of  $5mm$  in the axial plane. Here, we want to evaluate our method’s capability to identify hypothalamic atrophy with larger spacings, which are common in clinical MRI. Since no other method in the literature works with FLAIR images, we just compared the results from H-SynEx applied on FLAIR and T1w images.

## 13.4 Results

Here we present the final results for each of the presented experiments.

### 13.4.1 Inter-rater metrics

When analyzing inter-rater metrics, we can see that for all subregions the DC is lower or equal to 0.66 (Table 13.3). These low values are expected due to the different segmentation protocols used and the small size of the hypothalamus subregions.

Table 13.3 – Inter-rater metrics (median) for 10 subjects from FSM

Subregion \ Metric	DC	AVD
Anterior	0.63	0.41
Tuberal	0.66	0.43
Posterior	0.66	0.38

### 13.4.2 Direct comparison with manual segmentation on different sequences

Analyzing H-SynEx metrics on different sequences, we can see that the method presents a better performance on T1w images (Figure 13.6). Yet, it is capable of segmenting the hypothalamus and its subregions in all the proposed MRI sequences, as can be seen in Figure 13.7.

### 13.4.3 Comparing against other state-of-the-art methods

Given that Billot *et al.* (BILLOT *et al.*, 2020), the only available method for subregion segmentation works only on T1w images, we compared its results with H-SynEx on 22 T1w images from FSM (Table 13.4). Finally, to compare H-SynEx with ScLimbic (GREVE *et al.*, 2021) and Rodrigues *et al.* (RODRIGUES *et al.*, 2022) we used the whole structure (Table 13.5).

Since the datasets have few subjects and we can not assess with high significance that the distribution is Gaussian, the statistical analyses were conducted considering non-parametric distributions.

Table 13.4 – AVD and DC (median) for H-SynEx and Billot *et al.* on different subregions for FSM dataset. †indicates statistical significance on a two-sided Wilcoxon rank-sum test using Bonferroni correction for  $p < 0.05$

	Model		
	Subregion	H-SynEx	Billot <i>et al.</i>
AVD	Anterior	<b>0.54</b> †	1.32
	Tuberal	<b>0.49</b> †	0.66
	Posterior	<b>0.33</b> †	0.52
DICE	Anterior	<b>0.53</b> †	0.33
	Tuberal	<b>0.59</b>	<b>0.58</b>
	H-Posterior	<b>0.67</b> †	0.55

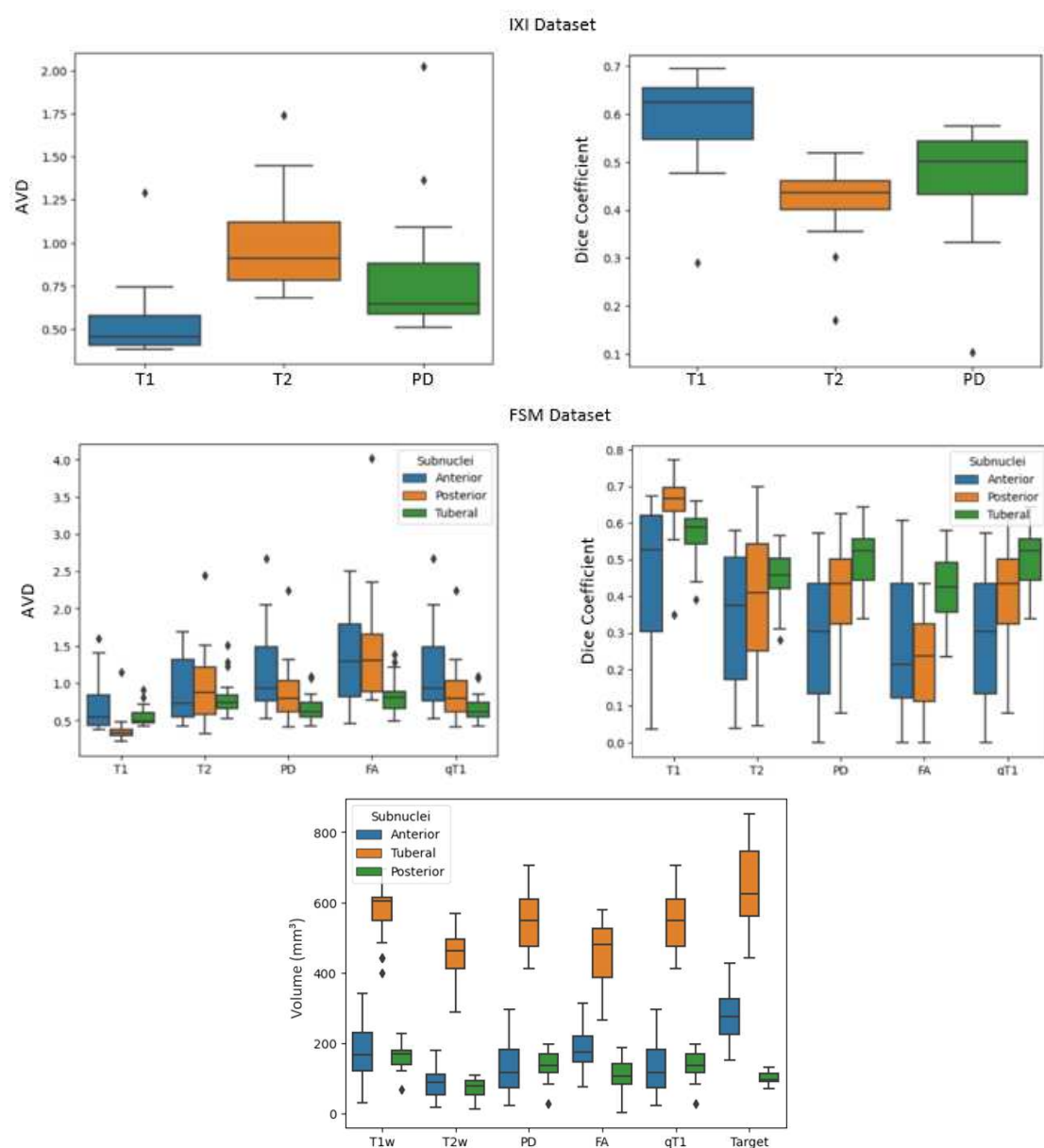


Figure 13.6 – Top row: IXI dataset, which only presents the segmentation of the whole structure (excluding the mammillary bodies). Middle row: FSM dataset, that contains the segmentation of the hypothalamus and its subregions. Bottom row: H-SynEx and manual segmentation (target) volumes for FSM dataset

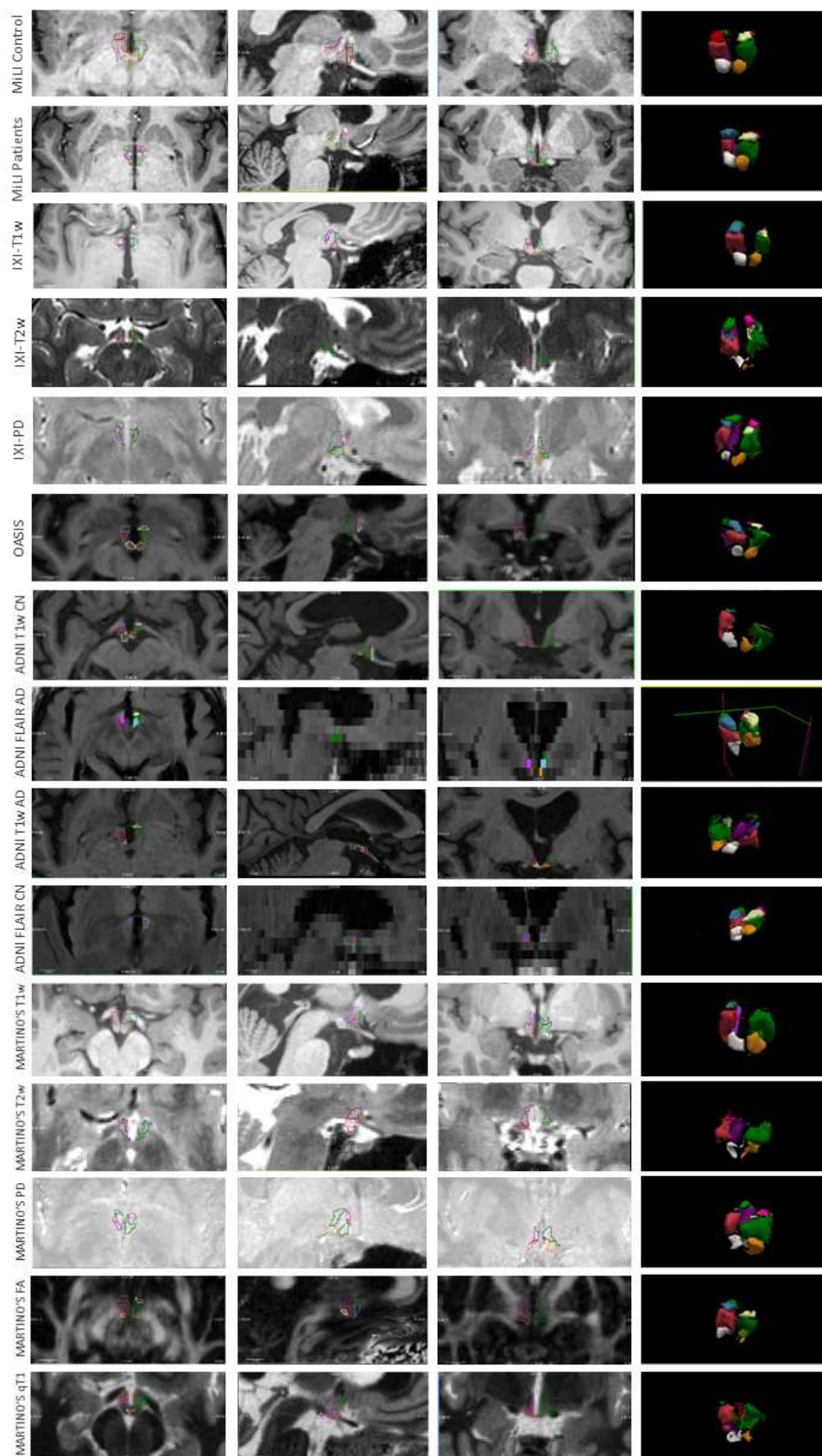


Figure 13.7 – Qualitative results in different datasets, sequences, and resolutions for H-SynEx. Other methods, when applied to sequences different from T1w, return no results

Table 13.5 – AVD and DC (median) for H-SynEx, ScLimbic (GREVE *et al.*, 2021) and Billot *et al.* (BILLOT *et al.*, 2020) on different datasets (MiLI, IXI, OASIS, and FSM) for the entire hypothalamus (except MB). The symbols indicate statistical significance on a two-sided Wilcoxon rank-sum test using Bonferroni correction for  $p < 0.05$ : (\*) Billot *vs* H-SynEx; (†) ScLimbic *vs* H-SynEx; (‡) Billot *vs* ScLimbic. Since ScLimbic was trained using the FSM dataset, we did not consider these results. Similarly, since HypAST was trained using data from MiLI, IXI and the same segmentation protocol as OASIS, we did not consider these results

	Dataset Model	MiLI	IXI	OASIS	FSM
AVD	Billot	0.46	0.61*‡	<b>0.47</b>	<b>0.40</b>
	HypAST	-	-	-	<b>0.41</b>
	ScLimbic	<b>0.39</b> †‡	<b>0.44</b>	<b>0.49</b>	-
	H-SynEx	0.45	<b>0.45</b>	<b>0.5</b>	<b>0.43</b>
DICE	Billot	0.66*	0.6	<b>0.65</b> *‡	<b>0.68</b>
	HypAST	-	-	-	<b>0.69</b>
	ScLimbic	<b>0.67</b> †‡	<b>0.64</b> †‡	0.59	-
	H-SynEx	0.63	0.62	0.58	<b>0.65</b>

#### 13.4.4 Application to group studies

Observing the applicability of the methods on group studies, H-SynEx achieved statistical significance ( $p < 0.05$ ) in the Wilcoxon rank-sum test in all hypothalamic subregions when comparing AD vs. controls, while Billot *et al.* was unable to detect differences in the tuberal-inferior region (Table 13.6). Additionally, in some cases, we observed a higher AUROC in H-SynEx, along with a  $p$ -value  $< 0.5$  for the DeLong test, indicating the ability of H-SynEx to better discern differences between the two groups in this dataset. Regarding NIFD, the results were similar for both models, except for the tuberal-inferior region.

#### 13.4.5 Resilience to large slice spacing

Finally, we analyzed H-SynEx segmentations on 5mm spacing FLAIR images from the same subjects from the ADNI dataset used in Experiment 4. When analyzing the volumes, H-SynEx returns statistically significant results (Ta-

ble 13.6) when comparing patient and control volumes normalized by TIV in all subregions, except for the posterior subregion.

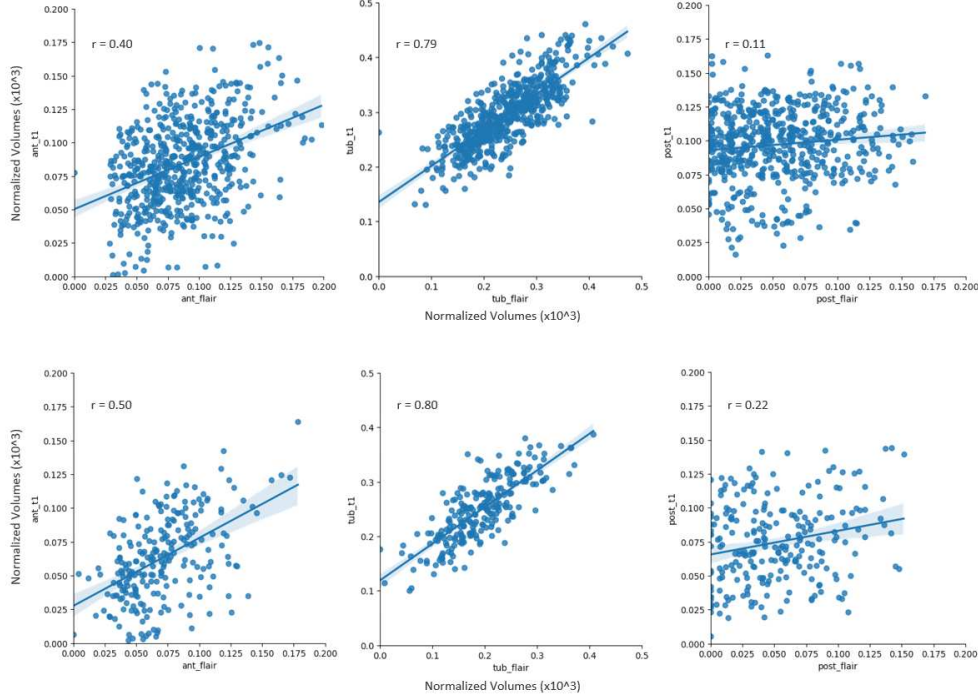


Figure 13.8 – Normalized volume correlation for FLAIRs *vs* T1w (ADNI Dataset) using H-SynEx segmentation. Up: Control subjects; Down: AD patients. We can see that besides the posterior subregion, we can find a positive correlation between FLAIR and T1w normalized volumes.

## 13.5 Discussion

In this part of the project, we addressed Hypotheses *b* and *f* by using ultra-high resolution *ex vivo* images to create more accurate segmentations, compared to traditional *in vivo* image segmentations. Additionally, using only 10 *ex vivo* images, we developed H-SynEx, a method capable of working across different MRI sequences and resolutions, thereby confirming Hypotheses *g* and *h*.

While other studies have utilized synthetic images (BILLOT *et al.*, 2023) and *ex vivo* data (IGLESIAS *et al.*, 2015) for developing segmentation meth-



Table 13.6 – AUROC Values for patients vs. controls (hypothalamus volume) for H-SynEx and Billot methods in ADNI and NIFD datasets. For ADNI dataset, we also analyze our method when applied to FLAIR images with spacing of  $5mm$ . Stars indicate the level of statistical significance (two-sided Wilcoxon rank-sum test) between both cohorts (\*  $p < 0.05$ , \*\*  $p < 0.01$ ).  $^\dagger$  indicates statistical significance on the DeLong test ( $p < 0.05$ ) between H-SynEx and Billot methods.  $^\ddagger$  indicates statistical significance on the DeLong test ( $p < 0.05$ ) between H-SynEx applied on T1w and H-SynEx applied on FLAIRS.

Dataset		ADNI			NIFD	
Model		H-SynEx FLAIR	H-SynEx T1w	Billot T1w	H-SynEx T1w	Billot T1w
Subregion						
Whole		0.66** $^\dagger$	0.74**	0.65** $^\dagger$	0.79**	0.74**
a-sHyp		0.60** $^\dagger$	0.69**	0.72**	0.76**	0.75**
a-iHyp		0.60**	0.64**	0.55* $^\dagger$	0.72**	0.62**
supTub		0.68** $^\dagger$	0.60**	0.67** $^\dagger$	0.76**	0.76**
infTub		0.67** $^\dagger$	0.73**	0.52 $^\dagger$	0.74**	0.59* $^\dagger$
postHyp		0.52 $^\ddagger$	0.72**	0.70**	0.7**	0.73**

ods applied to *in vivo* MRI, to the best of our knowledge, H-SynEx is the first to merge both techniques. Through this integration, we developed a method capable of effectively segmenting small structures, such as hypothalamus subregions, across various MRI sequences and resolutions, including FLAIRs with a spacing of  $5mm$ .

Typically, when evaluating how well a developed segmentation method generalizes, we compare it to others found in existing literature. To do this, it's common to use a dataset that none of the methods have seen during training. However, when these methods use training sets with different segmentation protocols, this difference can introduce bias, favoring the method trained under the same protocol as the test images. By using ex vivo images to construct the training set, the segmentation protocol used in training H-SynEx became different from any other in vivo image set. Consequently, the main challenge in analyzing the results lies in the difference between the training and test protocols. Focusing on that, we compared the manual segmentation of two raters who employed distinct protocols on 10 T1w images from the FSM dataset and found inter-rater DC values lower or equal to 0.66 and AVD higher or equal to 0.38. We use these values as a baseline

for analyzing the metrics in the subsequent experiments.

In Experiment 2, we analyzed H-SynEx usability across different MRI sequences. We could assess that T1w images presented the best results. However, despite the lower DC and higher AVD values for the other sequences, it is important to emphasize that the manual segmentations of the hypothalamus subregions in both FSM and IXI were done in T1w images, not being influenced by the different contrasts of other sequences. Also, despite FSM images for each subject have already been registered, the same is not true for IXI. Hence, the manual segmentations were registered to be used on the different sequences acquisitions of the same subject. Both registration and the use of a different sequence for manual segmentation may compromise the final results. Finally, we could notice a high variability on both metrics, which may be explained by the small size of the hypothalamus. This hypothesis is reassured by comparing the volumes delineated by H-SynEx and manual segmentation in the FSM dataset (Figure 13.6). We can see that both the posterior and anterior subregions, which show greater variability in the DC and AVD, are relatively smaller than the tuberal subregion. Furthermore, the variability in volumes across sequences and subregions appears to be less pronounced than the variability in the metrics. For instance, for the anterior subregions we can see a large variability in the DC, which is less pronounced in both AVD and the volumetric analysis. This may imply that the small size of the anterior subregion may be interfering in the final DC values. The same analysis is valid for the posterior region.

When comparing H-SynEx with other state-of-the-art methods, we can see that H-SynEx outperforms Billot *et al.* in almost every metric for subregion segmentation (Table 13.4). Here, it is important to highlight that despite DC values seeming to be low at first glance, they are not far from the values observed in the inter-rater analysis. The same is true for AVD. Observing AVD and DC for the whole structure (Table 13.5), H-SynEx outperforms Billot *et al.* and returns similar results to ScLimbic (GREVE *et al.*, 2021) on the former, despite not achieving the best performance on the latter. When comparing with HypAST, we can also find similar results on AVD and DC for FSM, but it is important to notice that since

HypAST metrics were computed without the posterior subregion, this is not a direct comparison. However, when dealing with small structures with complex boundaries, distance metrics such as AVD, are more suitable to compare different methods (TAHA; HANBURY, 2015). Also, it is important to emphasize that all other methods were exclusively trained on *in vivo* T1w images, not having to deal with domain gap. Therefore, even not presenting the best quantitative results in T1w images, H-SynEx presents the best generalization ability across MRI sequences.

When comparing volumes of the hypothalamus from patient and control groups on T1w images, we have confirmed that our method exhibits a statistically significant difference in all subregions in ADNI and NIFD datasets, with AUROCs of 0.74 and 0.79 respectively, and  $p - value < 0.05$  for the Wilcoxon signed-rank test in both cases. Notably, the AUROC values reported to NIFD are higher than those found in ADNI. This behavior is expected since bvFTD patients tend to exhibit more pronounced hypothalamic atrophy than AD patients (10-12% volume loss in AD and 15-20% in bvFTD) (VERCRUYSSSE *et al.*, 2018). Additionally, we determined that H-SynEx results differ statistically from Billot *et al.* for the entire hypothalamus and in most subregions in the ADNI dataset, with a  $p - value < 0.05$  for the DeLong test.

Finally, we analyzed the same subjects from ADNI used in experiment 4, but using FLAIR images with a spacing of  $5mm$ . It is possible to see that, similarly to when analyzing T1w images, the method was able to differentiate between patients and controls in almost all subregions, except for the posterior. This may be explained by the  $5mm$  spacing of the FLAIR images since it makes many images lack the mammillary bodies, or limit it to just one slice of the image. For this reason, the small AUROC values in this subregion are expected. Finally, we plotted the correlation among T1w and FLAIR normalized volumes (Figure 13.8) to investigate whether H-SynEx exhibits consistency among them. The anterior subregion displays a moderate correlation ( $r=0.40$  and  $r=0.50$ , respectively), and tuberal subregions have strong correlations ( $r=0.79$  and  $r=0.80$ , respectively), both for controls and AD subjects. As expected, the posterior correlation is weak in both

cases ( $r=0.11$  and  $r=0.22$ ). These results support the hypothesis that the method can be used in challenging resolutions and still detect differences among groups.

A limitation of the project is the difference in H-SynEx performance between T1w images and other sequences and the high variability of the metrics. However, besides the divergence in the manual segmentation protocols used for training and testing, in both IXI and FSM, we only have one label per subject. Therefore the manual segmentations were not influenced by different contrasts. Also, we could demonstrate that the smallest subregions (anterior and posterior) had the biggest variability, especially in DC, an overlap measure known for being sensitive to small structures (TAHA; HANBURY, 2015).

## 14 Considerations

In the process of **Method Expansion**, we established HELM, a comprehensive collection of label maps generated from ultra-high-resolution *ex vivo* MR images. Not only HELM enabled the development of H-SynEx, a state-of-the-art hypothalamus segmentation technique, but it also offers versatility for extending its application to different brain structures with minimal effort. HELM is openly available at [https://www.nitrc.org/projects/hsynex\\_data/](https://www.nitrc.org/projects/hsynex_data/)

Also, we developed, to the best of our knowledge, the first automated method for hypothalamic subregion segmentation capable of working across different *in vivo* MRI sequences and resolutions without retraining. By producing reliable and consistent segmentations, H-SynEx facilitates the analysis of the hypothalamus in various pre-existing datasets, whether in research or clinical settings. This contributes to an improved understanding of the roles played by the hypothalamus and its subregions in neurodegenerative diseases and other related conditions.

Finally, we will apply H-SynEx to real case study groups, exploring the role of the hypothalamus in three different conditions: Spinocerebellar ataxias, amyotrophic lateral sclerosis, and systemic lupus erythematosus.

# Part IV

## Applications

## 15 Applying H-SynEx in Group Studies

Finally, we applied H-SynEx to real case studies, aiming to better understand the role of the hypothalamus in different conditions. Here, we study:

- Spinocerebellar Ataxias;
- Amyotrophic Lateral Sclerosis; and
- Systemic Lupus Erythematosus.

### 15.1 Materials

All data came from subjects that were scanned at the University of Campinas hospital. The participants underwent an MR imaging scan utilizing a 3T Philips Achieva scanner (Philips, Best, The Netherlands) with standard 8-channel head coils. High-resolution 3D T1w volumetric images of the brain in sagittal orientation were acquired, featuring a voxel matrix of  $240 \times 240 \times 180$ , a voxel size of  $1 \times 1 \times 1 \text{ mm}^3$ , a TR/TE of 7/3.201 ms, and a flip angle of  $8^\circ$ . The data used for these studies received approval from the local ethical committee (CEP/Conep, number 3435027) and underwent complete anonymization. All participants were adequately informed and provided their consent by signing a consent form to participate in the study.

### 15.2 Methods

For each of the described conditions, we follow the same methodology, focusing on volumetry and texture attributes analysis.

### 15.2.1 Image Segmentation

The image processing procedures were executed employing H-SynEx. SynthSeg (BILLOT *et al.*, 2023), an advanced deep learning-based approach specifically designed for whole-brain segmentation, was used to find the total intracranial volume (TIV) of each subject in this study.

### 15.2.2 Image Analysis

The analysis conducted here are based on the hypothalamus subregions volumes and texture features extracted from the hypothalamus histogram.

The MR images are comprised of volumetric elements known as voxels. Each voxel is characterized by a numerical value representing its grayscale intensity. By examining the distribution of voxel values within a region of interest, the texture of that region is analyzed. Despite T1w MR imaging does not provide microscopic-level insights that can be visualized, certain pathologies may induce histological alterations that manifest as changes in grayscale patterns within the T1w MR images of the affected area. Such alterations can be discerned through texture analysis. One can find in the literature numerous studies investigating texture characteristics in MR images across different neurological conditions (CASTELLANO *et al.*, 2004). Here, a statistical texture analysis employing histogram-based parameters was applied.

The volumetric analysis, on the other hand, aims to assess variations in subregion volumes between patient and control groups. For this analysis, all subregion volumes are initially normalized by the TIV of each subject, ensuring a more accurate comparison by accounting for the individual variation in overall brain size.

There is no significant difference in the distributions of age and sex between the control and patient cohorts.



### 15.2.3 Statistical Analysis

All statistical analyses were conducted using Python. The statistical differences were assessed using Mann-Whitney U test with a confidence level of 95%. All tests were corrected for multiple comparisons (Bonferroni-corrected  $p < 0.05$ ). Additionally, the effect sizes were calculated for all findings, utilizing the area under the receiver operating characteristic curve (AUROC).

# 16 Spinocerebellar Ataxias

## 16.1 Description

Spinocerebellar ataxia (SCA) is an autosomal dominant degenerative disorder with remarkable phenotypic and genotypic heterogeneity. This is typically a late onset and very disabling condition. The unifying feature in all SCAs is the presence of slowly progressive cerebellar ataxia either in isolation or combined with other neurological manifestations (KLOCKGETHER *et al.*, 2019). To date, there are more than 40 genetic subtypes of SCAs; these are named sequentially in the chronological order of loci discovery (from SCA1 up to SCA50) (COARELLI *et al.*, 2023). The most common subtypes worldwide are SCA1, 2, 3, and 6 (SCOTT *et al.*, 2020). So far, little is known about hypothalamic involvement in SCAs. There is a single study that reported hypothalamic atrophy as well as positive correlation between atrophy and body mass index in Chinese patients with SCA3 (GUO *et al.*, 2022).

## 16.2 Materials

The study included a total of 135 adult patients with molecular confirmation of SCA (31 SCA1, 12 SCA2, 92 SCA3, and 17 SCA6) (Table 16.1). To match age and sex, two groups of healthy subjects were used as controls for comparison purposes. The first group included 100 subjects, and the second, 17 subjects. None of the control subjects had comorbid neurological and/or psychiatric disorders (Table 16.1).

Table 16.1 – SCA patients and controls demographics . SARA: scale for assessment and rating of ataxia

Group	n (male/female)	Age (mean $\pm$ std)	Mean SARA	Mean time of ataxia (years)
SCA1	<sup>31</sup> (21M/10F)	44.55 $\pm$ 8.63	15 $\pm$ 6.75	8 $\pm$ 6.58
SCA2	<sup>12</sup> (6M/6F)	40.5 $\pm$ 20.22	23.25 $\pm$ 11.38	9.6 $\pm$ 5.95
SCA3	<sup>92</sup> (40M/52F)	48.33 $\pm$ 12.2	13.35 $\pm$ 7.9	13.39 $\pm$ 7.15
SCA6	<sup>17</sup> (10M/7F)	67.88 $\pm$ 7.88	13.94 $\pm$ 6.34	14.13 $\pm$ 6.97
Control (SCA6)	17(8M/9F)	65.35 $\pm$ 3.22	–	–
Control (SCA1, SCA2, SCA3)	100(50M/50F)	47.29 $\pm$ 8.63	–	–

### 16.3 Results and Discussion

The volumetric analysis revealed a significant difference in the volume of the hypothalamus across various subregions in patients with SCA1 and SCA3 when compared with the control group (Table 16.2). Atrophy was evident in the entire anterior region (both superior and inferior) for both ataxias, while the tuberal inferior subregion remained unaffected in all cases. There was no significant difference in the subregion volume between controls and patients with SCA2 and SCA6.

Texture analysis indicated statistical differences in almost all hypothalamic subregions between the control group and patients for SCA1 and SCA3. Significant correlations were observed between volume and texture attributes and the scale for assessment and rating of ataxia (SARA) coefficient for SCA1 and SCA3. No significant correlations were found for SCA2. When analyzing the correlation between volume and texture attributes and the average time from ataxia onset, significant associations were observed for SCA1 and SCA3 (Figure 16.1). This may suggest that alterations in the hypothalamus in these conditions may initiate changes in tissue texture before progressing to atrophy, though further studies are necessary to validate this hypothesis.

Finally, the pattern of abnormalities is distinct across groups: severe

and widespread changes in SCA1 and SCA3, but not in SCA2 or SCA6.

Table 16.2 – Subregions that presented significant difference when comparing patient and control groups (Mann-Whitney test with confidence level of 95%)

Attribute	SCA1	SCA2	SCA3	SCA6
Volume	ant_inf, ant_sup, posterior	–	ant_inf, ant_sup, tub_sup	–
Mean	all subregions	all subregions	all subregions	–
Variance	all subregions	ant_sup, tub_inf, tub_sup, posterior	all subregions	tub_inf, tub_sup, posterior
Skweness	ant_sup, posterior	ant_sup, tub_inf	ant_sup	all subregions
Kurtosis	ant_sup	–	tub_sup, posterior	ant_sup
Entropy	all subregions	ant_inf, tub_inf, tub_sup, posterior	all subregions	tub_inf, tub_sup, posterior

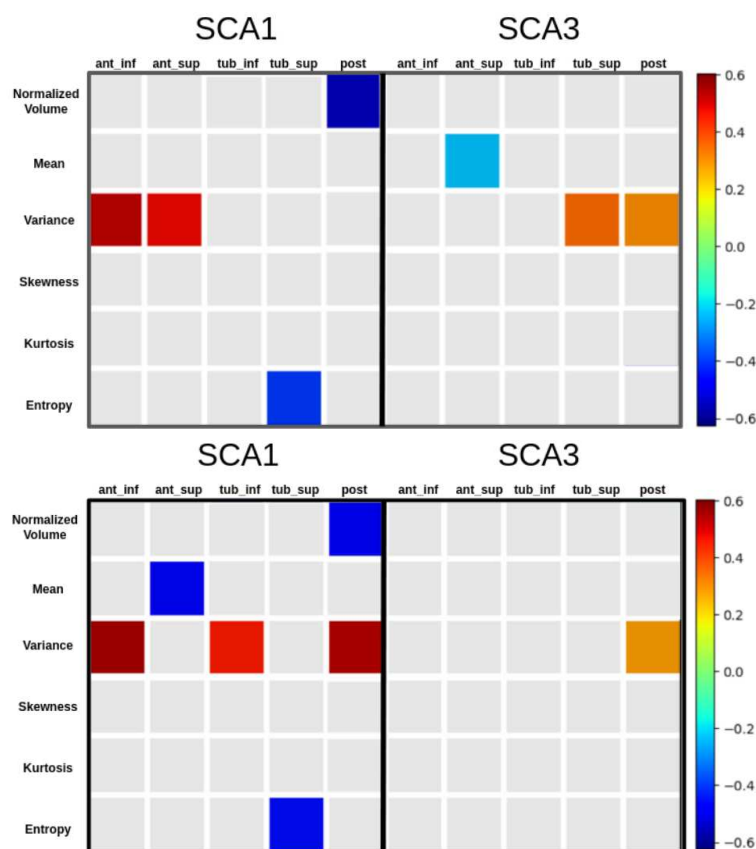


Figure 16.1 – Correlation between SARA scores (top) and disease duration (bottom) and hypothalamic imaging (subregion volumes and texture attributes). Interpretation of Pearson's  $r$  coefficient applied to medicine - analysis for absolute values (Akoglu, 2018): 0:None; 0.1-0.29:poor; 0.3-0.59:Fair; 0.6-0.79: Moderate; 0.8-0.99: Very Strong; 1: Perfect. Gray blocks mean no significant correlation.

# 17 Amyotrophic Lateral Sclerosis

## 17.1 Description

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder that affects the corticospinal tract, brainstem, and anterior horn cells of the spinal cord. The condition typically begins between the ages of 50 and 75, involving both upper and lower motor neurons. About 65% of patients present with limb involvement, 30% with bulbar muscle involvement, and 5% with respiratory muscle involvement. While ALS was initially thought to be confined to motor deficits, it is now increasingly recognized that patients experience significant changes in weight and eating behavior, which both result from and contribute to the underlying neurodegenerative process. These changes include alterations in metabolism, lipid levels, and insulin resistance (AHMED *et al.*, 2021b).

The hypothalamus, the primary brain region regulating energy balance, controls food intake and energy expenditure through homeostatic mechanisms. Given that energy expenditure and eating behavior are profoundly affected in ALS, emerging research suggests that these alterations may be mediated by changes in hypothalamic function. Notably, atrophy of the hypothalamus has been observed in both ALS patients and presymptomatic individuals at genetic risk for the disease (AHMED *et al.*, 2021b; VERCRUYSSSE *et al.*, 2018).

## 17.2 Materials

This study involved 139 subjects, including 83 patients (49 male and 34 female, with an average age of  $56.86 \pm 12.31$  years) and 56 controls (31 male and 25 female, with an average age of  $56.91 \pm 7.67$  years).

## 17.3 Results and Discussion

Upon analyzing the differences in volume and texture attributes between patient and control groups, we discovered significant disparities in both the tuberal inferior and posterior subregions. Conversely, no significant differences in volume or texture were observed in the other subregions. When reviewing the literature, we can find studies indicating hypothalamic atrophy in ALS (GORGES *et al.*, 2017; AHMED *et al.*, 2021b; CHANG *et al.*, 2023); however, few have examined subregions (LIU *et al.*, 2022a), and none have analyzed texture. By assessing both texture and volume and examining the subregions individually, we can gain greater confidence in identifying affected regions. The concurrent alterations in texture and volume within the same regions suggest that the tuberal inferior and posterior regions are indeed more affected in ALS patients.

# 18 Systemic Lupus Erythematosus

## 18.1 Description

Systemic lupus erythematosus (SLE) is an autoimmune disease that can affect people of different genders and ethnicities, despite being more prevalent in black women. The disease can affect different organs such as the skin, central nervous system, and kidneys (PISETSKY, 2008).

Several authors have shown alterations in the Hypothalamic–pituitary–adrenal (HPA) axis in SLE patients when compared with controls (APPENZELLER *et al.*, 2007; MONTERO-LÓPEZ *et al.*, 2017; GOES *et al.*, 2010). Besides, the SLE may affect the central nervous system (CNS), resulting in neuropsychiatric SLE (NSLE) in 15-66% of SLE patients. The NSLE comprehends psychiatric and neurological syndromes of SLE patients (OTA *et al.*, 2022). One of the primary mechanisms is the alteration of cytokine balance in the CNS, which may affect the hypothalamus.

Being a crucial part of the neuroendocrine system, some studies have defended the need to further study the hypothalamus and HPA in autoimmune and chronic rheumatic disease patients.

## 18.2 Materials

The study was divided into two subgroups: Juvenile SLE and Adult SLE. Juvenile SLE encompasses patients who received an SLE diagnosis before the age of 18 years, while adult SLE encompasses patients who received an SLE diagnosis with or after the age of 18 years.

The juvenile SLE group included a total of 126 subjects, 92 patients (12 male and 80 female with average age of  $17.71 \pm 3.84$ ), and 34 controls (6 male and 28 female with average age of  $18.76 \pm 4.56$ ). The adult SLE group included a



total of 325 subjects, 207 patients (12 male and 195 female with an average age of  $41.16 \pm 11.88$ ), and 97 controls (6 male and 81 female with an average age of  $39.80 \pm 12.14$ ) (Table 18.1).

Table 18.1 – Adult and Juvenile lupus patients and control demographics

Group	n (male/female)	Age (mean $\pm$ std)
Adult Lupus	207 (12M/195F)	$41.16 \pm 11.88$
Juvenile Lupus	92 (12M/80F)	$17.71 \pm 3.84$
Adult Lupus Control	118 (38M/80F)	$39.80 \pm 12.14$
Juvenile Lupus Control	34 (6M/28F)	$18.76 \pm 4.56$

### 18.3 Results and Discussion

When analyzing juvenile SLE, no significant differences are found in the hypothalamus when considered as a whole. However, when examining the subregions individually, a difference in the volume of the posterior subregion becomes apparent. Texture analysis also reveals differences in all attributes for the posterior subregion and in the variance of the tuberal superior subregion.

In adult SLE, differences in hypothalamus volumetry can be detected without subregion analysis. However, significant differences are observed in the volumes of the anterior superior, tuberal inferior, and posterior subregions. Texture analysis indicates significant differences in all attributes for the anterior superior, tuberal superior, and posterior subregions.

These results suggest that the hypothalamus may be more affected by the disease in cases of late-onset symptoms, as seen in adult lupus (Figure 18.1), and reinforce the importance of analyzing the subregions individually.

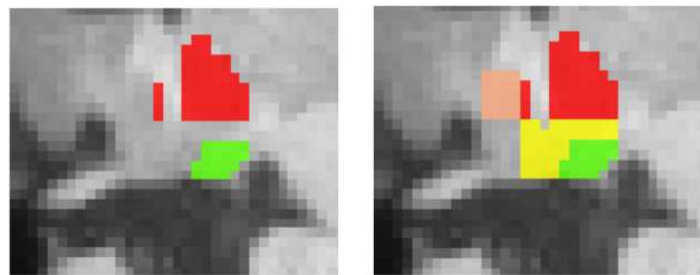


Figure 18.1 – Subregions where we could find volume and/or texture differences between patients and controls. Left: Juvenile Lupus; Right: Adult Lupus. Pink: anterior superior, red: tuberal superior, yellow: tuberal inferior, and green: posterior

## Part V

### Conclusion

## 19 Conclusions

The hypothalamus is a small brain structure that plays an essential role in maintaining body homeostasis and is linked to neurodegenerative diseases. Despite its importance, there was a gap in the literature regarding hypothalamus segmentation, and no automated method existed at the beginning of this project. In this thesis, we present two datasets focused on the hypothalamus (MiLI and HELM) along with two segmentation methods (HypAST and H-SynEx).

MiLI is the first public dataset for hypothalamus segmentation, comprising four distinct image groups: three public datasets (OASIS, IXI and CC-359) and a newly created dataset specifically for this benchmark. MiLI includes over 1300 acquisitions, with images from controls and ataxia patients. MiLI contains both manual and automated labels of the hypothalamus. Using MiLI, we developed HypAST, a tool that embeds Correct2Seg, a method consisting of two CNN blocks that can handle noise from different labels and segment small structures such as the hypothalamus. HypAST provides access to hypothalamus volumes and histogram texture attributes. Despite its good performance, returning better dice scores compared to state-of-the-art methods, HypAST is limited to T1w 1mm isotropic images. This limitation led us to the second part of this thesis.

In the second part, we introduced HELM and H-SynEx. HELM is a dataset of label maps derived from 10 *ex vivo* images. The label maps were created using both manual and automated segmentation: manual for delineating the hypothalamus and its subregions, and automated for segmenting the rest of the brain. H-SynEx is a segmentation method trained with synthetic images derived from HELM, capable of segmenting the hypothalamus and its subregions.

Finally, we presented an analysis of three study groups consisting of patients with ataxias (SCA1, SCA2, SCA3, and SCA6), lupus (juvenile and adult), and ALS. In all cases, we found significant differences in some hypothalamus subregions between patient and control groups.

To the best of our knowledge, we have presented the first automated method for hypothalamic segmentation of the literature (RODRIGUES *et al.*, 2020) and the first method capable of segmenting the hypothalamus and its subregions across different in vivo MRI sequences and resolutions without retraining. By producing reliable and consistent segmentations, H-SynEx facilitates the analysis of the hypothalamus in various preexisting datasets, whether in research or clinical settings. This contributes to an improved understanding of the roles played by the hypothalamus and its individual subregions in neurodegenerative diseases and other related conditions.

## 20 Next Studies

Despite our work have helped fill a gap at the hypothalamus literature, we can list some works that can be done to improve even more the studies on this structure:

- **Expand HELM and H-SynEx to Different Structures:** Current methods such as SynthSeg have demonstrated the ability to generalize across different MRI sequences, focusing primarily on the entire brain (BILLOT *et al.*, 2023). However, leveraging the ultra-high resolution of *ex vivo* images could enhance the segmentation of smaller structures such as the nucleus accumbens, amygdala, and pituitary gland.
- **Analyze the Application in Neurodegenerative Diseases Across Other Sequences:** Recent studies indicate that obese patients exhibit higher T2w relaxation times compared to control groups (SANDE-LEE *et al.*, 2020). A subsequent step for this project is to compare T2w images from control groups with those from lupus patients, who often experience an increase in body mass index due to medication intake (CHAIAMNUAY *et al.*, 2007).
- **Merge the Texture and Volume Analysis from HypAST with H-SynEx:** Integrating a graphical user interface with H-SynEx can streamline the application of this method, making it more user-friendly and enhancing its functionality by combining texture and volume analysis.

## 21 Publications, Collaborations, Awards

### 21.1 Publications on the Hypothalamus

- Rodrigues, Livia, et al. "High-resolution segmentations of the hypothalamus and its subregions for training of segmentation models". *Scientific Data* (2024) (*Accepted*)
- Rodrigues, Livia, et al. "H-SynEx: Using synthetic images and ultra-high resolution ex vivo MRI for hypothalamus subregion segmentation." *arXiv preprint arXiv:2401.17104* (2024). *Submitted to AI in Medicine*
- Rodrigues, Livia, et al. "A benchmark for hypothalamus segmentation on T1w MR images." *NeuroImage* 264 (2022): 119741.
- Rodrigues, Livia, et al. "Hypothalamus fully automatic segmentation from MR images using a U-Net based architecture." 15th International Symposium on Medical Information Processing and Analysis. Vol. 11330. SPIE, 2020.
- Rodrigues, Livia, et al. "Hypothalamus semi-automatic segmentation from MR images using Convolutional Neural Networks. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM) 27th Annual Meeting Exhibition*.
- Rodrigues, Livia, et al. "Hypothalamus Segmentation From MRI Using Convolutional Neural Networks." In: 6th Brainn Congress, 2019, Campinas. *Journal of Epilepsy and Clinical Neurophysiology*, 2019.

### 21.2 Publications on collaborative projects

- \*Carmo, D., \*Pinheiro, G., \*Rodrigues, L., Abreu, T., Lotufo, R., Rittner, L. (2023). Automated computed tomography and magnetic resonance imag-

ing segmentation using deep learning: a beginner's guide. arXiv preprint arXiv:2304.05901.

- Beauferris, Y., Teuwen, J., Karkalousos, D., Moriakov, N., Caan, M., Yiasemis, G., Rodrigues, L., ... Souza, R. (2022). Multi-coil MRI reconstruction challenge—assessing brain MRI reconstruction models and their generalizability to varying coil configurations. *Frontiers in Neuroscience*, 16, 919186.
- Carmo, D., Campiotti, I., Fantini, I., Rodrigues, L., Rittner, L., Lotufo, R. (2021, December). Multitasking segmentation of lung and COVID-19 findings in CT scans using modified EfficientDet, UNet and MobileNetV3 models. In *17th International Symposium on Medical Information Processing and Analysis* (Vol. 12088, pp. 65-74). SPIE.
- Carmo, D., Campiotti, I., Rodrigues, L., Fantini, I., Pinheiro, G., Moraes, D., ... Lotufo, R. (2021). Rapidly deploying a COVID-19 decision support system in one of the largest Brazilian hospitals. *Health Informatics Journal*, 27(3), 14604582211033017.
- Phellan, R., Rodrigues, L., Pinheiro, G. R., Soto, A. Q., Rodrigues, I. D., Rittner, L., ... Bento, M. (2019). Automatic detection of age-and sex-related differences in human brain morphology. In *Proceedings of International Society for Magnetic Resonance in Medicine (ISMRM) 27th Annual Meeting Exhibition*.

## 21.3 Colaborations

The following collaborations were essential for the development of this project:

- **LNI:** The development of MiLI and HypAST was in collaboration with the Laboratory of NeuroImage from the School of Medical Sciences (Faculdade de Ciências Médicas - FCM), Unicamp, which provided us with images and manual segmentations of public datasets.



- **University of Calgary:** Currently, MiLI is hosted on the Calgary-Campinas dataset, in a collaboration with the University of Calgary.
- **Athinoula A. Martinos Center for Biomedical Imaging:** The collaboration with the Martinos Center (Massachusetts General Hospital - MGH/Harvard) made possible the development of H-SynEx and HELM.

## 21.4 Awards

During the PhD, the candidate received the following awards:

- **SIPAIM Society Award:** Best Paper on SIPAIM Conference, 2021
- **MICCAI:** MICCAI 2020 Startups Village Pitch Competition - COVID Solution, 2020
- **IdeaGov (Brazilian Government Challenge):** "How the use of artificial intelligence algorithms can assist radiologists in diagnosing Covid-19 through Computed Tomography images and chest x-rays?", São Paulo-Alberta-BrainHack 2020
- **São Paulo-Alberta-BrainHack:** First place in the "Age classification based on magnetic resonance images of the brain" challenge

# Bibliography

- AGARWAL. *Complete Architectural Details of all EfficientNet Models*. <<https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142>>. Accessed: 2024-05-12. Citado 2 vezes nas páginas 9 and 49.
- AHMED, R. M.; STEYN, F.; DUPUIS, L. Hypothalamus and weight loss in amyotrophic lateral sclerosis. *Handbook of Clinical Neurology*, Elsevier, v. 180, p. 327–338, 2021. Citado 2 vezes nas páginas 24 and 32.
- AHMED, R. M.; STEYN, F.; DUPUIS, L. Hypothalamus and weight loss in amyotrophic lateral sclerosis. *Handbook of Clinical Neurology*, Elsevier, v. 180, p. 327–338, 2021. Citado 2 vezes nas páginas 110 and 111.
- ALFARO-ALMAGRO, F.; JENKINSON, M.; BANGERTER, N. K.; ANDERSSON, J. L.; GRIFFANTI, L.; DOUAUD, G.; SOTIROPOULOS, S. N.; JBABDI, S.; HERNANDEZ-FERNANDEZ, M.; VALLEE, E. *et al.* Image processing and quality control for the first 10,000 brain imaging datasets from uk biobank. *Neuroimage*, Elsevier, v. 166, p. 400–424, 2018. Citado na página 30.
- APPENZELLER, S.; BONILHA, L.; RIO, P. A.; LI, L. M.; COSTALLAT, L. T. L.; CENDES, F. Longitudinal analysis of gray and white matter loss in patients with systemic lupus erythematosus. *Neuroimage*, Elsevier, v. 34, n. 2, p. 694–701, 2007. Citado na página 112.
- ATALOGLU, D.; DIMOU, A. *et al.* Fast and precise hippocampus segmentation through deep convolutional neural network ensembles and transfer learning. *Neuroinformatics*, Springer, v. 17, n. 4, p. 563–582, 2019. Citado na página 27.
- BAHETI, B.; INNANI, S. *et al.* Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In: *Proceedings of the IEEE/CVF CVPR Workshops*. [S.l.: s.n.], 2020. p. 358–359. Citado na página 49.
- BARTLETT, D. M.; REYES, A. *et al.* Investigating the relationships between hypothalamic volume and measures of circadian rhythm and habitual sleep in premanifest huntington’s disease. *Neurobiology of sleep and circadian rhythms*, Elsevier, v. 6, p. 1–8, 2019. Citado 3 vezes nas páginas 24, 25, and 32.
- BEBER, B. C.; CHAVES, M. L. Evaluation of patients with behavioral and cognitive complaints: Misdiagnosis in frontotemporal dementia and Alzheimer’s

disease. *Dementia & neuropsychologia*, SciELO Brasil, v. 7, p. 60–65, 2013. Citado na página 32.

BILLOT, B.; BOCCHETTA, M. *et al.* Automated segmentation of the hypothalamus and associated subunits in brain MRI. *NeuroImage*, Elsevier, v. 223, p. 117287, 2020. Citado 11 vezes nas páginas 15, 30, 31, 45, 46, 61, 62, 78, 90, 92, and 95.

BILLOT, B.; GREVE, D. N.; PUONTI, O.; THIELSCHER, A.; LEEMPUT, K. V.; FISCHL, B.; DALCA, A. V.; IGLESIAS, J. E. *et al.* Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical image analysis*, Elsevier, v. 86, p. 102789, 2023. Citado 7 vezes nas páginas 28, 34, 84, 90, 96, 104, and 118.

BOCCHETTA, M.; GORDON, E. *et al.* Detailed volumetric analysis of the hypothalamus in behavioral variant frontotemporal dementia. *Journal of Neurology*, Springer, v. 262, p. 2635–2642, 2015. Citado 6 vezes nas páginas 24, 26, 73, 74, 83, and 90.

BORTSOVA, G.; DUBOST, F.; HOGEWEG, L.; KATRAMADOS, I.; BRUIJNE, M. D. Semi-supervised medical image segmentation via learning consistency under transformations. In: SPRINGER. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*. [S.l.], 2019. p. 810–818. Citado na página 27.

BRETELER, M. M.; STÖCKER, T.; PRACHT, E.; BRENNER, D.; STIRNBERG, R. Ic-p-165: Mri in the rhineland study: a novel protocol for population neuroimaging. *Alzheimer's & Dementia*, Wiley Online Library, v. 10, p. P92–P92, 2014. Citado na página 30.

CABEZAS, M.; OLIVER, A.; LLADÓ, X.; FREIXENET, J.; CUADRA, M. B. A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine*, Elsevier, v. 104, n. 3, p. e158–e177, 2011. Citado na página 27.

CARDOZO-HERNÁNDEZ, A. L. de C.; REZENDE, T. J. R.; JR, M. C. F. Hereditary spastic paraplegia type 11 (SPG11) is associated with obesity and hypothalamic damage. *Journal of the Neurological Sciences*, Elsevier, v. 416, p. 116982, 2020. Citado na página 47.

CARMO, D.; SILVA, B. o. Hippocampus segmentation on epilepsy and Alzheimer's disease studies with multiple convolutional neural networks. *Heliyon*, Elsevier, v. 7, n. 2, p. e06226, 2021. Citado na página 28.

CASTELLANO, G.; BONILHA, L.; LI, L.; CENDES, F. Texture analysis of medical images. *Clinical radiology*, Elsevier, v. 59, n. 12, p. 1061–1069, 2004. Citado na página 104.

CHAIAMNUAY, S.; BERTOLI, A. M.; FERNÁNDEZ, M.; APTE, M.; VILÁ, L. M.; REVEILLE, J. D.; ALARCÓN, G. S.; GROUP, L. S. *et al.* The impact of increased body mass index on systemic lupus erythematosus: Data from lumina, a multiethnic cohort. *JCR: Journal of Clinical Rheumatology*, LWW, v. 13, n. 3, p. 128–133, 2007. Citado na página 118.

CHANG, J.; SHAW, T. B.; HOLDOM, C. J.; MCCOMBE, P. A.; HENDERSON, R. D.; FRIPP, J.; BARTH, M.; GUO, C. C.; NGO, S. T.; STEYN, F. J. *et al.* Lower hypothalamic volume with lower body mass index is associated with shorter survival in patients with amyotrophic lateral sclerosis. *European Journal of Neurology*, Wiley Online Library, v. 30, n. 1, p. 57–68, 2023. Citado na página 111.

CHEN, H.; DOU, Q. *et al.* VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, Elsevier, v. 170, p. 446–455, 2018. Citado na página 27.

CHO, H.; SHUKLA, S. Role of edaravone as a treatment option for patients with amyotrophic lateral sclerosis. *Pharmaceuticals*, MDPI, v. 14, n. 1, p. 29, 2020. Citado na página 33.

ÇIÇEK, Ö.; ABDULKADIR, A.; LIENKAMP, S. S.; BROX, T.; RONNEBERGER, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In: SPRINGER. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19*. [S.l.], 2016. p. 424–432. Citado na página 85.

COARELLI, G.; COUTELIER, M.; DURR, A. Autosomal dominant cerebellar ataxias: new genes and progress towards treatments. *The Lancet Neurology*, Elsevier, v. 22, n. 8, p. 735–749, 2023. Citado na página 106.

COSTANTINI, I.; MORGAN, L.; YANG, J.; BALBASTRE, Y.; VARADARAJAN, D.; PESCE, L.; SCARDIGLI, M.; MAZZAMUTO, G.; GAVRYUSEV, V.; CASTELLI, F. M. *et al.* A cellular resolution atlas of broca’s area. *Science Advances*, American Association for the Advancement of Science, v. 9, n. 41, p. eadg3844, 2023. Citado na página 70.

DELONG, E. R.; DELONG, D. M.; CLARKE-PEARSON, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves:

a nonparametric approach. *Biometrics*, JSTOR, p. 837–845, 1988. Citado na página 91.

DO, N.; JOO, S. *et al.* Knee bone tumor segmentation from radiographs using Seg-Unet with dice loss. In: *25th (IW-FCV), Gangneung, South Korea*. [S.l.: s.n.], 2019. Citado na página 28.

ESTRADA, S.; KÜGLER, D.; BAHRAMI, E.; XU, P.; MOUSA, D.; BRETELER, M.; AZIZ, N. A.; REUTER, M. Fastsurfer-hypvinn: Automated sub-segmentation of the hypothalamus and adjacent structures on high-resolutioal brain mri. *arXiv preprint arXiv:2308.12736*, 2023. Citado 3 vezes nas páginas 30, 31, and 78.

FAYJIE, A. R.; DUTTA, R.; KASHYAP, P.; KUMAR, U. R.; VANDEWALLE, P. Semi-supervised adversarial few-shot learning for medical image segmentation. 2022. Citado na página 27.

GABERY, S.; GEORGIOU-KARISTIANIS, N. *et al.* Volumetric analysis of the hypothalamus in huntington disease using 3T MRI: The image-hd study. *PloS one*, Public Library of Science, v. 10, n. 2, p. e0117593, 2015. Citado 4 vezes nas páginas 24, 25, 32, and 47.

GOES, M. C. van der; BOSSEMA, E. R.; HARTKAMP, A.; GODAERT, G. L.; JACOBS, J. W.; KRUIZE, A. A.; DERKSEN, R. H.; BIJLSMA, J. W.; GEENEN, R. Cortisol during the day in patients with systemic lupus erythematosus or primary sjögren's syndrome. *The Journal of rheumatology*, The Journal of Rheumatology, 2010. Citado na página 112.

GOLDSTEIN, J. M.; SEIDMAN, L. o. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biological psychiatry*, Elsevier, v. 61, n. 8, p. 935–945, 2007. Citado na página 25.

GORGES, M.; VERCRUYSSSE, P. *et al.* Hypothalamic atrophy is related to body mass index and age at onset in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, BMJ Publishing Group Ltd, v. 88, n. 12, p. 1033–1041, 2017. Citado 3 vezes nas páginas 24, 47, and 111.

GREVE, D. N.; BILLOT, B.; CORDERO, D.; HOOPES, A.; HOFFMANN, M.; DALCA, A. V.; FISCHL, B.; IGLESIAS, J. E.; AUGUSTINACK, J. C. A deep learning toolbox for automatic segmentation of subcortical limbic structures from mri images. *Neuroimage*, Elsevier, v. 244, p. 118610, 2021. Citado 9 vezes nas páginas 15, 30, 31, 81, 83, 90, 92, 95, and 98.

GUO, J.; JIANG, Z.; BISWAL, B. B.; ZHOU, B.; XIE, D.; GAO, Q.; SHENG, W.; CHEN, H.; ZHANG, Y.; FAN, Y. *et al.* Hypothalamic atrophy, expanded

cag repeat, and low body mass index in spinocerebellar ataxia type 3. *Movement Disorders*, Wiley Online Library, v. 37, n. 7, p. 1541–1546, 2022. Citado na página 106.

GUTIERREZ, M.; GARCIA, M.; RODRIGUEZ, J.; RIVERO, S.; JACOBELLI, S. Hypothalamic-pituitary-adrenal axis function and prolactin secretion in systemic lupus erythematosus. *Lupus*, Sage Publications Sage CA: Thousand Oaks, CA, v. 7, n. 6, p. 404–408, 1998. Citado na página 24.

HARDIMAN, O.; BERG, L. H. V. D.; KIERNAN, M. C. Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nature reviews neurology*, Nature Publishing Group, v. 7, n. 11, p. 639–649, 2011. Citado na página 32.

HENSCHL, L.; CONJETI, S.; ESTRADA, S.; DIERS, K.; FISCHL, B.; REUTER, M. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*, Elsevier, v. 219, p. 117012, 2020. Citado na página 30.

HENSCHL, L.; KÜGLER, D.; REUTER, M. Fastsurfervinn: Building resolution-independence into deep learning segmentation methods—a solution for highres brain mri. *NeuroImage*, Elsevier, v. 251, p. 118933, 2022. Citado na página 30.

IGLESIAS, J. E. Easyreg: A ready-to-use deep learning tool for symmetric affine and nonlinear brain mri registration. 2023. Citado 3 vezes nas páginas 28, 79, and 87.

IGLESIAS, J. E.; AUGUSTINACK, J. C.; NGUYEN, K.; PLAYER, C. M.; PLAYER, A.; WRIGHT, M.; ROY, N.; FROSCH, M. P.; MCKEE, A. C.; WALD, L. L. *et al.* A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution mri: application to adaptive segmentation of in vivo mri. *Neuroimage*, Elsevier, v. 115, p. 117–137, 2015. Citado 3 vezes nas páginas 28, 71, and 96.

IGLESIAS, J. E.; BILLOT, B.; BALBASTRE, Y.; TABARI, A.; CONKLIN, J.; GONZÁLEZ, R. G.; ALEXANDER, D. C.; GOLLAND, P.; EDLOW, B. L.; FISCHL, B. *et al.* Joint super-resolution and synthesis of 1 mm isotropic mp-rage volumes from clinical mri exams with scans of different orientation, resolution and contrast. *Neuroimage*, Elsevier, v. 237, p. 118206, 2021. Citado na página 79.

IGLESIAS, J. E.; INSAUSTI, R.; LERMA-USABIAGA, G.; BOCCHETTA, M.; LEEMPUT, K. V.; GREVE, D. N.; KOUWE, A. Van der; FISCHL, B.; CABALLERO-GAUDES, C.; PAZ-ALONSO, P. M. *et al.* A probabilistic atlas of the human thalamic nuclei combining ex vivo mri and histology. *Neuroimage*, Elsevier, v. 183, p. 314–326, 2018. Citado 2 vezes nas páginas 28 and 71.

IGLESIAS, J. E.; SABUNCU, M. R. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, Elsevier, v. 24, n. 1, p. 205–219, 2015. Citado na página 27.

JOHNEN, A.; BERTOUX, M. Psychological and cognitive markers of behavioral variant frontotemporal dementia—A clinical neuropsychologist’s view on diagnostic criteria and beyond. *Frontiers in neurology*, Frontiers Media SA, v. 10, p. 594, 2019. Citado na página 32.

JUAREZ A. GARCIA, U. o. A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs. In: SPRINGER. *International Workshop on Machine Learning in Medical Imaging*. [S.l.], 2019. p. 583–591. Citado na página 28.

KARIMI, D.; DOU, H. *et al.* Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Medical Image Analysis*, Elsevier, v. 65, p. 101759, 2020. Citado na página 44.

KISS, D. S.; TOTH, I.; JOCSAK, G.; BARANY, Z.; BARTHA, T.; FRENYO, L. V.; HORVATH, T. L.; ZSARNOVSZKY, A. Functional aspects of hypothalamic asymmetry. *Brain Sciences*, MDPI, v. 10, n. 6, p. 389, 2020. Citado na página 71.

KLOCKGETHER, T.; MARIOTTI, C.; PAULSON, H. L. Spinocerebellar ataxia. *Nature reviews Disease primers*, Nature Publishing Group UK London, v. 5, n. 1, p. 24, 2019. Citado na página 106.

KLOMP, A.; KOOLSCHIJN, P. . *et al.* Hypothalamus and pituitary volume in schizophrenia: a structural MRI study. *IJNP*, Cambridge University Press Cambridge, UK, v. 15, n. 2, p. 281–288, 2012. Citado na página 25.

LAMONTAGNE, P. J.; BENZINGER, T. L.; MORRIS, J. C.; KEEFE, S.; HORNBECK, R.; XIONG, C.; GRANT, E.; HASSENSTAB, J.; MOULDER, K.; VLASSENKO, A. *et al.* OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *MedRxiv*, Cold Spring Harbor Laboratory Press, 2019. Citado 3 vezes nas páginas 42, 44, and 82.

LANDMAN, B. A.; HUANG, A. J.; GIFFORD, A.; VIKRAM, D. S.; LIM, I. A. L.; FARRELL, J. A.; BOGOVIC, J. A.; HUA, J.; CHEN, M.; JARSO, S. *et al.* Multi-parametric neuroimaging reproducibility: a 3-T resource study. *NeuroImage*, Elsevier, v. 54, n. 4, p. 2854–2866, 2011. Citado na página 44.

LIU, S.; REN, Q.; GONG, G.; SUN, Y.; ZHAO, B.; MA, X.; ZHANG, N.; ZHONG, S.; LIN, Y.; WANG, W. *et al.* Hypothalamic subregion abnormalities

are related to body mass index in patients with sporadic amyotrophic lateral sclerosis. *Journal of Neurology*, Springer, p. 1–9, 2022. Citado na página 111.

LIU, X.; YANG, L.; CHEN, J.; YU, S.; LI, K. Region-to-boundary deep learning model with multi-scale feature fusion for medical image segmentation. *Biomedical Signal Processing and Control*, Elsevier, v. 71, p. 103165, 2022. Citado na página 84.

MAKRIS, N.; SWAAB, D. F.; KOUWE, A. van der; ABBS, B.; BORIEL, D.; HANDA, R. J.; TOBET, S.; GOLDSTEIN, J. M. Volumetric parcellation methodology of the human hypothalamus in neuroimaging: normative data and sex differences. *NeuroImage*, Elsevier, v. 69, p. 1–10, 2013. Citado 2 vezes nas páginas 25 and 26.

MARCUS, D. S.; WANG, T. H.; PARKER, J.; CSERNANSKY, J. G.; MORRIS, J. C.; BUCKNER, R. L. Open access series of imaging studies (OASIS): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ..., v. 19, n. 9, p. 1498–1507, 2007. Citado na página 62.

MATHEWS, M.; ANZAR, S. *et al.* EfficientNet for retinal blood vessel segmentation. In: IEEE. (*ICSPIS*). [S.l.], 2020. p. 1–4. Citado 2 vezes nas páginas 29 and 48.

MENG, X.-L. Reproducibility, Replicability, and Reliability. *Harvard Data Science Review*, v. 2, n. 4, oct 29 2020. <https://hdsr.mitpress.mit.edu/pub/hn51kn68>. Citado na página 56.

MILES, K. A.; GANESHAN, B.; HAYBALL, M. P. Ct texture analysis using the filtration-histogram method: what do the measurements mean? *Cancer Imaging*, BMC, v. 13, n. 3, p. 400, 2013. Citado na página 64.

MINGXING, T.; QUOC, V. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. *CoRR*, abs/1905.11946, 2019. Disponível em: <<http://arxiv.org/abs/1905.11946>>. Citado na página 28.

MODI, S.; THAPLOO, D. *et al.* Individual differences in trait anxiety are associated with gray matter alterations in hypothalamus: Preliminary neuroanatomical evidence. *Psychiatry Research: Neuroimaging*, Elsevier, v. 283, p. 45–54, 2019. Citado na página 24.

MONTERO-LÓPEZ, E.; SANTOS-RUIZ, A.; GONZÁLEZ, R.; NAVARRETE-NAVARRETE, N.; ORTEGO-CENTENO, N.; MARTÍNEZ-AUGUSTÍN, O.;



RODRÍGUEZ-BLÁZQUEZ, M.; PERALTA-RAMÍREZ, M. I. Analyses of hair and salivary cortisol for evaluating hypothalamic–pituitary–adrenal axis activation in patients with autoimmune disease. *Stress*, Taylor & Francis, v. 20, n. 6, p. 541–548, 2017. Citado na página 112.

MUELLER, S. G.; WEINER, M. W.; THAL, L. J.; PETERSEN, R. C.; JACK, C.; JAGUST, W.; TROJANOWSKI, J. Q.; TOGA, A. W.; BECKETT, L. The alzheimer’s disease neuroimaging initiative. *Neuroimaging Clinics*, Elsevier, v. 15, n. 4, p. 869–877, 2005. Citado na página 82.

NGUYEN, D.; TRAN, T. *et al.* Skin Lesion Segmentation based on Integrating EfficientNet and Residual block into U-Net Neural Network. In: IEEE. *2020 (GTSD)*. [S.l.], 2020. p. 366–371. Citado na página 29.

OLIVEIRA, A. S. B.; PEREIRA, R. D. B. Amyotrophic lateral sclerosis (ALS): three letters that change the people’s life. for ever. *Arquivos de neuro-psiquiatria*, SciELO Brasil, v. 67, p. 750–782, 2009. Citado na página 33.

OTA, Y.; SRINIVASAN, A.; CAPIZZANO, A. A.; BAPURAJ, J. R.; KIM, J.; KUROKAWA, R.; BABA, A.; MORITANI, T. Central nervous system systemic lupus erythematosus: pathophysiologic, clinical, and imaging features. *Radiographics*, Radiological Society of North America, v. 42, n. 1, p. 212–232, 2022. Citado na página 112.

PARK, G.; HONG, J. *et al.* Corpus callosum segmentation using deep neural networks with prior information from multi-atlas images. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*. [S.l.], 2018. v. 10579, p. 105791B. Citado na página 27.

PHARWAHA, A. P. S.; SINGH, B. Shannon and non-shannon measures of entropy for statistical texture feature extraction in digitized mammograms. In: *Proceedings of the world congress on engineering and computer science*. [S.l.: s.n.], 2009. v. 2, n. 6. Citado na página 64.

PIGUET, O.; PETERSÉN, Å.; LAM, B. Y. K.; GABERY, S.; MURPHY, K.; HODGES, J. R.; HALLIDAY, G. M. Eating and hypothalamus changes in behavioral-variant frontotemporal dementia. *Annals of Neurology*, Wiley Online Library, v. 69, n. 2, p. 312–319, 2011. Citado 2 vezes nas páginas 24 and 32.

PISETSKY, D. S. Systemic lupus erythematosus: epidemiology, pathology, and pathogenesis. *Primer on the Rheumatic Disease*, Springer, p. 246–251, 2008. Citado na página 112.

PIYUSH, R.; RAMAKRISHNAN, S. Analysis of sub-anatomic volume changes in Alzheimer brain using diffusion tensor imaging. In: IEEE. *2014 40th Annual Northeast Bioengineering Conference (NEBEC)*. [S.l.], 2014. p. 1–2. Citado na página 24.

PUONTI, O.; IGLESIAS, J. E.; LEEMPUT, K. V. Fast and sequence-adaptive whole-brain segmentation using parametric bayesian modeling. *NeuroImage*, Elsevier, v. 143, p. 235–249, 2016. Citado 2 vezes nas páginas 72 and 73.

RODRIGUES, L.; BOCCHETTA, M.; PUONTI, O.; GREVE, D.; LONDE, A. C.; FRANÇA, M.; APPENZELLER, S.; IGLESIAS, J. E.; RITTNER, L. *H-SynEx: Using synthetic images and ultra-high resolution ex vivo MRI for hypothalamus subregion segmentation*. 2024. Citado 2 vezes nas páginas 30 and 31.

RODRIGUES, L.; REZENDE, T. *et al.* Hypothalamus fully automatic segmentation from MR images using a U-Net based architecture. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *15th SIPAIM*. [S.l.], 2020. v. 11330, p. 113300J. Citado 6 vezes nas páginas 29, 31, 38, 45, 46, and 117.

RODRIGUES, L.; REZENDE, T. J. R.; WERTHEIMER, G.; SANTOS, Y.; FRANÇA, M.; RITTNER, L. A benchmark for hypothalamus segmentation on t1-weighted mr images. *NeuroImage*, Elsevier, v. 264, p. 119741, 2022. Citado 6 vezes nas páginas 29, 73, 82, 83, 90, and 92.

ROMÁN, K. L.-L.; OCAÑA, M. I. G.; URZELAI, N. L.; BALLESTER, M. Á. G.; OLIVER, I. M. Medical image segmentation using deep learning. *Deep Learning in Healthcare: Paradigms and Applications*, Springer, p. 17–31, 2020. Citado na página 27.

RONNEBERGER, O.; FISCHER, P. *et al.* U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *MICCAI*. [S.l.], 2015. p. 234–241. Citado 2 vezes nas páginas 27 and 28.

SANDE-LEE, S. van de; MELHORN, S. J.; RACHID, B.; RODOVALHO, S.; DE-LIMA-JUNIOR, J. C.; CAMPOS, B. M.; PEDRO, T.; BELTRAMINI, G. C.; CHAIM, E. A.; PAREJA, J. C. *et al.* Radiologic evidence that hypothalamic gliosis is improved after bariatric surgery in obese women with type 2 diabetes. *International Journal of Obesity*, Nature Publishing Group UK London, v. 44, n. 1, p. 178–185, 2020. Citado na página 118.

SAYGIN, Z. M.; KLIEMANN, D.; IGLESIAS, J. E.; KOUWE, A. J. van der; BOYD, E.; REUTER, M.; STEVENS, A.; LEEMPUT, K. V.; MCKEE, A.;

FROSCH, M. P. *et al.* High-resolution magnetic resonance imaging reveals nuclei of the human amygdala: manual segmentation to automatic atlas. *Neuroimage*, Elsevier, v. 155, p. 370–382, 2017. Citado na página 28.

SCHINDLER, S.; SCHÖNKNECHT, P.; SCHMIDT, L.; ANWANDER, A.; STRAUSS, M.; TRAMPEL, R.; BAZIN, P.-L.; MÖLLER, H. E.; HEGERL, U.; TURNER, R. *et al.* Development and evaluation of an algorithm for the computer-assisted segmentation of the human hypothalamus on 7-Tesla magnetic resonance images. *PloS one*, Public Library of Science San Francisco, USA, v. 8, n. 7, p. e66394, 2013. Citado 2 vezes nas páginas 25 and 26.

SCHÖNKNECHT, P.; ANWANDER, A.; PETZOLD, F.; SCHINDLER, S.; KNÖSCHE, T. R.; MÖLLER, H. E.; HEGERL, U.; TURNER, R.; GEYER, S. Diffusion imaging-based subdivision of the human hypothalamus: a magnetic resonance study with clinical implications. *European archives of psychiatry and clinical neuroscience*, Springer, v. 263, p. 497–508, 2013. Citado 2 vezes nas páginas 24 and 25.

SCHUR, E. A.; MELHORN, S. J.; OH, S.-K.; LACY, J. M.; BERKSETH, K. E.; GUYENET, S. J.; SONNEN, J. A.; TYAGI, V.; ROSALYNN, M.; LEON, B. D. *et al.* Radiologic evidence that hypothalamic gliosis is associated with obesity and insulin resistance in humans. *Obesity*, Wiley Online Library, v. 23, n. 11, p. 2142–2148, 2015. Citado na página 24.

SCOTT, S. S. de O.; PEDROSO, J. L.; BARSOTTINI, O. G. P.; FRANÇA-JUNIOR, M. C.; BRAGA-NETO, P. Natural history and epidemiology of the spinocerebellar ataxias: Insights from the first description to nowadays. *Journal of the Neurological Sciences*, Elsevier, v. 417, p. 117082, 2020. Citado na página 106.

SEONG, J.; KANG, J. Y.; SUN, J. S.; KIM, K. W. Hypothalamic inflammation and obesity: a mechanistic review. *Archives of pharmacal research*, Springer, v. 42, p. 383–392, 2019. Citado na página 24.

SOUZA, R.; LUCENA, O. *et al.* An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, Elsevier, v. 170, p. 482–494, 2018. Citado 3 vezes nas páginas 42, 43, and 45.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, BioMed Central, v. 15, n. 1, p. 1–28, 2015. Citado 4 vezes nas páginas 52, 53, 99, and 100.

TAO, A.; MYSLINSKI, Z.; PAN, Y.; IADECOLA, C.; DYKE, J.; CHIANG, G.; ISHII, M. *Hypothalamic Atrophy in Alzheimer's Disease (1819)*. [S.l.]: AAN Enterprises, 2021. Citado na página 90.

TERLEVIC, R.; ISOLA, M. *et al.* Decreased hypothalamus volumes in generalized anxiety disorder but not in panic disorder. *Journal of affective disorders*, Elsevier, v. 146, n. 3, p. 390–394, 2013. Citado na página 25.

THAMBAWITA, V.; SALEHI, P.; SHESHKAL, S. A.; HICKS, S. A.; HAMMER, H. L.; PARASA, S.; LANGE, T. d.; HALVORSEN, P.; RIEGLER, M. A. Singan-seg: Synthetic training data generation for medical image segmentation. *PloS one*, Public Library of Science San Francisco, CA USA, v. 17, n. 5, p. e0267976, 2022. Citado na página 28.

TOGNIN, S.; RAMBALDELLI, G. *et al.* Enlarged hypothalamic volumes in schizophrenia. *Psychiatry Research: Neuroimaging*, Elsevier, v. 204, n. 2-3, p. 75–81, 2012. Citado 2 vezes nas páginas 25 and 26.

VERCRUYSSSE, P.; VIEAU, D. *et al.* Hypothalamic alterations in neurodegenerative diseases and their relation to abnormal energy metabolism. *Front. Mol. Neurosci.*, Frontiers, v. 11, p. 2, 2018. Citado 3 vezes nas páginas 24, 99, and 110.

WACHINGER, C.; REUTER, M. *et al.* DeepNAT: Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*, Elsevier, v. 170, p. 434–445, 2018. Citado na página 27.

WANG, R.; LEI, T.; CUI, R.; ZHANG, B.; MENG, H.; NANDI, A. K. Medical image segmentation using deep learning: A survey. *IET Image Processing*, Wiley Online Library, v. 16, n. 5, p. 1243–1267, 2022. Citado na página 27.

WARFIELD, S.; ZOU, K. *et al.* Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging*, IEEE, v. 23, n. 7, p. 903–921, 2004. Citado na página 46.

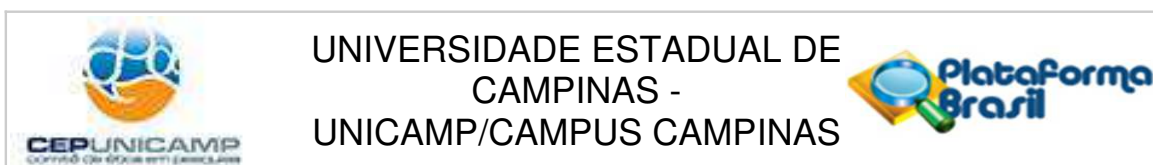
WOLFE, F. H.; AUZIAS, G. *et al.* Focal atrophy of the hypothalamus associated with third ventricle enlargement in autism spectrum disorder. *Neuroreport*, LWW, v. 26, n. 17, p. 1017–1022, 2015. Citado na página 24.

WOLFF, J.; SCHINDLER, S. *et al.* A semi-automated algorithm for hypothalamus volumetry in 3 tesla magnetic resonance images. *Psychiatry Research: Neuroimaging*, Elsevier, v. 277, p. 45–51, 2018. Citado 3 vezes nas páginas 24, 26, and 29.

WOLNY, A.; CERRONE, L.; VIJAYAN, A.; TOFANELLI, R.; BARRO, A. V.; LOUVEAUX, M.; WENZL, C.; STRAUSS, S.; WILSON-SÁNCHEZ, D.; LYMBOURIDOU, R.; STEIGLEDER, S. S.; PAPE, C.; BAILONI, A.; DURAN-NEBREDÁ, S.; BASSEL, G. W.; LOHMANN, J. U.; TSIANTIS, M.; HAMPRECHT, F. A.; SCHNEITZ, K.; MAIZEL, A.; KRESHUK, A. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *eLife*, eLife Sciences Publications, Ltd, v. 9, p. e57613, jul 2020. ISSN 2050-084X. Disponível em: <<https://doi.org/10.7554/eLife.57613>>. Citado na página 85.

YU, L.; WANG, S. *et al.* Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: SPRINGER. *MICCAI*. [S.l.], 2019. p. 605–613. Citado na página 50.

## ANNEX A – Ethical Approvals

**PARECER CONSUBSTANCIADO DO CEP****DADOS DA EMENDA**

**Título da Pesquisa:** Métodos quantitativos em ressonância magnética

**Pesquisador:** LETICIA RITTNER

**Área Temática:**

**Versão:** 2

**CAAE:** 15505319.4.0000.5404

**Instituição Proponente:** Faculdade de Engenharia Elétrica e de Computação

**Patrocinador Principal:** Financiamento Próprio

**DADOS DO PARECER**

**Número do Parecer:** 6.773.799

**Apresentação do Projeto:**

Trata-se da apresentação de uma Emenda ao projeto CAAE 15505319.4.0000.5404

**Objetivo da Pesquisa:**

- Atualização dos integrantes na equipe de pesquisa (inclusões e exclusões)
- Atualização do cronograma de execução da pesquisa

**Avaliação dos Riscos e Benefícios:**

Os riscos e benefícios não foram alterados em relação ao projeto original

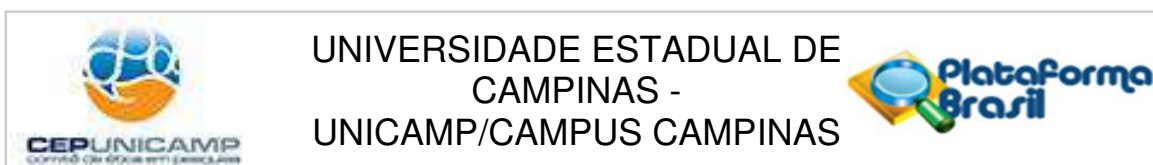
**Comentários e Considerações sobre a Pesquisa:**

Data da aprovação do projeto por este CEP: 03/07/2019 (parecer número 3.435.027, em 'PB\_PARECER\_CONSUBSTANCIADO\_CEP\_3435027.pdf', de 03/07/2019 13:36:25)

Através da Emenda apresentada, o pesquisador informa que (item 'Justificativa da Emenda')

- "Com o grande avanço das técnicas de Inteligência Artificial (AI) no processamento e análise de imagens médicas, muitos dos métodos tradicionais se mostraram obsoletos e com resultados inferiores do ponto de vista de acurácia, confiabilidade e reprodutibilidade"
- "Para que possamos desenvolver novas técnicas baseadas em aprendizado profundo (Deep Learning) que possam de fato servir para uso na pesquisa médica, necessitamos de mais tempo"

**Endereço:** Rua Tessália Vieira de Camargo, 126, 1º andar do Prédio I da Faculdade de Ciências Médicas  
**Bairro:** Barão Geraldo **CEP:** 13.083-887  
**UF:** SP **Município:** CAMPINAS  
**Telefone:** (19)3521-8936 **Fax:** (19)3521-7187 **E-mail:** cep@unicamp.br



Continuação do Parecer: 6.773.799

Dessa forma, um novo cronograma de execução foi proporcionado (item 'Outras informações, justificativas ou considerações a critério do pesquisador')

- "Organização dos dados 01/08/2019 - 30/09/2019 (2 meses)"
- Implementação de técnicas de análise de imagens existentes - 01/10/2019 - 31/03/2020 (6 meses)"
- "Desenvolvimento de técnicas baseadas em aprendizado profundo - 01/04/2020 - 30/09/2024 (4 anos e 6 meses)"
- "Avaliação e análise dos resultados - 01/10/2024 - 30/09/2025 (1 ano) "
- "Publicação de resultados por artigos"

Equipe de Pesquisa (cf. arquivo 'Carta\_emenda\_CEP\_2024.docx', de 25/03/2024 19:59:07)

- Bruno Rangel Balbino dos Santos
- Caio Pinheiro Santana
- Diedre Santos do Carmo
- Gabriel Santos Martins Dias
- Gustavo Retuci Pinheiro
- Jean Antônio Ribeiro
- Joany do Socorro Santa Rosa Rodrigues
- Leticia Rittner
- Livia Maria de Aguiar Rodrigues
- Mateus Oliveira da Silva
- Roberto de Alencar Lotufo
- Suellen Sena da Silva
- Wesna Simone Bulla de Araújo
- Yanna Jannetti Ognibine

#### **Considerações sobre os Termos de apresentação obrigatória:**

Para avaliação desta Emenda foi analisado o documento intitulado 'PB\_INFORMAÇÕES\_BÁSICAS\_2097529\_E1.pdf' (de 26/03/2024 08:01:50)

#### **Conclusões ou Pendências e Lista de Inadequações:**

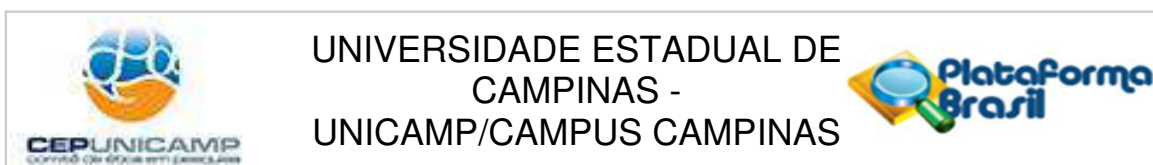
Emenda aprovada

#### **Considerações Finais a critério do CEP:**

- O participante da pesquisa deve receber uma via do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado (quando aplicável).

**Endereço:** Rua Tessália Vieira de Camargo, 126, 1º andar do Prédio I da Faculdade de Ciências Médicas  
**Bairro:** Barão Geraldo **CEP:** 13.083-887  
**UF:** SP **Município:** CAMPINAS  
**Telefone:** (19)3521-8936 **Fax:** (19)3521-7187 **E-mail:** cep@unicamp.br

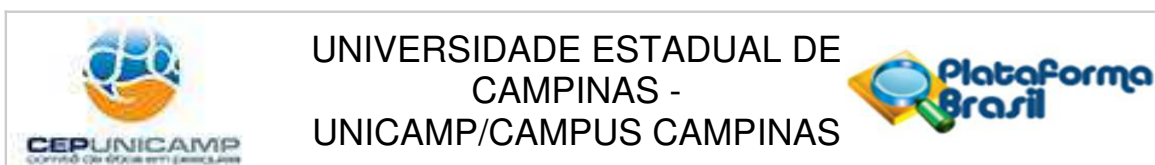




Continuação do Parecer: 6.773.799

- O participante da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (quando aplicável).
- O pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado. Se o pesquisador considerar a descontinuação do estudo, esta deve ser justificada e somente ser realizada após análise das razões da descontinuidade pelo CEP que o aprovou. O pesquisador deve aguardar o parecer do CEP quanto à descontinuação, exceto quando perceber risco ou dano não previsto ao participante ou quando constatar a superioridade de uma estratégia diagnóstica ou terapêutica oferecida a um dos grupos da pesquisa, isto é, somente em caso de necessidade de ação imediata com intuito de proteger os participantes.
- O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo. É papel do pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária – ANVISA – junto com seu posicionamento.
- Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas e aguardando a aprovação do CEP para continuidade da pesquisa. Em caso de projetos do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma, junto com o parecer aprovatório do CEP, para serem juntadas ao protocolo inicial.
- Relatórios parciais e final devem ser apresentados ao CEP, inicialmente seis meses após a data deste parecer de aprovação e ao término do estudo.
- Lembramos que segundo a Resolução 466/2012, item XI.2 letra e, “cabe ao pesquisador apresentar dados solicitados pelo CEP ou pela CONEP a qualquer momento”.
- O pesquisador deve manter os dados da pesquisa em arquivo, físico ou digital, sob sua guarda

**Endereço:** Rua Tessália Vieira de Camargo, 126, 1º andar do Prédio I da Faculdade de Ciências Médicas  
**Bairro:** Barão Geraldo **CEP:** 13.083-887  
**UF:** SP **Município:** CAMPINAS  
**Telefone:** (19)3521-8936 **Fax:** (19)3521-7187 **E-mail:** cep@unicamp.br



Continuação do Parecer: 6.773.799

e responsabilidade, por um período de 5 anos após o término da pesquisa.

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_2097529_E1.pdf	26/03/2024 08:01:50		Aceito
Outros	2019_ProjetoCEP_LeticiaRittner_destaque.docx	25/03/2024 20:00:47	Yanna Jannetti Ognibine	Aceito
Outros	2019_ProjetoCEP_LeticiaRittner_Sem_destaque.docx	25/03/2024 19:59:47	Yanna Jannetti Ognibine	Aceito
Outros	Carta_emenda_CEP_2024.docx	25/03/2024 19:59:07	Yanna Jannetti Ognibine	Aceito
Outros	relatorio.pdf	11/06/2019 13:46:30	LETICIA RITTNER	Aceito
Folha de Rosto	FolhaDeRosto_assinada.pdf	10/06/2019 10:23:29	LETICIA RITTNER	Aceito
Outros	TCUD_LeticiaRittner.pdf	10/06/2019 09:55:44	LETICIA RITTNER	Aceito
Projeto Detalhado / Brochura Investigador	2019_ProjetoCEP_LeticiaRittner.pdf	10/06/2019 09:50:50	LETICIA RITTNER	Aceito
Outros	Termo_de_Autorizacao_LETICIA_RIBEIRO.pdf	09/06/2019 22:18:07	LETICIA RITTNER	Aceito
Outros	Leticia0001.pdf	09/06/2019 22:16:12	LETICIA RITTNER	Aceito
Outros	Autorizacao_uso_Leticia.pdf	09/06/2019 22:15:46	LETICIA RITTNER	Aceito
Outros	Template_autorizacao_uso.pdf	09/06/2019 22:12:51	LETICIA RITTNER	Aceito
Outros	Balthazar_autorizacao_uso.pdf	08/06/2019 11:19:22	LETICIA RITTNER	Aceito

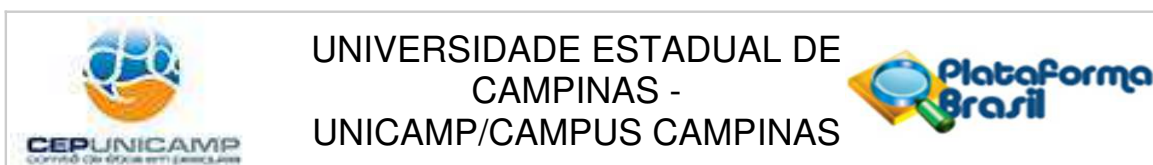
**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

**Endereço:** Rua Tessália Vieira de Camargo, 126, 1º andar do Prédio I da Faculdade de Ciências Médicas  
**Bairro:** Barão Geraldo **CEP:** 13.083-887  
**UF:** SP **Município:** CAMPINAS  
**Telefone:** (19)3521-8936 **Fax:** (19)3521-7187 **E-mail:** cep@unicamp.br



Continuação do Parecer: 6.773.799

CAMPINAS, 18 de Abril de 2024

---

**Assinado por:**  
**Renata Maria dos Santos Celeghini**  
**(Coordenador(a))**

**Endereço:** Rua Tessália Vieira de Camargo, 126, 1º andar do Prédio I da Faculdade de Ciências Médicas  
**Bairro:** Barão Geraldo **CEP:** 13.083-887  
**UF:** SP **Município:** CAMPINAS  
**Telefone:** (19)3521-8936 **Fax:** (19)3521-7187 **E-mail:** cep@unicamp.br