



Universidade Estadual de Campinas
Faculdade de Ciências Médicas

Diogo Hideki Shiraishi

**Segmentação Cerebelar Automatizada e Otimizada
Utilizando Deep Learning**

***Automated and Optimized Cerebellar Segmentation
Using Deep Learning***

CAMPINAS
2024

Diogo Hideki Shiraishi

**Segmentação Cerebelar Automatizada e Otimizada Utilizando
Deep Learning**

***Automated and Optimized Cerebellar Segmentation Using Deep
Learning***

Dissertação apresentada à Faculdade de Ciências Médicas da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciências, na área de concentração em Neurologia.

Dissertation submitted to the Faculty of Medical Sciences of the University of Campinas in partial fulfillment of the requirements for the degree of Master of Science, with a concentration in Neurology.

Orientador: Dr. Thiago Junqueira Ribeiro de Rezende

Este exemplar corresponde à versão final da Dissertação defendida por Diogo Hideki Shiraishi e orientada pelo Dr. Thiago Junqueira Ribeiro de Rezende.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca da Faculdade de Ciências Médicas
Maristella Soares dos Santos - CRB 8/8402

Sh65s Shiraishi, Diogo Hideki, 1993-
Segmentação cerebelar automatizada e otimizada utilizando deep learning /
Diogo Hideki Shiraishi. – Campinas, SP : [s.n.], 2024.

Orientador: Thiago Junqueira Ribeiro de Rezende.
Dissertação (mestrado) – Universidade Estadual de Campinas (UNICAMP),
Faculdade de Ciências Médicas.

1. Cerebelo. 2. Imageamento por ressonância magnética. 3. Aprendizado profundo. 4. Ataxias espinocerebelares. 5. Ataxia de Friedreich. I. Rezende, Thiago Junqueira Ribeiro de, 1988-. II. Universidade Estadual de Campinas (UNICAMP). Faculdade de Ciências Médicas. III. Título.

Informações Complementares

Título em outro idioma: Automated and optimized cerebellar segmentation using deep learning

Palavras-chave em inglês:

Cerebellum

Magnetic resonance imaging

Deep learning

Spinocerebellar ataxias

Friedreich's ataxia

Área de concentração: Neurologia

Titulação: Mestre em Ciências

Banca examinadora:

Thiago Junqueira Ribeiro de Rezende [Orientador]

Edson Amaro Junior

Andre Monteiro Paschoal

Data de defesa: 22-08-2024

Programa de Pós-Graduação: Ciências Médicas

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

- ORCID do autor: <https://orcid.org/0000-0002-6826-573X>

- Currículo Lattes do autor: <http://lattes.cnpq.br/3543415467158569>

COMISSÃO EXAMINADORA DA DEFESA DE MESTRADO

DIOGO HIDEKI SHIRAISHI

Orientador: Dr. Thiago Junqueira Ribeiro de Rezende

Membros titulares

1. Dr. Thiago Junqueira Ribeiro de Rezende

2. Prof. Dr. Edson Amaro Junior

3. Prof. Dr. Andre Monteiro Paschoal

Programa de Pós-Graduação em Ciências Médicas da Faculdade de Ciências Médicas da Universidade Estadual de Campinas.

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Data da defesa: 22 de agosto de 2024

Agradecimentos

A Deus por me guiar e proporcionar as oportunidades.

Aos meus pais, Marli e Júlio (*in memoriam*), pelo incentivo, pela paciência e pelos conselhos.

Ao meu orientador Dr. Thiago Rezende pela confiança, pelo convite, pela parceria nesses anos todos e pelos ensinamentos que aos poucos vou absorvendo. É sua referência que incentiva e inspira.

Aos meus familiares, em especial meus padrinhos e primos, que, além de serem profissionais e pessoas excelentes, foram um suporte essencial no decorrer da jornada.

Ao Ian Harding e à Susmita Saha, nossos parceiros de pesquisa, que propuseram, discutiram e viabilizaram este projeto em parceria conosco.

Novamente ao meu orientador pelo suporte em algumas situações complicadas e à Mariana pela indicação do Dr. Leonardo.

Aos colegas do Laboratório de Neuroimagem que compartilharam ensinamentos, dúvidas e que, de alguma forma, contribuíram com o trabalho.

Aos gestores e colegas do Itaú pelo incentivo e interesse. E à empresa por estruturar uma carreira na qual é possível conciliar as atividades do trabalho com as atividades acadêmicas.

À Friedreich's Ataxia Research Alliance (FARA) pelo apoio financeiro concedido através de *General Research Grant* com vigência de março de 2023 a março de 2025.

Aos pacientes, voluntários e seus familiares.

Resumo

A segmentação do cerebelo e do núcleo denteado (do inglês, *dentate nucleus*, DN) em imagens de ressonância magnética (do inglês, *magnetic resonance imaging*, MRI) é desafiadora devido às suas estruturas complexas e à proximidade com as regiões cerebrais adjacentes. O córtex cerebelar apresenta dobras intrincadas, e o DN baixa relação contraste-ruído por ser uma pequena estrutura localizada internamente na substância branca do cerebelo. As ferramentas tradicionais apresentam desempenho satisfatório em imagens de indivíduos saudáveis, mas menor performance quando aplicadas em imagens provenientes de indivíduos com lesão cerebelar. A fim de aumentar a generalização e acurácia das segmentações automatizadas das estruturas cerebelares, propomos implementar um modelo de aprendizado profundo para segmentação do córtex cerebelar em imagens ponderadas em T1 (do inglês, *T1-weighted*, T1w) e do DN em imagens de mapeamento quantitativo de suscetibilidade magnética (do inglês, *quantitative susceptibility mapping*, QSM), tanto em indivíduos saudáveis quanto em pacientes que apresentam dano cerebelar.

Para a segmentação do cerebelo, utilizamos sequências T1w, incluindo imagens de controles saudáveis e pacientes com ataxia espinocerebelar (do inglês, *spinocerebellar ataxia*, SCA) tipos 1 e 3 e ataxia de Friedreich (FRDA). Neurorradiologistas experientes criaram anotações das imagens (máscaras de segmentação), garantindo a identificação de qualidade das fissuras cerebelares. Treinamos uma arquitetura U-Net 3D modificada e o desempenho foi comparado com os resultados dos modelos ACAPULCO e CERES. Para a segmentação do DN, analisamos imagens QSM de 132 controles saudáveis e 170 pacientes com ataxia cerebelar ou esclerose múltipla de nove conjuntos de dados. A segmentação manual do DN (hilus da substância branca e cinzenta) por anotadores experientes forneceu as máscaras de referência, aliado a um processo robusto de controle de qualidade. Uma abordagem para segmentação automatizada de dois passos foi implementada com aprendizado

profundo, composta por um modelo de localização seguido de segmentação do DN.

Nosso modelo de segmentação do cerebelo atingiu uma métrica Dice de $0,928 \pm 0,030$ no conjunto de teste e superou o ACAPULCO (Dice: $0,909 \pm 0,043$) e o CERES (Dice: $0,918 \pm 0,039$). Em pacientes com maior nível de dano cerebelar, essa diferença se manteve (Dice do modelo: $0,924 \pm 0,033$, ACAPULCO: $0,904 \pm 0,047$, CERES: $0,917 \pm 0,044$). Diferenças volumétricas significativas foram encontradas entre o modelo proposto e o ACAPULCO e o CERES para controles, FRDA e pacientes com SCA3 ($p < 0,01$). O protocolo de segmentação manual do DN produziu máscaras padrão-ouro com alta concordância intra-avaliador (ICC médio: $0,906$) e inter-avaliador (ICC médio: $0,776$). O *pipeline* de localização e segmentação em dois passos atingiu um Dice de $0,898 \pm 0,031$ e $0,894 \pm 0,036$ para o DN esquerdo e direito, respectivamente. Em validação externa, nosso algoritmo superou a principal ferramenta automatizada, MRICloud (Dice para DN esquerdo/direito: $0,863 \pm 0,038 / 0,843 \pm 0,066$ vs. $0,568 \pm 0,222 / 0,582 \pm 0,239$). O modelo demonstrou excelente capacidade de generalização em conjuntos de dados não vistos em treinamento e correlação significativa com anotações manuais.

Os resultados iniciais mostraram precisão e confiabilidade promissoras na segmentação do cerebelo em imagens T1w. Continuamos nossos esforços para obter mais segmentações cerebelares de qualidade em conjuntos diversificados de imagens para retomar o treinamento do modelo. Além disso, fornecemos um modelo que segmenta com precisão o DN de imagens cerebrais QSM, que pode ser prontamente implementado. Essas ferramentas podem melhorar a busca por biomarcadores e a avaliação de terapias para doenças neurodegenerativas, fornecendo metodologias robustas para estudos observacionais e de história natural.

Palavras-chave: cerebelo; imageamento por ressonância magnética; aprendizado profundo; ataxias espinocerebelares; ataxia de Friedreich.

Abstract

The segmentation of the cerebellum and dentate nucleus (DN) in magnetic resonance imaging (MRI) presents unique challenges due to their complex structures and proximity to surrounding brain regions. The cerebellar cortex is intricately folded, and the DN has a low contrast-to-noise ratio as a consequence of being a small structure buried deep within the surrounding white matter. Traditional tools perform well on healthy controls but show lower performance when applied to images from individuals with cerebellar lesions. To improve the generalization and accuracy of automated segmentations of cerebellar structures, we propose implementing a deep learning model for the cerebellar cortex segmentation in T1-weighted (T1w) images and the DN in quantitative susceptibility mapping (QSM) images, for both healthy individuals and patients with cerebellar damage.

For cerebellum segmentation, we used T1w sequences, including images from healthy controls and patients with spinocerebellar ataxia (SCA) 1 and 3 and Friedreich's ataxia (FRDA). Expert neuroradiologists manually annotated segmentation masks, ensuring quality identification of cerebellar fissures. We trained a modified 3D U-Net architecture and performance was compared with ACAPULCO and CERES model results. For DN segmentation, we analyzed QSM images from 132 healthy controls and 170 patients with cerebellar ataxia or multiple sclerosis across nine datasets worldwide. Manual delineation of the DN (gray matter and white matter hilus) by experienced raters provided the ground truth data with a robust quality control process. A two-step approach for automated segmentation was implemented, composed of a localization model followed by DN segmentation.

Our cerebellum segmentation model achieved a Dice score of 0.928 ± 0.030 on the test set and outperformed ACAPULCO (Dice: 0.909 ± 0.043) and CERES (Dice: 0.918 ± 0.039). This difference is sustained in patients with the highest level of cerebellar damage (our model Dice: 0.924 ± 0.033 , ACAPULCO Dice: 0.904 ± 0.047 , CERES

Dice: 0.917 ± 0.044). Significant volumetric differences were found between our model and ACAPULCO and CERES for controls, FRDA, and SCA3 patients ($p < 0.01$). The DN manual tracing protocol produced ground truth data with high intra-rater (average ICC 0.906) and inter-rater reliability (average ICC 0.776). The two-step localization plus segmentation pipeline achieved a Dice score of 0.898 ± 0.031 and 0.894 ± 0.036 for left and right DN, respectively. In external validation, our algorithm outperformed the leading automated tool, MRICloud (left/right DN Dice: $0.863 \pm 0.038/0.843 \pm 0.066$ vs. $0.568 \pm 0.222/0.582 \pm 0.239$, respectively). The model demonstrated superior generalizability across unseen datasets and showed significant correlation with manual annotations.

Initial results showed promising accuracy and reliability in cerebellum segmentation in T1w images. We continue our effort to gather more expert cerebellar segmentations on a diverse dataset to resume the model training. Also, we provide a model that accurately segments the DN from brain QSM images and can be readily deployed. These tools can enhance biomarker discovery and therapeutic assessment in neurodegenerative disorders, providing robust methodologies for observational and natural history studies.

Keywords: cerebellum; magnetic resonance imaging; deep learning; spinocerebellar ataxias; Friedreich's ataxia.

Lista de Figuras

Figura 1	Slice coronal de imagem ponderada em T1 e QSM de um mesmo indivíduo.	27
Figura 2	Exemplo de segmentação do ACAPULCO em paciente com SCA1. Presença de artefato inferior ao cerebelo, delimitado inferiormente pelo recorte do localizador.	28
Figura 3	Exemplo de participante controle em plano coronal (superior) e axial (inferior). Segmentação manual à esquerda e resultado do MRICloud à direita. Erros de segmentação indicados pelas setas. Sobre-segmentações indicadas pelas setas laranja e sub-segmentação em azul.	32
Figura 4	Exemplo de imagem de um paciente com ataxia de Friedreich. (a) imagem ponderada em T1. (b) é o resultado da ferramenta Cereb-Net. (c) é o resultado da segmentação manual.	40
Figura 5	Slice coronal de imagem ponderada em T1 e QSM de um mesmo indivíduo.	42
Figura 6	Processo de reconstrução QSM incluindo <i>phase unwrapping</i> , <i>background field removal</i> e a inversão de dipolo.	43
Figura 7	Amostras de imagens QSM de cada <i>dataset</i> . Selecionados aleatoriamente. As imagens foram recortadas centralizadas no cerebelo. Intensidades dos <i>voxels</i> foram normalizados utilizando <i>z-score</i> para melhorar a visualização.	46
Figura 8	Fluxograma dos dados utilizados no estudo.	47
Figura 9	Pipeline do modelo de segmentação do núcleo denteado.	55
Figura 10	Métrica Dice no conjunto de teste. As barras de erro indicam intervalo de confiança de 95%. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$	59
Figura 11	Volume em cm^3 para amostras <i>out-of-sample</i> em pacientes com FRDA (a) e SCA3 (b). As barras de erro indicam intervalo de confiança de 95%. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$	59

Figura 12 Comparação de segmentações de um slice coronal para um paciente com ataxia espinocerebelar tipo 1. **(a)** é a máscara de segmentação produzida por um anotador experiente, o *ground truth*. **(b)** é o resultado do modelo treinado. Em **(c)** e **(d)** têm-se os resultados do ACAPULCO e CERES, respectivamente. Nota-se uma maior similaridade qualitativa entre o *ground truth* e a máscara predita pelo modelo U-Net. Já ACAPULCO e CERES apresentam problemas de sobre-segmentação. Essa imagem de ressonância magnética não fez parte do treinamento. 60

Lista de Tabelas

Tabela 1	Conjuntos de dados anotados manualmente por especialistas para segmentação do córtex cerebelar.	41
Tabela 2	Protocolos de aquisição para cada <i>dataset</i>	44
Tabela 3	Dados demográficos dos participantes para cada <i>dataset</i>	45
Tabela 4	Métricas de reprodutibilidade em conjunto de dados de teste-reteste Kirby-21.	61

Lista de Abreviações e Siglas

ACAPULCO	Automatic Cerebellum Anatomical Parcellation using U-Net with Locally Constrained Optimization
ANN	do inglês, <i>artificial neural network</i> ou rede neural artificial
ANTs	Advanced Normalization Tools
BraTS	Brain Tumor Segmentation
CERES	CEREbellum Segmentation
CHoP	Children's Hospital of Philadelphia
CNN	do inglês, <i>convolutional neural networks</i> ou rede neural convolucional
CSF	do inglês, <i>cerebrospinal fluid</i> ou líquido ou líquido cefalorraquidiano
CT	do inglês, <i>computed tomography</i> ou tomografia computadorizada
DICOM	Digital Imaging and Communications in Medicine
DN	do inglês, <i>dentate nucleus</i> ou núcleo denteado
DS	do inglês, <i>deep supervision</i>
FARA	Friedreich's Ataxia Research Alliance
FCN	do inglês, <i>fully convolutional network</i> ou rede totalmente convolucional
FoV	do inglês, <i>field of view</i>
FRDA	ataxia de Friedreich
FWHM	do inglês, <i>full-width half maximum</i> ou largura à meia altura
GM	do inglês, <i>gray matter</i> ou substância cinzenta
GPU	do inglês, <i>graphics processing unit</i> ou unidade de processamento gráfico
GRE	do inglês, <i>gradient-recalled echo</i>
ICC	do inglês, <i>intraclass correlation coefficient</i> ou coeficiente de correlação intraclasse

IRB	do inglês, <i>institutional review board</i>
LLM	do inglês, <i>large language model</i> ou grande modelo de linguagem
LSTM	do inglês, <i>long short-term memory</i>
MLP	Multi-Layer Perceptron
MNI	Montreal Neurological Institute
MPRAGE	do inglês, <i>magnetization-prepared rapid gradient echo</i>
MRI	do inglês, <i>magnetic resonance imaging</i>
MS	do inglês, <i>multiple sclerosis</i> ou esclerose múltipla
NIfTI	Neuroimaging Informatics Technology Initiative
NLP	do inglês, <i>natural language processing</i> ou processamento de linguagem natural
OPAL	Optimized PatchMatch Label <i>fusion</i>
QSM	do inglês, <i>quantitative susceptibility mapping</i> ou mapeamento quantitativo de suscetibilidade magnética
ReLU	do inglês, <i>rectified linear unit</i>
RM	ressonância magnética
ROBEX	Robust Brain Extraction
RoI	do inglês, <i>region of interest</i>
RSNA	Radiological Society of North America
SAM	Segment Anything Model
SARA	Scale for the Assessment and Rating of Ataxia
SCA	do inglês, <i>spinocerebellar ataxia</i>
SD	do inglês, <i>standard deviation</i> ou desvio padrão
SGD	do inglês, <i>stochastic gradient descent</i>
SPM	Statistical Parametric Mapping
SUIT	Spatially Unbiased Infratentorial Template
SWI	do inglês, <i>susceptibility-weighted images</i> ou imagens ponderadas em suscetibilidade
T	Tesla
T1w	do inglês, <i>T1-weighted</i> ou ponderada em T1
TE	tempo de eco

TR tempo de repetição

ViT Vision Transformer

WM do inglês, *white matter* ou substância branca

Sumário

1	Introdução	18
2	Objetivos	20
2.1	Geral	20
2.2	Específicos	20
3	Revisão da Literatura	21
3.1	Segmentação do Córtex Cerebelar	23
3.1.1	SUIT	23
3.1.2	CERES	24
3.1.3	ACAPULCO	25
3.1.4	CerebNet	27
3.2	Segmentação do Núcleo Denteado	29
3.2.1	MRICloud	31
3.3	Deep Learning	32
4	Materiais e Métodos	37
4.1	Dados	38
4.1.1	Definição da Casuística	38
4.1.2	Córtex Cerebelar	40
4.1.3	Mapeamento Quantitativo de Suscetibilidade Magnética	41
4.1.4	Núcleo Denteado	43
4.1.5	Arquiteturas de Deep Learning	49
4.2	<i>Data Augmentation</i>	53
4.3	Pipeline de Segmentação	53
4.4	Métricas e Análise Estatística	56
4.5	Disponibilização dos Modelos	57
5	Resultados	58
5.1	Segmentação do Cerebelo	58
5.2	Segmentação do Núcleo Denteado	61
6	Discussão	101
7	Conclusão	108
8	Referências Bibliográficas	109
9	Anexos	120
9.1	Parecer do Comitê de Ética	120

9.2 Comprovante de Submissão do Manuscrito	130
--	-----

1 Introdução

O cerebelo é uma estrutura localizada na região posterior do encéfalo e está relacionada ao controle motor. No entanto, estudos recentes da literatura têm mostrado sua relação também com a cognição (1). A relação do cerebelo em diferentes doenças neurodegenerativas e sua função relacionada à cognição têm aumentado o interesse de pesquisas nessa estrutura cerebral como alvo primário de investigação (2, 3, 4, 5, 6). De modo especial, o núcleo denteado (do inglês, *dentate nucleus*, DN), o maior dos núcleos cerebelares profundos que está localizado na porção central de cada um dos hemisférios cerebelares, tem ganhado atenção por estar envolvido no processo sensoriomotor e não-motores (4, 6).

O estudo *in vivo* dessas estruturas pode ser realizado com o imageamento por ressonância magnética (do inglês, *magnetic resonance imaging*, MRI). Essa técnica permite a obtenção de imagens anatômicas estruturais e funcionais, com o uso de radiação não-ionizante, e tem se destacado na identificação de potenciais biomarcadores (7, 8). A segmentação precisa das estruturas cerebelares é uma tarefa desafiadora, consequência da presença de invaginações e fissuras no córtex cerebelar, além da proximidade com o cérebro. No caso do núcleo denteado, seu pequeno tamanho e a alta concentração de ferro acabam por dificultar o correto delineamento das bordas e sua separação dos demais núcleos cerebelares. Este problema é, de fato, agravado em condições de doenças neurodegenerativas, como as ataxias cerebelares, que podem alterar e reduzir o volume e o formato do DN.

A fim de superar esses desafios supracitados, a comunidade acadêmica criou ferramentas de primeira geração baseados em neuroimagem (9, 10). Porém, essas soluções apresentam problemas quanto à precisão — sobre e sub-segmentação —, reprodutibilidade e limitações de generalização, particularmente quando aplicados em conjuntos de dados de MRI os quais possuem imagens de estruturas atrofiadas. Isso resulta em quantificações de volume com baixa precisão, impactando estudos

longitudinais e a identificação de potenciais biomarcadores de imagem para ensaios clínicos.

De fato, a ausência de biomarcadores, principalmente de imagem, em doenças cerebelares é uma séria limitação para o desenvolvimento de novos ensaios clínicos, rastreamento farmacodinâmico e estratificação dos pacientes (11, 12). Para endereçar tais desafios, há um reconhecimento crescente na comunidade acadêmica do potencial da neuroimagem como meio de atender essas necessidades. Esse advento tem levado a investimentos substanciais na criação de conjuntos de dados de imagens de ressonância, como o TRACK-FA¹, e no compromisso da comunidade internacional em esforços de pesquisa em larga escala, como o ENIGMA-Ataxia². No entanto, a identificação de medidas de desfecho sensíveis, confiáveis e escaláveis depende não apenas da existência abundante de dados de alta qualidade, mas também da disponibilidade de ferramentas analíticas otimizadas, automatizadas e precisas. Enquanto a necessidade de recursos de dados está sendo amplamente atendida, a disponibilidade de ferramentas personalizadas otimizadas para avaliar de forma acurada o cerebelo representa uma lacuna crítica.

Portanto, o objetivo deste trabalho é aproveitar a disponibilidade e maximizar a utilidade dos recursos de *datasets* de alta qualidade que estão sendo coletados e organizados (TRACK-FA, ENIGMA-Ataxia e outros estudos acadêmicos internos) para desenvolver e fornecer ferramentas analíticas atualizadas e otimizadas da substância cinzenta (do inglês, *gray matter*, GM) do cerebelo (córtex e núcleo denteado). Em particular, modelos baseados em aprendizado de máquina profundo, a fim de obter ferramentas automatizadas que possam ser prontamente implementadas para uso em ensaios clínicos observacionais, estudos de história natural e de tratamentos. Uma avaliação quantitativa precisa e sensível das estruturas cerebelares é essencial para maximizar a probabilidade de descoberta bem-sucedida de biomarcadores de neuroimagem para acompanhar o estágio e a progressão de doenças e, inclusive, levar a novas possíveis compreensões sobre mecanismos fisiopatológicos subjacentes.

¹<https://clinicaltrials.gov/study/NCT04349514>, <https://www.monash.edu/medicine/trackfa>

²<https://enigma.ini.usc.edu/ongoing/enigma-ataxia/>

2 Objetivos

2.1 Geral

O objetivo deste trabalho é o desenvolvimento de ferramentas de segmentação automáticas e precisas para o córtex cerebelar e os núcleos denteados utilizando algoritmos de *deep learning* a partir de imagens de ressonância magnética.

2.2 Específicos

- Organizar bases de imagens de ressonância magnética ponderadas em T1 (do inglês, *T1-weighted*, T1w) e de mapeamento quantitativo de suscetibilidade magnética (do inglês, *quantitative susceptibility mapping*, QSM) para treinamento e para teste-reteste composto por indivíduos com e sem atrofia cerebelar;
- Criar e organizar um conjunto de segmentações manuais de alta qualidade das estruturas de interesse, conduzido por neurorradiologistas experientes;
- Explorar diferentes arquiteturas de redes neurais estado da arte para a tarefa objetivo de segmentação de imagens (U-Net (13, 14), nnU-Net (15) e outras baseadas em Transformers (16, 17, 18));
- Criar modelos de segmentação usando algoritmos de *deep learning* para o córtex cerebelar (T1w) e os núcleos denteados (QSM);
- Avaliar o desempenho e a reprodutibilidade dos modelos propostos e comparar com os atuais métodos de segmentação estado da arte para o cerebelo e para o núcleo denteado.

3 Revisão da Literatura

O cerebelo e o núcleo dentado têm recebido cada vez mais atenção da comunidade acadêmica devido à sua importância fisiológica em uma ampla gama de funções motoras, cognitivas, sociais e de linguagem, mas também seu envolvimento com doenças neurodegenerativas (19, 20, 21). Nestas, a degeneração pode levar a distúrbios de movimento e danos cognitivos (22, 23, 24, 25). Estudos recentes têm evidenciado os impactos de atrofia do cerebelo em epilepsia (26), na doença de Alzheimer (27), esquizofrenia (28), além das ataxias espinocerebelares. Já anormalidades no núcleo dentado podem comprometer as redes cerebrais em larga escala, afetando a conectividade do córtex cerebelar com o córtex cerebral, como em pacientes com doença de Alzheimer, autismo e esquizofrenia (29, 4, 6). Além disso, estudos neuropatológicos forneceram evidências de que o DN desempenha um papel central na patogênese de várias doenças cerebelares, especialmente nas ataxias cerebelares hereditárias (30).

Anatomicamente, o cerebelo está conectado ao córtex cerebral por meio de conexões bem organizadas, estando ligado por núcleos cerebelares profundos que recebem entradas das células de Purkinje do córtex cerebelar e originam fibras eferentes que se projetam no córtex cerebral através do tálamo, mesencéfalo, tronco encefálico e medula espinhal (31). O artigo do atlas Schmahmann (32) define as nomenclaturas para as subdivisões do cerebelo em regiões, fissuras e lóbulos. No cerebelo há uma porção mediana denominada vérmis, ao qual estão lateralmente conjuntos dois hemisférios cerebelares. Estruturalmente, possui uma estrutura foliar, no qual existem sulcos mais evidenciados, denominados fissuras cerebelares. Uma consequência disso é que, dado um mesmo volume, a sua superfície é aumentada por essas inúmeras invaginações. O corpo medular cerebelar, um centro de substância branca (do inglês, *white matter*, WM) com suas ramificações irradiando-se em direção à superfície cerebelar externa, é envolto por uma camada de substância cinzenta, o córtex cerebelar,

de espessura variada e constricta. Essa descrição evidencia a principal explicação da dificuldade de prover uma delimitação manual acurada das bordas do cerebelo em uma imagem de ressonância magnética (RM). Facilmente é possível ignorar indevidamente determinadas fissuras, maiores ou menores, resultando em uma quantificação volumétrica imprecisa. Em geral, esses erros resultam em uma mensuração superior à verdadeira, consequência da consideração de espaços de líquido (fissuras) como tecido de substância cinzenta.

Nesse contexto, diversas ferramentas de neuroimagem foram desenvolvidas com o constante aprimoramento de técnicas e abrangência de dados. Dentre eles, para o cerebelo existe o Spatially Unbiased Infratentorial Template (SUIT) (9), o CEREBellum Segmentation (CERES) (10), o Automatic Cerebellum Anatomical Parcellation using U-Net with Locally Constrained Optimization (ACAPULCO) (33) e, mais recentemente, o CerebNet (34). Sendo que os dois primeiros são baseados em métodos probabilísticos baseados em atlas e os dois últimos em aplicação de modelos baseados em *deep learning*. Já para o núcleo denteado, apenas uma solução existente é considerada, o MRICloud (35, 36). Contudo, as segmentações produzidas por essas soluções, para ambos alvos anatômicos, sofrem, em maior ou menor frequência, com diversos problemas, desde erros de segmentação, tempo de processamento, necessidade de pós-processamento, limitações de dados de entrada, entre outros. Isso impacta na não adequabilidade do emprego de tais soluções em cenários em que medidas quantitativas precisas sejam cruciais, como em estudos longitudinais com pacientes de doenças neurodegenerativas que acometem os órgãos de interesse. Desse modo, há uma oportunidade para a proposição e o desenvolvimento de novos modelos acurados e validados que atendam às exigências atuais de precisão e reprodutibilidade.

A constante evolução dos equipamentos de MRI também é acompanhada por novos protocolos de aquisição. Nestes, os parâmetros estão permitindo um imageamento com cada vez maior resolução espacial, ou seja, o tamanho dos *voxels* tende a diminuir, conseqüentemente permitindo maior nível de detalhamento. Todavia, os métodos de segmentação automatizados disponíveis atualmente, em sua maioria, compreendem aprendizado ou referência restrito a imagens com tamanho de *voxel* igual ou superior a 1 mm. Desse modo, não são otimizados para os novos padrões de

protocolos com aquisições de maior resolução.

A seguir são abordados os desafios e atuais soluções disponíveis para segmentação do cerebelo e do núcleo denteado.

3.1 Segmentação do Córtex Cerebelar

A segmentação do córtex cerebelar ocorre, a princípio, em imagens de RM ponderadas em T1, que permite um contraste adequado entre substância branca, cinzenta e líquido. Porém, demais técnicas e sequências podem ser utilizadas de forma complementar à distinção de algumas estruturas.

Com a necessidade de se avaliar a estrutura morfológica do cerebelo de modo mais específico e com a falta de ferramentas dedicadas para tal, em meados de 2006, Diedrichsen e colegas propuseram o primeiro método de análise dedicado, especificamente, para o cerebelo, o SUIT (9). Entretanto, desde então, tal tarefa tem se mostrado desafiadora já que o cerebelo é um órgão irregular e compacto. Não obstante, a limitação da resolução espacial pelos equipamentos de ressonância magnética também dificulta a segmentação das estruturas cerebelares. Apesar disto, outros métodos foram propostos, tais como o CERES (10) e o ACAPULCO (33), cada qual com sua metodologia particular. Por fim, mais recentemente, o CerebNet foi proposto, baseado em uma abordagem de aprendizado profundo (34).

3.1.1 SUIT

O SUIT foi desenvolvido como *toolbox* para MATLAB e SPM¹ (Statistical Parametric Mapping), no qual foram utilizadas imagens do cerebelo de 20 indivíduos saudáveis para criar um atlas a partir da geometria média desse grupo. Exalta-se por ser um método não-viesado espacialmente, no qual são preservadas características anatômicas do cerebelo (9). Isso é alcançado, pois na etapa de geração do atlas, os cerebelos de todos os indivíduos são alinhados, após serem isolados, e são normalizados, obtendo-se uma média das deformações. No entanto, um dos pontos negativos é a necessidade de correção, em muitos casos, na etapa de isolamento do cerebelo e tronco cerebral, demandando muito tempo para a execução dessa tarefa.

¹<https://www.fil.ion.ucl.ac.uk/spm/>

As imagens são alinhadas, inicialmente, ao longo da comissura anterior, a fim de minimizar erros de processamento, o que é uma exigência da ferramenta. Em resumo, as etapas do processamento do SUIT consistem, primeiramente, no isolamento das estruturas infratentoriais e, então, na identificação dos mapas de segmentação para a substância cinzenta, branca e líquido ou líquido cefalorraquidiano (do inglês, *cerebrospinal fluid*, CSF), dado por um mapa de probabilidades. Na sequência são feitas correções para a segmentação da GM. Por fim, o mapa de GM será normalizado para o *template* do SUIT e um teste de homogeneidade é conduzido para descartar sujeitos que possuem uma variância muito grande, quando comparados à média dos indivíduos (identificação de *outliers*) (37). A normalização compreende uma transformação não-linear para o cerebelo de um indivíduo particular, o Colin 27 (38, 32). Com o intuito de homogeneizar ainda mais os indivíduos, uma suavização por um filtro gaussiano de *kernel* com 3 mm de largura à meia altura (do inglês, *full-width half maximum*, FWHM) usando o software SPM pode ser utilizada.

3.1.2 CERES

Por sua vez, o CERES possui um *pipeline* totalmente automatizado, ou seja, sem interação manual, que permite a segmentação e parcelamento do cerebelo. Diferentemente do SUIT, nesse projeto a abordagem é multi-atlas, isto é, a segmentação é feita em comparação com mais de um atlas, sendo estes oriundos de sujeitos saudáveis. Apesar disso, o conjunto de imagens utilizado foi registrado no espaço MNI (Montreal Neurological Institute) a 1 mm^3 , portanto não otimizado para imagens com maior resolução. O CERES aplica uma adaptação do método Optimized PatchMatch Label fusion (OPAL) (39). Neste, a vantagem expressa no artigo é que a complexidade do algoritmo não aumenta conforme o tamanho da biblioteca de *templates*, mas sim com o tamanho da imagem que se está analisando. Isso acontece, pois o PatchMatch implementa uma busca aleatória iterativa no conjunto de *templates*. Ao não explorar todos os exemplos da biblioteca, diferentes subconjuntos de *patches* são utilizados a cada execução, o que também o define como não-determinístico. Em comparação com o SUIT, o CERES apresenta um melhor desempenho na segmentação dos lóbulos, ainda que o tempo de processamento seja equivalente. No SUIT foi observada tendência à segmentação dos lóbulos além dos limites e ao preenchimento das fis-

suras cerebelares. Por fim, o CERES foi atualizado para o CERES2 (40), no qual os autores citam uma correção de erros com uso de rede neural, porém essa solução não está disponível para uso público.

O CERES é uma ferramenta desenvolvida para segmentação e análise automatizada do cerebelo em MRI disponibilizada em interface web. A sequência de passos inicia-se com uma imagem T1w em formato NIfTI, o qual é submetido a um pré-processamento com várias etapas. Primeiramente, a imagem passa por uma redução de ruído seguido pela correção de inomogeneidade no espaço nativo com o algoritmo *N4 bias field correction* (41). Em sequência, um registro linear (transformação *affine*) para o *template* MNI152 é realizado. Novamente é aplicada uma correção de inomogeneidade, mas agora no espaço MNI. Por fim, a imagem é reduzida, sendo recortada contemplando-se a área do cerebelo a partir de limites calculados nas imagens de referência no espaço MNI. Após, um registro não-linear para o *template* MNI152 é calculado e é feita uma normalização de intensidade. Para a atribuição dos *labels*, é construído um conjunto específico para o caso analisado, a partir de transformações não-lineares das imagens segmentadas da biblioteca. No entanto, uma limitação é que essa biblioteca de casos anotados manualmente foi construída com apenas cinco imagens T1w. Por fim, ocorre a atribuição das classes aos *voxels* com o procedimento multi-atlas, combinando as segmentações de referência da biblioteca (42, 39). Como pós-processamento, é aplicada uma regularização para suavização de possíveis irregularidades nas bordas. Ao final, é produzido um relatório detalhado com volumes do cerebelo e dos lóbulos, assim como espessura do córtex cerebelar e índices de assimetria entre os tecidos no espaço nativo.

3.1.3 ACAPULCO

O ACAPULCO, por outro lado, destaca-se pela utilização de *deep learning*, dedicado a segmentar e quantificar os lóbulos cerebelares (33). Dado que é um método baseado em redes neurais convolucionais (do inglês, *convolutional neural networks*, CNNs), diferentemente dos modelos mostrados anteriormente, não há aqui uma segmentação baseada em atlas, mas um modelo treinado a partir de um conjunto de treinamento manualmente segmentado. É composto por duas redes CNN 3D, uma de localização (obtenção de *bounding box* da região cerebelar) e uma de

parcelamento propriamente dito. Essa segunda rede é baseada na arquitetura U-Net, na qual a saída é um mapa de probabilidades para cada um dos *labels* das regiões para todos os *voxels*. Em comparação com os métodos anteriores, aqui tem-se um treinamento que compreende exemplos de indivíduos saudáveis (seis) e pacientes com atrofia cerebelar (nove). Além disso, um *dataset* pediátrico com 20 crianças, 10 saudáveis e 10 pacientes, é utilizado para prover um segundo modelo. O ACAPULCO apresenta performance satisfatória em conjuntos de teste com exemplos de atrofia severa, como o Kwyjibo (43). Por fim, o tempo de processamento é inferior ao das ferramentas anteriores, sendo um pouco mais de um minuto. No entanto, em estudo de reprodutibilidade (44), após duas análises de um mesmo conjunto de dados em hardware e software idênticos, a diferença entre duas execuções do ACAPULCO foi alta, relativamente maiores que no caso do CERES. Desse modo, diz-se que a replicabilidade é prejudicada. Por fim, o ACAPULCO foi descontinuado e não receberá mais atualizações após a última versão publicada (0.3.0) há mais de dois anos.

O *pipeline* da ferramenta segue o protocolo descrito em (33). Em resumo, cria-se uma máscara do cérebro extraída com Robust Brain Extraction (ROBEX) (45) seguida da correção de *bias field* com o algoritmo N4 (41). Em seguida, a imagem é linearmente registrada no espaço MNI utilizando ANTs (Advanced Normalization Tools) (46, 47). A solução é composta por duas redes convolucionais 3D. A primeira é a rede de localização que define um *bounding box* em torno da região cerebelar e a segunda, uma U-Net, cujo objetivo é obter o parcelamento, fornecendo, ao final do processamento, um mapa de probabilidades de cada *voxel* da imagem pertencer a uma determinada classe ou *label*. A ferramenta ACAPULCO retorna o volume dos lóbulos cerebelares para ambos hemisférios. Uma lista com as regiões cerebelares identificadas pelo ACAPULCO pode ser encontrada no artigo publicado (33).

Apesar de o CERES e o ACAPULCO representarem avanços na qualidade da segmentação do cerebelo em comparação com as soluções de primeira geração como o FreeSurfer e o SUIT, ainda há uma lacuna para melhoria. Dentre os problemas encontrados, está a presença de artefatos de sobre e sub-segmentação, especialmente nas fissuras cerebelares, mais pronunciadas e pacientes com algum grau de atrofia cerebelar. Nesses casos, os modelos acabam ignorando algumas fissuras, classificando-as incorretamente como parte do córtex cerebelar (Figura 1), apesar de

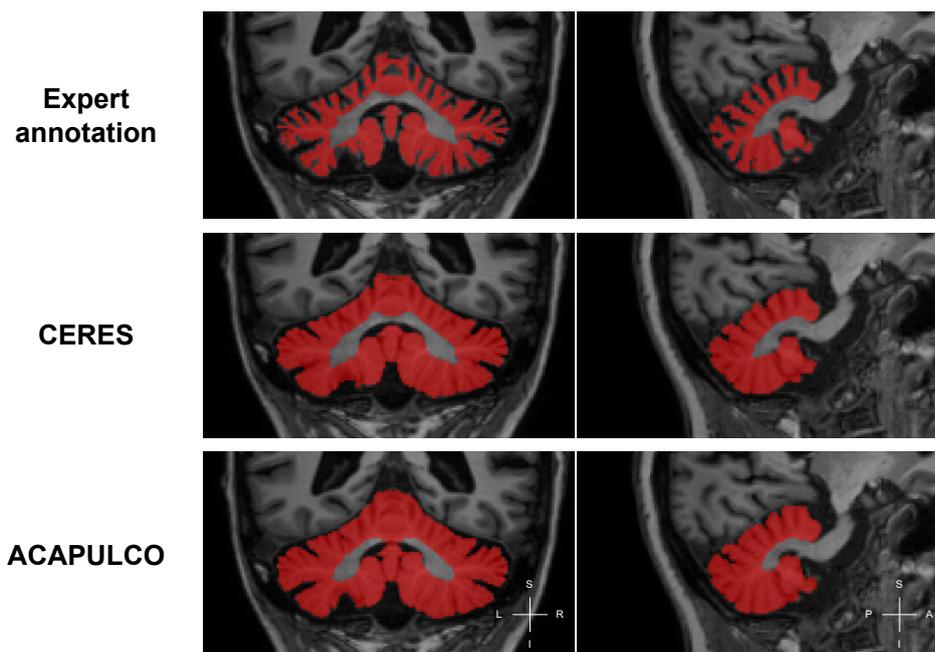


Figura 1 – Slice coronal de imagem ponderada em T1 e QSM de um mesmo indivíduo.

serem espaços preenchidos por líquido. Outro problema constatado foram alguns artefatos extra-cerebelares como apresentado na Figura 2, onde há uma massa na região do pescoço, nesse exemplo de uma análise com o ACAPULCO.

3.1.4 CerebNet

O CerebNet é um modelo de segmentação e parcelamento do cerebelo que utiliza *deep learning*. Diferentemente do ACAPULCO, cujo modelo é uma rede CNN 3D, o CerebNet adota redes baseadas em U-Net 2.5D, o FastSurferCNN (48). Metodologicamente, a ferramenta é composta por três redes U-Net 2.5D, uma para cada plano (axial, coronal e sagital) possuindo como entrada, além do *slice* central, três *slices* anteriores e três posteriores. Essa técnica de *spatial information aggregation* (SPI) permite a recuperação de informação contextual dos *voxels* centrais. A partir dessa estratégia de combinação de modelos, tem-se uma combinação dos três mapas de probabilidade gerados. Os autores citam como vantagens dessa abordagem, frente à adoção de uma arquitetura 3D, é a redução de requisito de memória e maior quantidade de exemplos apresentados a cada modelo para treinamento, visto que um volume 3D é repartido em vários *slices* 2D. Além disso, há um intenso uso de *data augmentation* em uma estratégia que combina deformações não-lineares e registro para

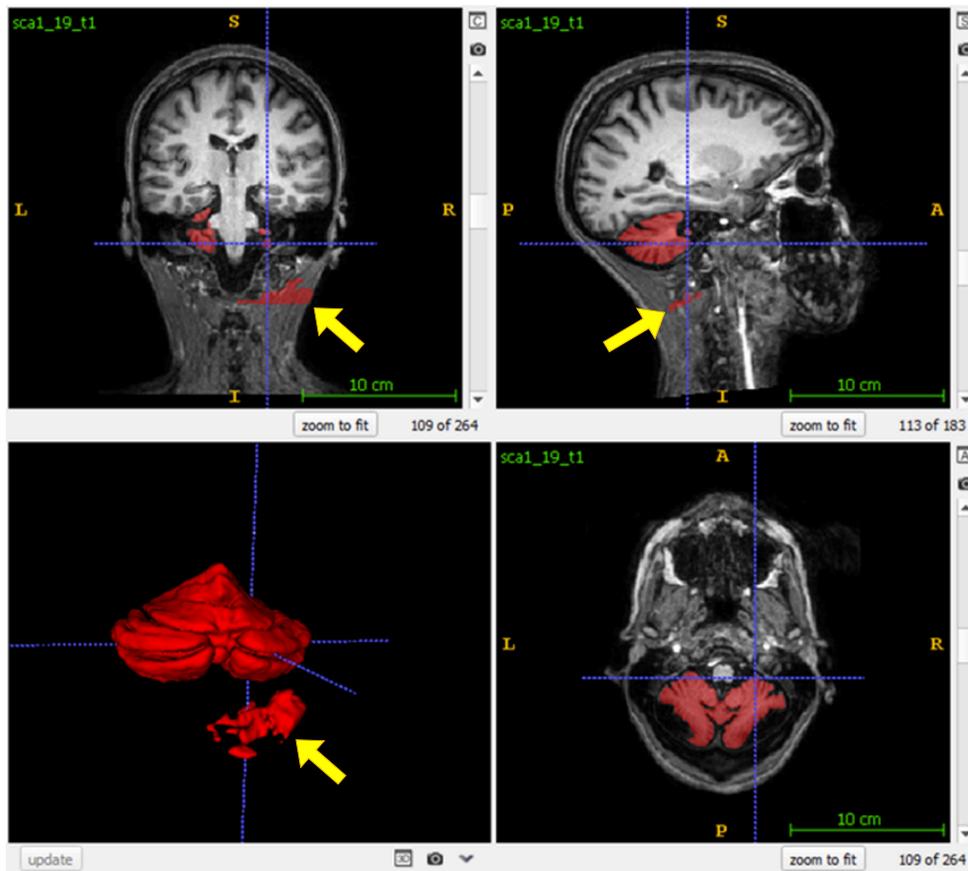


Figura 2 – Exemplo de segmentação do ACAPULCO em paciente com SCA1. Presença de artefato inferior ao cerebelo, delimitado inferiormente pelo recorte do localizador.

geração de amostras artificiais com intuito de incrementar a variabilidade anatômica para treinamento. Uma etapa relevante do estudo é a realização de um pré-treino do modelo utilizando máscaras de segmentação produzidas por outros softwares, o SUIT (9) e o FreeSurfer (49). Para isso, foi realizada uma fusão dos *labels* das saídas desses dois modelos, tendo como principal o *output* do SUIT e aproveitando apenas a anotação para substância branca do FreeSurfer. Esse estudo foi realizado com um conjunto de dados de 30 imagens T1w, dos quais 20 eram de pacientes com ataxia espinocerebelar (do inglês, *spinocerebellar ataxia*, SCA) do tipo 3 (SCA3), abrangendo desde a fase pré-atáxica até estágios avançados. Apesar dos dados serem provenientes de quatro centros distintos, três europeus e um americano, todas as imagens foram adquiridas em ressonâncias Siemens (Siemens Medical Systems, Erlangen, Alemanha) de 3 Tesla (T) com o mesmo protocolo. Além disso, tal qual o CERES, as imagens utilizadas são todas em 1 mm, inclusive cita-se que o *pipeline* aceita apenas como entrada imagens nessa resolução.

O *pipeline* do CerebNet está incluído como um módulo do FastSurfer (48). Partindo de uma imagem estrutural ponderada em T1, não é necessário realizar nenhum pré-processamento, como normalização espacial ou correção de inhomogeneidade. Ao final, a saída é um mapa de probabilidades para 27 classes, incluindo classes para os lóbulos cerebelares. Os autores reportam uma performance superior ao ACAPULCO retreinado e a uma combinação SUIT+FS (FreeSurfer) (9, 49), em métrica Dice, distância Hausdorff e similaridade de volume. Uma limitação do estudo é que todas as aquisições foram realizadas em ressonâncias de um único fabricante. Além disso, o *pipeline* aceita apenas imagens com resolução de 1 mm.

3.2 Segmentação do Núcleo Denteado

Apesar da importância do núcleo denteado em circuitos cérebro-cerebelares e de sólidas evidências do seu envolvimento em doenças neurológicas, há uma escassez de estudos quantitativos *in vivo* em neuroimagem dessa estrutura e de demais núcleos cerebelares profundos (50, 51, 52, 23). O estudo do DN é particularmente desafiador devido às características de seu tecido e seu reduzido tamanho que dificultam a visualização em sequências de imagens ponderadas em T1 e

T2. Há exceções, como casos específicos de pacientes com repetições de aquisições auxiliadas por administração de contraste com base em gadolínio, cujo acúmulo permite uma rasa visibilidade em T1, por exemplo (53). Já em imagens ponderadas em suscetibilidade (do inglês, *susceptibility-weighted images*, SWI), dado o acúmulo de ferro no tecido do DN, pode-se visualizar, ainda que de forma limitada, seu contorno (54, 55, 56). Embora útil em análises qualitativas com aplicações clínicas em detecção de anomalias cardíacas, aquisições SWI estão sujeitas a efeitos de *blooming*, distorcendo as bordas, justamente consequência dos depósitos de ferro. Recentemente, com o desenvolvimento do mapeamento quantitativo de suscetibilidade magnética (57), permitiu-se uma análise quantitativa, em contraste com o SWI. Além disso, o QSM permite um melhor imageamento anatômico do denteado e uma mensuração direta quantitativa da concentração de ferro do tecido (58) e, por consequência, a avaliação dessa estrutura (50, 23, 59).

O QSM tem sido utilizado no estudo e acompanhamento de alterações no DN em indivíduos com ataxias cerebelares hereditárias, incluindo ataxia de Friedreich (FRDA) (24, 23) e ataxias espinocerebelares (50). Além disso, estudos investigam as mudanças estruturais e de suscetibilidade magnética no DN tanto relacionadas ao envelhecimento normal quanto a distúrbios de movimento, incluindo a doença de Parkinson (51, 60, 61). Isso tudo demonstra a utilidade do QSM na neuroimagem para a análise qualitativa e quantitativa do DN em indivíduos saudáveis e pacientes, o que motiva também seu emprego em investigações envolvendo outras doenças neurológicas, do desenvolvimento e psiquiátricas que afetam estruturas cerebelares.

Embora o QSM seja relevante para estudos do DN e rastreamento das alterações em doenças, uma barreira é a dependência de anotações manuais, principalmente em estudos de larga escala. Para contornar essa limitação, métodos automatizados são necessários. Desse modo, dos estudos na literatura envolvendo a segmentação automática em imagens QSM dos núcleos cerebelares profundos, destaca-se apenas um, o MRICloud (35, 36), contrapondo a maior disponibilidade de soluções para a segmentação do cerebelo.

3.2.1 MRICloud

O MRICloud (35, 36) é uma ferramenta disponibilizada pela web que utiliza uma abordagem multi-atlas para realizar a segmentação em imagens QSM de forma automática. É considerado o *pipeline* estado da arte para segmentação em QSM, particularmente para o DN. O multi-atlas foi constituído com 10 atlas, ainda que na plataforma web haja menção e possibilidade de escolha de um processamento com 23 atlas. Após o processamento pelos atlas, uma etapa de multi-atlas *fusion* (62, 63) é executada para compor o resultado final. O conjunto de dados utilizados para formação dos atlas compreende apenas indivíduos saudáveis escaneados com uma ressonância Philips (Philips Healthcare, Países Baixos) 3 Tesla. Além do núcleo denteado, o MRICloud provê segmentações para o núcleo caudado, putâmen, globo pálido interno e externo, tálamo, pulvinar, núcleo subtalâmico, substância negra e núcleo rubro.

O MRICloud requer que seja fornecido, além da imagem QSM, uma imagem T1w. Além de estarem co-registradas, é necessário que a T1w tenha sido processada por uma ferramenta de *skull-stripping* previamente. Ainda, ambas imagens devem estar no formato ANALYZE. O processamento do *pipeline* se dá de forma assíncrona. Isto é, após submeter uma requisição pela interface web, é necessário aguardar a finalização do processamento para então ser realizado o download dos resultados. Os autores, no entanto, reportam um tempo de $1,6 \pm 0,5$ h em um *cluster* próprio. Além disso, destaca-se que os resultados são retornados também no formato ANALYZE. Em relação ao desempenho dessa ferramenta, nota-se uma acurácia inconsistente em diferentes conjuntos de dados. Aliás, não foram utilizadas imagens de pacientes de doenças neurodegenerativas. Apesar dos autores citarem que o conjunto de dados abrange razoavelmente variações anatômicas, divergências acentuadas de um novo exemplo com os demais atlas pode comprometer a qualidade final da segmentação (64).

Em um exemplo comparativo do resultado do MRICloud com uma segmentação manual conduzida por especialista, pode-se notar a presença de problemas de sobre e sub-segmentação dos núcleos denteados (Figura 3), o que afeta a precisão, ainda mais se considerando de uma estrutura pequena e compacta.

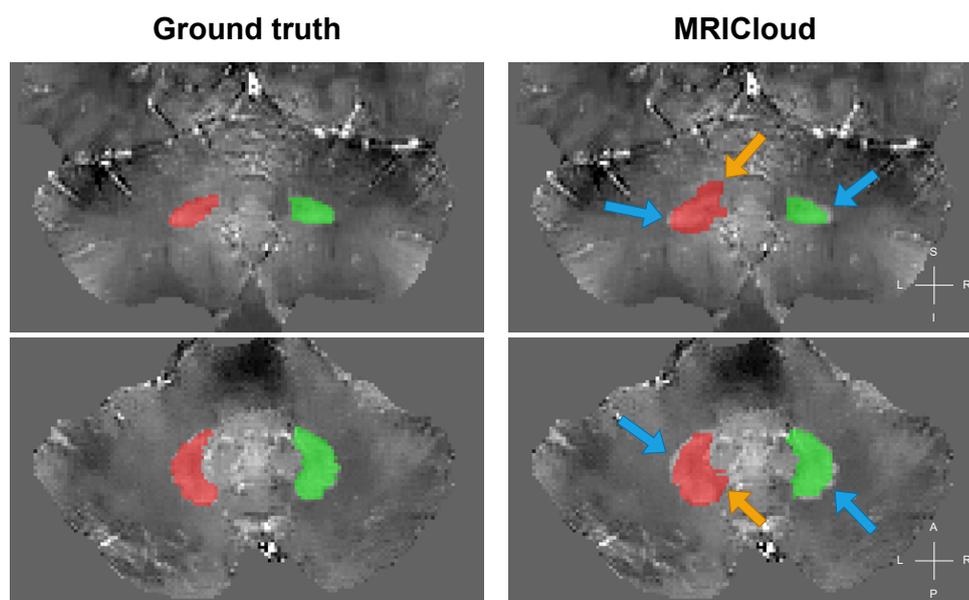


Figura 3 – Exemplo de participante controle em plano coronal (superior) e axial (inferior). Segmentação manual à esquerda e resultado do MRICloud à direita. Erros de segmentação indicados pelas setas. Sobre-segmentações indicadas pelas setas laranja e sub-segmentação em azul.

3.3 Deep Learning

No contexto da inteligência artificial, o aprendizado de máquina engloba o subcampo de aprendizado profundo (do inglês, *deep learning*) (65). Trata-se de um método no qual as representações dos dados são aprendidas com o uso de redes neurais artificiais (do inglês, *artificial neural networks*, ANNs) profundas. Recentemente, o *deep learning* tem recebido atenção devido a resultados satisfatórios, seja em aplicações em visão computacional, processamento de linguagem natural e na medicina (66).

Nas últimas décadas, é evidente a evolução das unidades de processamento gráfico (do inglês, *graphics processing unit*, GPUs) que têm na multiplicação de matrizes sua maior vantagem, o que as tornam essenciais tanto na renderização em computação gráfica quanto na aceleração de processamento de redes neurais. Com isso, o avanço e o desenvolvimento de novos modelos foi possível através do atendimento da necessidade intrínseca de poder computacional para processamento dos milhares de parâmetros pertencentes a essas redes. Tais modelos são compostos por um grande número de unidades, ou nós, inspirados em neurônios e, assim, agrupados em camadas (do inglês, *layers*), baseiam-se na organização estrutural de redes

interconexas e funcionalidade do cérebro humano. Em cada nó ocorre a soma das entradas multiplicadas por pesos que, após adicionar um valor de *bias*, é submetida a uma função de ativação, cujo resultado é repassado adiante na rede. Em geral, tem-se uma camada de entrada, uma ou mais camadas intermediárias, ou escondidas (do inglês, *hidden layers*), e uma camada de saída (do inglês, *output layer*). Essa é uma descrição sucinta e simplificada de uma rede neural artificial, denominada *Multi-Layer Perceptron* (MLP). Em uma MLP, uma camada possui seus nós conectados a todos da próxima camada, ou seja, é uma camada densamente conectada (do inglês, *fully connected layer*). Quando uma ANN é composta por duas ou mais camadas escondidas, convencionou-se denominá-la uma rede neural profunda.

Definir uma arquitetura de uma rede neural não envolve apenas determinar a quantidade de camadas ou o número de nós em cada uma delas. Existem diversas funções de ativação, como a sigmoide e a ReLU (do inglês, *rectified linear unit*). Esta última é uma função linear por partes — não linear, portanto —, amplamente adotada, principalmente pela rapidez no cálculo do gradiente. No entanto, como a saída é zero para valores negativos, caso um neurônio ative com esse valor, pode-se considerar esse nó praticamente morto. Para superar esse inconveniente, há a variante Leaky ReLU. Além disso, existem outras estratégias na forma de operações ou camadas que podem compor uma rede, tais como blocos LSTM (do inglês, *long short-term memory*) e de atenção, camadas convolucionais, *instance* e *batch normalization*, *pooling*, *early stopping*, *dropout*, *data augmentation*, entre outros. Particularmente, esses três últimos citados compõem técnicas utilizadas para prevenir o sobre-ajuste (do inglês, *overfitting*) dos modelos no treinamento.

O aprendizado de uma rede acontece por meio de um processo iterativo de minimização de uma função de custo (do inglês, *loss function*). Isso deriva do fato que a relação entre a função objetivo e os parâmetros do modelo é não-linear, ou seja, não é possível obter uma solução fechada. Em uma situação de aprendizado supervisionado, ao ter disponíveis as respostas esperadas para todas as entradas, o modelo deve aprender a mapear tal entrada à sua resposta ou alvo correspondente. Durante o treinamento, os pesos da rede são atualizados conforme o erro quantificado numa operação denominada *back-propagation*. Ambos os problemas abordados neste trabalho são dessa categoria. As funções de custo são um modo de quantificar o erro

entre a saída do modelo e o valor real esperado. Tal qual em funções de ativação, diversos otimizadores podem ser utilizados para encontrar o mínimo da função de custo, como o *Stochastic Gradient Descent* (SGD) e o otimizador Adam (67). Para isso, em geral, se define uma taxa de aprendizado (do inglês, *learning rate*), o que pode não ser um ajuste trivial. Uma taxa muito baixa pode fazer com que o treinamento — a convergência — demande muitas iterações, eventualmente atingindo um valor ótimo, ou, ainda, que estacione em mínimos locais. Por outro lado, um *learning rate* alto pode fazer com que o modelo divirja. Além da definição de um valor de taxa de aprendizado inicial, é possível aplicar funções que alteram esse valor de taxa com o progresso do treinamento, o chamado *learning rate scheduler*. Essa técnica é benéfica, podendo auxiliar na convergência ao valor ótimo em tempo reduzido.

Em visão computacional, o emprego de CNNs tem se destacado, inclusive em segmentação de imagens médicas (68, 69, 70, 48, 71, 72). As redes neurais são consideradas CNNs quando são principalmente compostas por camadas convolucionais, ao invés de camadas totalmente conectadas. Ao contrário das *fully connected layers*, as quais em processamento de imagens requereriam uma quantidade de parâmetros que facilmente poderia extrapolar os recursos computacionais razoavelmente disponíveis, cada filtro de um bloco convolucional se restringe a um campo de receptores. Assim, os neurônios de uma camada se conectam apenas aos *pixels* ou *voxels* que estão sob esse campo. A partir dessa inspiração que advém dos campos de receptores visuais (73), compartilham de certa invariância espacial, isto é, possuem a capacidade de lidar com padrões a despeito de translações, reescala, rotações, etc. Além disso, uma sequência de camadas convolucionais, permite desde pequenos campos de receptores até grandes que auxiliam na extração de atributos de baixa a alta ordem.

Em segmentação semântica de imagens médicas, a arquitetura U-Net (13) é largamente utilizada. Essa *fully convolutional network* (FCN) e sua variante 3D (14) desempenham relativamente bem em problemas de imagens médicas com restrita disponibilidade de dados. A partir dessas arquiteturas muitas outras foram propostas, como a V-Net (74), que utiliza blocos residuais (75), o UNETR (17) e Swin UNETR (18), que empregam *transformers* (16, 76, 77), e o *framework* nnU-Net (15). Por fim, modelos fundacionais têm surgido. O Segment Anything Model (SAM) (78) é uma fer-

ramenta que realiza a segmentação de objetos a partir de um *prompt*, como um ponto ou um *bounding box* definido pelo usuário, mesmo que não tenha sido treinado com a *target* específica. Outras soluções baseadas nesse tipo de modelo surgiram, como o MedSAM (79), com objetivo de prover uma solução de segmentação dita “universal”. Essa classe de abordagem para segmentação não foi explorada neste trabalho.

O campo de aplicação de aprendizado de máquina na área de imagens médicas, principalmente em imagens de ressonância magnética, tem desafios particulares. Talvez, o mais relevante deles seja a baixa disponibilidade, em geral, de dados e, também, seus rótulos e anotações de segmentação com qualidade por especialistas experientes, seguindo minimamente um protocolo, convenção ou consenso. Além disso, há o impacto causado pela diversidade de fabricantes e modelos de equipamentos de imagem médica, acompanhado por distintos protocolos de aquisição com diferentes especificações de espaçamento, dimensões, resoluções, tempos e orientações.

Há um consenso que o padrão-ouro é a segmentação manual por um anotador experiente ou o consenso entre vários especialistas. Contudo, esse processo está sujeito às variações entre anotadores, a variabilidade *inter-rater*, especialmente no detalhamento de subdivisões anatômicas, quando presentes. Além disso, um mesmo anotador pode apresentar variações com suas próprias anotações — variabilidade *intra-rater* — a depender do momento e de fatores como a qualidade da imagem reproduzida por monitores e o ferramental usado para marcação manual. Ou seja, além de apresentar variações entre avaliadores, o delineamento manual tende a possuir baixa repetibilidade pelo mesmo operador. Por fim, a própria anotação em si pode ser uma tarefa desgastante e demorada, além de exigir um prévio treinamento. Logo, essas limitações motivaram o surgimento de métodos, semi ou totalmente automatizados, para segmentação de estruturas cerebrais, com as anotações manuais por especialistas consideradas o *ground truth* ou o padrão de referência.

Nos últimos anos, diversos desafios propostos pela comunidade científica tiveram como foco a segmentação em imagens médicas, dentre eles o Medical Segmentation Decathlon (80), o desafio Brain Tumor Segmentation (BraTS) (81, 82), os desafios anuais dos encontros da Radiological Society of North America (RSNA AI

Challenge)² e o Desafio de Inteligência Artificial da Jornada Paulista de Radiologia³. Nesses desafios, é notável a aplicação de *deep learning*, compreendendo diversas arquiteturas e técnicas. Recentemente, foi apresentado o TotalSegmentator (83), uma ferramenta capaz de realizar a segmentação de até 117⁴ classes em imagens de tomografia computadorizada (do inglês, *computed tomography*, CT). Desse modo, compreendendo uma técnica bem estabelecida na literatura recente e capaz de atender aos requisitos dos objetivos deste trabalho, emprega-se o uso de *deep learning* no desenvolvimento dos modelos de segmentação para os alvos anatômicos de interesse.

Contudo, como mostrado anteriormente, as soluções atuais apresentam problemas quanto à acurácia, reprodutibilidade e generalização, principalmente quando aplicado em *datasets* de imagens com atrofia cerebelar. Assim, a proposta deste trabalho versa sobre o aproveitamento da disponibilidade e o uso de conjuntos de dados representativos e de qualidade aliados a técnicas de aprendizado profundo para obtenção de soluções para segmentação automatizada do cerebelo e o núcleo denteado que apresentem desempenho superior aos modelos estado da arte.

²<https://www.rsna.org/annual-meeting/awards-recognition/ai-challenge>

³<https://www.spr.org.br/concursos-e-premiacoes/285/desafio-de-inteligencia-artificial-da-spr/sobre-o-desafio>, <https://www.kaggle.com/competitions/spr-head-ct-age-prediction-challenge>

⁴O artigo publicado na RSNA AI indica 104 classes, porém, no repositório (github.com/wasserth/TotalSegmentator), há menção de 117 classes principais.

4 Materiais e Métodos

No capítulo são apresentados os conjuntos de dados utilizados para os problemas de segmentação do cerebelo e do núcleo denteado. Em seguida, há uma breve apresentação das arquiteturas e *frameworks* de *deep learning* que foram utilizadas no decorrer do desenvolvimento do trabalho. Adicionalmente, são expostas as principais técnicas de aumento de dados adequadas ao domínio de neuroimagem, as métricas de avaliação e a análise estatística. Dado que ambos os desafios deste trabalho são de segmentação, as abordagens são similares, de tal modo que não há distinção de metodologia aplicada, a despeito de pré e pós-processamentos adotados, de modo particular, a cada problema.

Ressalta-se que um dos princípios adotados no projeto é o da abordagem centrada em dados (do inglês, *data-centric approach*)¹ (84). Ao contrário de concentrar-se preferencialmente em modelos e suas arquiteturas (do inglês, *model-centric approach*), é direcionado um esforço inicial para melhorar a qualidade do conjunto de dados de forma sistemática com a intenção final de incrementar a performance dos modelos treinados. Por melhora da qualidade dos dados, pode-se derivar diferentes ações, tais como a supervisão e curadoria das máscaras de segmentação alvo do aprendizado, a observância quanto a ruídos nos dados e a atenção à amostragem, garantindo uma adequada representação da variabilidade encontrada no mundo real. Isso tudo acontece sem que se deixe de lado a busca pelo aumento do tamanho do conjunto de dados, muito menos a exploração de arquiteturas de redes neurais. No entanto, diferencia-se ao não colocar uma dedicação em demasia na proposição e experimentação de alterações altamente sofisticadas e complexas nas arquiteturas de modelos. Portanto, ao invés de procurar obter o melhor modelo para os dados, procura-se conseguir o melhor conjunto de dados para alimentar o aprendizado.

¹<https://dcai.csail.mit.edu/lectures/data-centric-model-centric/>; <https://mitsloan.mit.edu/ideas-made-to-matter/why-its-time-data-centric-artificial-intelligence>

4.1 Dados

Para os dois problemas abordados neste trabalho, os conjuntos de dados foram distintos, principalmente devido à metodologia e sequência de imagens utilizadas. No caso do cerebelo, aquisições ponderadas em T1 foram coletadas. Já para o núcleo denteado, é o QSM que fornece o mapeamento quantitativo da estrutura anatômica. Contudo, em ambos, há o cuidado para obtenção de dados representativos, tanto em relação à variedade de centros de imageamento médico e de protocolos de aquisição, quanto à diversidade populacional dos indivíduos que foram escaneados. Isto é, incluindo imagens de pessoas saudáveis e pacientes de diferentes doenças neurodegenerativas.

Os dados foram armazenados preferencialmente em formato NIfTI. Alternativamente, foram recebidas imagens em formato DICOM, que foram convertidos para NIfTI usando a ferramenta `dcm2niix` (85). As imagens T1w foram pré-processadas para correção de inomogeneidade de campo com o algoritmo N4 (41) implementado no software ANTs. Quando necessário, registro entre imagens ou entre imagens e templates foram conduzidas com o ANTs. Caso este não fosse bem-sucedido, a ferramenta de registro do FSL foi adotada. Para template, foi determinado o MNI152 (86). A orientação preferencial foi a convenção neurológica, especificamente a RAS, na qual o eixo X (*left-right*) é positivo para direita (*right*, R) do paciente, o eixo Y (*posterior-anterior*) é positivo para parte anterior (*anterior*, A) e o eixo Z (*inferior-superior*) é positivo para parte superior (*superior*, S). Isso garante consistência e padronização nas visualizações e análises.

4.1.1 Definição da Casuística

As ataxias hereditárias são um grupo de doenças que compartilham a ataxia cerebelar como um achado clínico comum, porém achados não-motores e déficits cognitivos também podem estar presentes (87). Tal grupo é classificado de duas formas de herança autossômica dominante e recessiva. Dentre as formas dominantes, as mais prevalentes, principalmente no Brasil, estão a SCA3 (também conhecida como doença de Machado-Joseph), SCA1, SCA2 e SCA6. Já a ataxia de Friedreich é a ataxia de herança autossômica recessiva mais comum (88). Ambas as formas afetam

adultos e crianças, levando a severas deficiências e elevados gastos financeiros (89). Além disso, essas doenças não possuem tratamento aprovado, de modo que a elaboração de ensaios clínicos se torna extremamente essencial e necessária. Dessa forma, vários grupos têm trabalhado em possíveis agentes modificadores da doença, mas a falta de variáveis respostas sensíveis gera uma grande limitação na avaliação do real potencial dessas terapias. Os estudos atuais, de fato, usam escalas clínicas como variáveis resposta primárias, tais como a SARA (Scale for the Assessment and Rating of Ataxia) (90). Embora sejam instrumentos úteis para avaliar e quantificar a severidade da doença, são pouco sensíveis para detectar alterações longitudinais em um curto espaço de tempo (91). Sendo assim, esses estudos que empregam escalas clínicas necessitam recrutar um número elevado de pacientes, da ordem de 150 a 200 para cada braço do estudo. Considerando que ataxias hereditárias são raras e progridem lentamente, isso faz com que essas análises possuam uma séria limitação.

Patologicamente, as SCAs são caracterizadas por uma severa perda neuronal no cerebelo e tronco encefálico (92). Contudo, estudos histológicos mostram que entre as SCAs há diferentes padrões de danos cerebelares, por exemplo: pacientes com SCA6 possuem uma acentuada redução do córtex cerebelar com perdas importantes de células de Purkinje, já na ataxia de Friedreich ele é bem preservado, especialmente no início da doença. Por outro lado, o núcleo denteado do cerebelo está, agora, atrofiado em FRDA e preservado em SCA6 (30). Além disso, estudos de MRI recentes têm mostrado que a progressão no dano cerebelar em tais doenças se dá de forma diferente, mas em todas elas o *effect size* tem sido maior se comparado à evolução da gravidade da doença medida pela escala clínica (93, 94).

Portanto, as ataxias hereditárias foram definidas como um modelo adequado para a utilização de imagens de ressonância magnética no treinamento de modelos de segmentação automatizada devido à característica comum de atrofia cerebelar e, em alguns casos, alterações nos núcleos denteados em diversos graus. Além das ataxias hereditárias, outras patologias compartilham do achado de neurodegeneração, como a esclerose múltipla (do inglês, *multiple sclerosis*, MS) (95, 96) e a atrofia de múltiplos sistemas (do inglês, *multiple system atrophy*, MSA) (50, 97). Ao incluir também imagens de indivíduos saudáveis, pretende-se obter uma representação com variabilidade anatômica suficiente para generalizar os modelos de segmentação

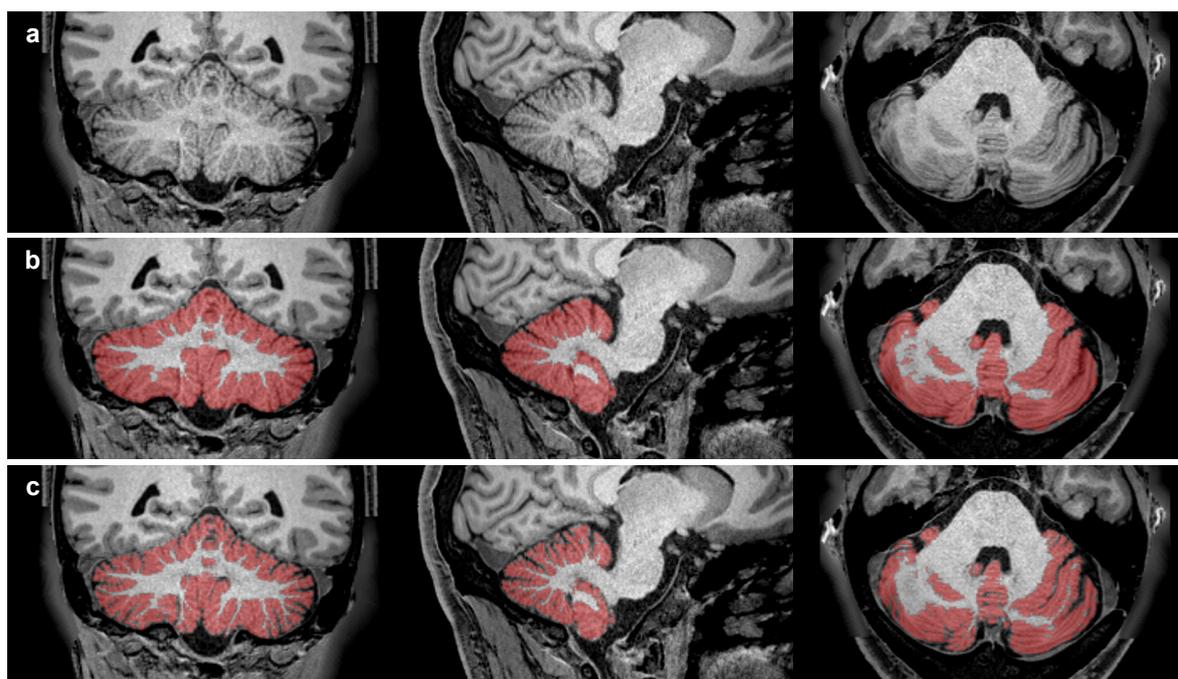


Figura 4 – Exemplo de imagem de um paciente com ataxia de Friedreich. (a) imagem ponderada em T1. (b) é o resultado da ferramenta CerebNet. (c) é o resultado da segmentação manual.

a diferentes contextos. Isso visa criar ferramentas robustas e precisas, capazes de detectar mudanças anatômicas em diversos níveis e atender às demandas das pesquisas em andamento, superando a performance dos métodos atuais.

4.1.2 Córtex Cerebelar

Inicialmente, foram utilizadas imagens T1w de estudos internos do Laboratório de Neuroimagem da Unicamp. Todas as sequências foram obtidas com uma ressonância Philips Achieva (Philips Healthcare, Países Baixos) 3 Tesla. Ao todo, o *dataset* era composto, aproximadamente, por 1000 aquisições de indivíduos saudáveis e 500 pacientes, todos com confirmação molecular para ataxia hereditária. Em um primeiro momento, um subconjunto de 607 imagens foi utilizado para um teste de adequabilidade do uso de técnicas de *deep learning* para o problema da segmentação do cerebelo. Isto é, na circunstância inicial do projeto, se queria avaliar se modelos de DL eram adequados. Para isso, segmentações “grosseiras” produzidas pelo software FreeSurfer (49), binarizadas para o cerebelo completo, considerando corpo medular e córtex como uma única classe, foram empregadas como anotação alvo.

Tabela 1 – Conjuntos de dados anotados manualmente por especialistas para segmentação do córtex cerebelar.

Grupo	Treino	Teste	Total (anotado)	Out-of-sample
Controle	9	3	12	35
FRDA	11	3	14	71
SCA1	26	5	31	-
SCA3	7	3	10	72
Total	53	14	67	178

FRDA: ataxia de Friedreich; SCA: ataxia espinocerebelar.

Conforme já apresentado, a construção de um *dataset* de alta qualidade depende, além de boas imagens, de máscaras manualmente anotadas por experientes colaboradores. Para isso, um grupo de neurorradiologistas delineou manualmente o cerebelo, com foco especial nas bordas, onde fissuras e invaginações do córtex demandam rigorosa atenção (Figura 4). Nessa tarefa, o *dataset* foi composto por 67 imagens de indivíduos controle e de pacientes com SCA3, SCA6 e FRDA. A Tabela 1 mostra as quantidades de imagens anotadas por grupo (controles e doenças), já com a divisão em treino e teste, além de um conjunto *out-of-sample* utilizado posteriormente para avaliação das redes treinadas.

Com objetivo de avaliar a reprodutibilidade, o *dataset* Kirby-21 (98) foi selecionado para teste-reteste. Ele consiste em um conjunto de imagens de 21 indivíduos saudáveis (11 homens e 10 mulheres) que foram escaneados duas vezes, após serem reposicionados, seguindo o mesmo protocolo, em um intervalo menor que um dia. Dado que o Kirby-21 inclui uma sequência MPRAGE (do inglês, *magnetization-prepared rapid gradient echo*) ponderada em T1, estas foram utilizadas para a análise. As imagens foram adquiridas em um equipamento Philips Achieva (Philips Healthcare, Países Baixos) de 3 Tesla e os dados demográficos e parâmetros de aquisição estão disponíveis no estudo relacionado (98).

4.1.3 Mapeamento Quantitativo de Suscetibilidade Magnética

O QSM é uma técnica que permite a obtenção de um mapeamento quantitativo de suscetibilidade magnética. A suscetibilidade magnética é uma propriedade

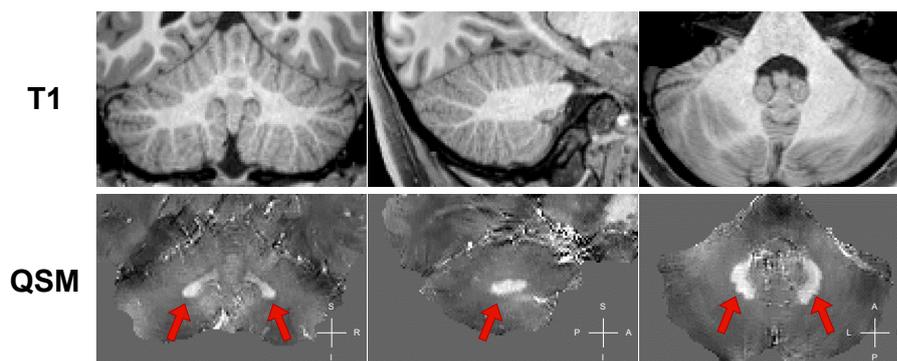


Figura 5 – Slice coronal de imagem ponderada em T1 e QSM de um mesmo indivíduo.

que mensura grau com que um material pode ser magnetizado na presença de um campo magnético externo, ou seja, quão magnetizável um material é. Ao contrário do SWI, que provê uma imagem qualitativa, o QSM possibilita a diferenciação de materiais paramagnéticos e diamagnéticos. Além de ser particularmente útil para a detecção e a caracterização de lesões cerebrais, como hemorragias e calcificações, é uma ferramenta ideal para a visualização dos núcleos denteados, dado o seu tecido rico em ferro. Logo, o DN é identificado em QSM por hipersinal na imagem, permitindo sua análise estrutural anatômica, ao contrário de uma imagem ponderada em T1, como exemplificado na Figura 5.

O QSM é obtido por um processamento aplicado sobre os dados de magnitude e fase, resultantes de aquisições *multi-echo gradient-recalled echo* (GRE) ou *echo planar imaging* (EPI). Dentre as etapas necessárias estão o *phase unwrapping*, *masking*, *background field removal* e a inversão de dipolo magnético, obtendo a reconstrução do mapeamento de susceptibilidade (Figura 6). Existem disponíveis diversas ferramentas que reúnem essas etapas em pipelines que facilitam a sua utilização. Entretanto, a ampla gama de métodos desenvolvidos para cada das fases, permite inúmeras combinações, além de disporem de diversos parâmetros que podem, e devem, ser ajustados para que forneçam resultados adequados. Dessa forma, é necessário contemplar, ainda que parcialmente, diferentes metodologias de reconstrução com o intuito de que o aprendizado das redes neurais abranja imagens de distintas qualidades.

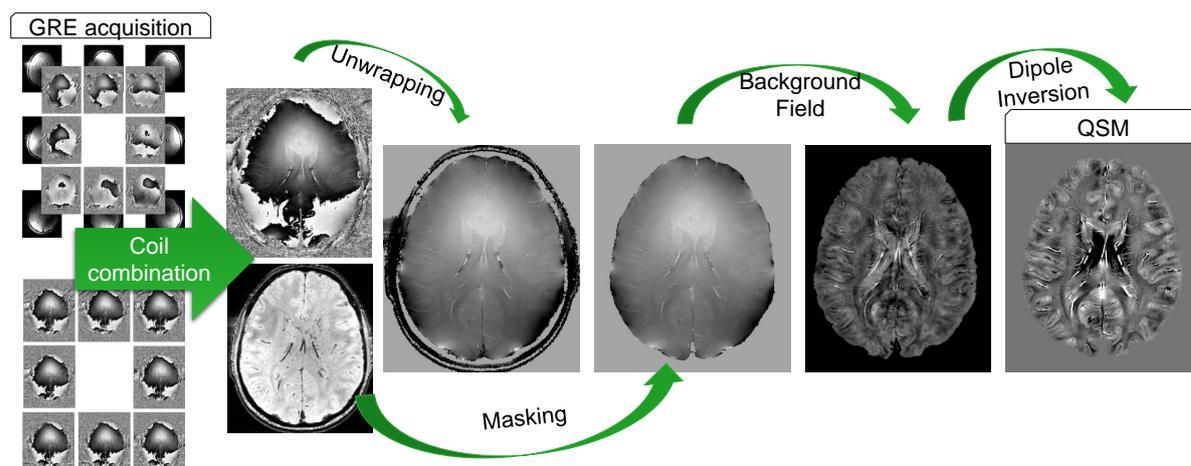


Figura 6 – Processo de reconstrução QSM incluindo *phase unwrapping*, *background field removal* e a inversão de dipolo. Adaptado de (99).

4.1.4 Núcleo Denteado

As aquisições *multi-echo gradient-recalled echo* foram conduzidas usando quatro diferentes protocolos e realizadas em dez centros de imagem ao redor do mundo com equipamentos de ressonância magnética de 3 Tesla de fabricação Philips ou Siemens. Os dados de pesquisa coletados estavam desidentificados já na origem, garantindo aderência a um *pipeline* de dados sem informações que permitissem identificação pessoal. O conjunto de dados incluiu um total de 132 indivíduos controle, 154 pacientes com FRDA e 16 com esclerose múltipla, como disposto na Tabela 3. Aquisições de *follow-up* após 12 meses estavam disponíveis para 55 desses participantes. As imagens de pacientes com FRDA e MS foram escolhidas para abranger a variabilidade anatômica no DN causada por neurodegeneração no curso das doenças. Os protocolos de aquisição foram publicados anteriormente (100, 101, 23) e estão resumidos na Tabela 2. Os centros identificados como Aachen, Campinas, Children’s Hospital of Philadelphia (CHoP), Melbourne, Flórida e Minnesota correspondem às unidades de pesquisa participantes do TRACK-FA Neuroimaging Consortium, compartilhando o mesmo protocolo de aquisição. Os dados demográficos são apresentados na Tabela 3.

Tabela 2 – Protocolos de aquisição para cada *dataset*.

Conjunto de dados	Equipamento de ressonância	Seqüência	TR (ms)	TE1 (ms)	Δ TE (ms)	Número de ecos	FoV (mm)	Matriz da imagem (voxels)	Tamanho do voxel (mm)	Tempo de aquisição
Aachen	3T Siemens Skyra	GRE	27	3,7	6	4	220 x 220 x 176	208 x 256 x 176	0,86 iso	7' 22"
Campinas	3T Siemens Prisma									
CHoP	3T Philips Ingenia									
Melbourne										
Flórida										
Minnesota										
Naples	3T Siemens Magnetom Trio	GRE	32	7,38	14,76	2	230 x 194 x 160	378 x 448 x 160	0,51 x 0,51 x 1,0	9' 53"
IMAGE-FRDA	3T Siemens Skyra	GRE	30	7,38	14,76	2	230 x 230 160 axial slices	232 x 256 x 160	0,90 iso	11' 30"
INFLAM-FRDA	3T Siemens Biograph	GRE	31	5,70	5,27	5	230 x 230 104 axial slices	384 x 384 x 104	0,60 x 0,60 x 1,2	6' 50"

GRE: *gradient recalled echo*; TR: tempo de repetição; TE: tempo de eco; TE1: primeiro tempo de eco; FoV: *field of view*; T: Tesla; iso: isotrópico.

Tabela 3 – Dados demográficos dos participantes para cada *dataset*.

	Aachen, Campinas, CHoP, Melbourne, Flórida, Minnesota		Naples			IMAGE-FRDA		INFLAM-FRDA	
	Controle	FRDA	Controle	FRDA	MS	Controle	FRDA	Controle	FRDA
Participantes	44	98	48	12	16	31	30	9	14
Crianças (<18 anos)	14	31	0	3	0	0	0	0	0
Idade (média±SD)	23,6±7,7	23,9±8,7	37,8±12,7	32,6±15,8	46,8±5,6	37,6±13,1	35,7±12,2	28,6±5,8	27,8±7,6
Idade (intervalo)	[11, 41]	[12, 42]	[19, 64]	[14, 62]	[35, 57]	[19, 62]	[18, 56]	[19, 40]	[18, 42]
Sexo (M/F)	24/20	49/49	19/29	4/8	5/11	16/15	17/13	4/5	9/5

FRDA: ataxia de Friedreich; MS: esclerose múltipla; SD: desvio padrão; M: masculino; F: feminino.

Uma imagem QSM de amostra para cada *dataset* é apresentada na Figura 7. Na Figura 8, tem-se um fluxograma dos dados utilizados. As imagens QSM foram reconstruídas através dos *toolboxes* JHU/KKI (102, 35, 103) e STI Suite², utilizado *Laplacian unwrapping* para *phase aliasing*, V-SHARP (104) para a remoção do *background field* e tanto MEDI (105, 106) quanto iLSQR (107) para reconstrução do mapeamento de suscetibilidade.

Para o treinamento dos modelos de *deep learning*, os dados foram divididos em conjuntos de treino, validação e teste na proporção 70/10/20%, respectivamente. Foi aplicada uma estratégia estratificada de amostragem, de modo que os conjuntos criados mantivessem a proporção relativa das distribuições por centro de aquisição. Além disso, caso um indivíduo possuísse mais de uma imagem adquirida, estas foram alocadas a um único subconjunto, minimizando a possibilidade de vazamento de dados.

O comitê de ética ou *institutional review board* (IRB) de cada projeto ou centro de pesquisa originador dos dados compartilhados aprovaram o uso, ou apresentaram uma dispensa ética para este estudo retrospectivo. Todos os participantes forneceram consentimento por escrito previamente à aquisição das imagens. O TRACK-FA *steering committee* aprovou o uso dos dados, e os números de registro dos IRB foram previamente publicados (100) (Monash Health Human Research Ethics Committee: RES-20-0000-139A; Children’s Hospital of Philadelphia: IRB 20-017611; University of Minnesota: IRB STUDY00009047; University of Florida: IRB202000399; RWTH Aachen University: EK195/20; University of Campinas CAAE: 83241318.3.1001.5404;

²<https://people.eecs.berkeley.edu/~chunlei.liu/software.html>

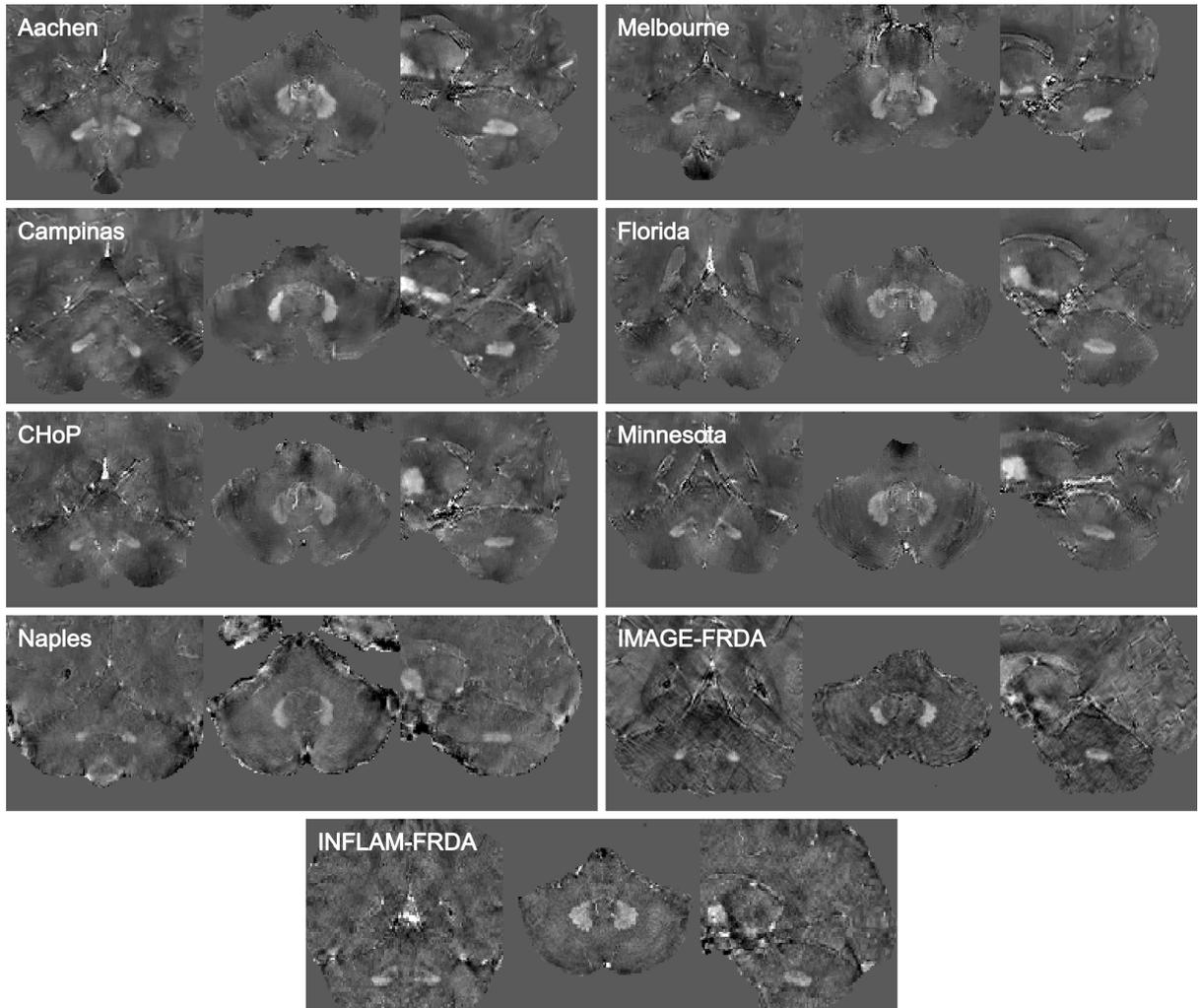


Figura 7 – Amostras de imagens QSM de cada *dataset*. Selecionados aleatoriamente. As imagens foram recortadas centralizadas no cerebelo. Intensidades dos *voxels* foram normalizados utilizando *z-score* para melhorar a visualização.

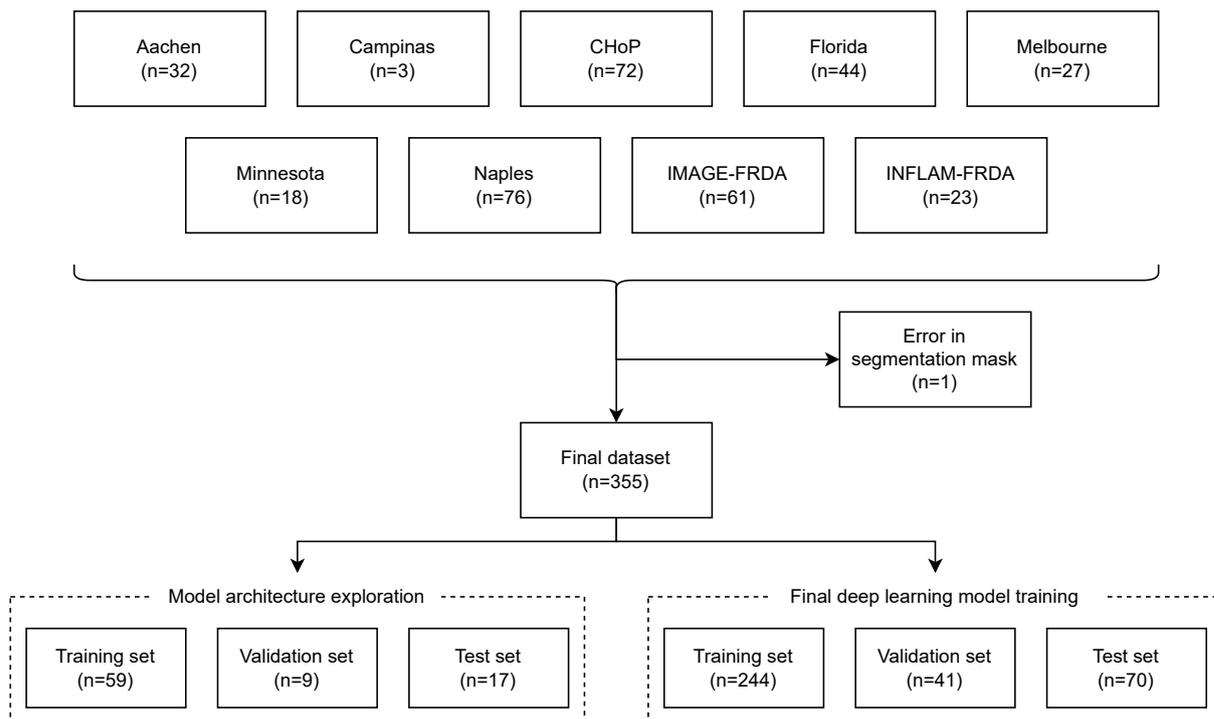


Figura 8 – Fluxograma dos dados utilizados no estudo.

McGill IRB Approved Project Number: 2022-8676). Para os demais estudos, a aprovação ética foi obtida independentemente: Ethical Committee “Carlo Romano” da University of Naples “Federico II” (Naples A: 209/13, Naples B: 47/15), Monash University Human Research Ethics Committee (IMAGE-FRDA: 13201B, INFLAM-FRDA: 7810) e University of Minnesota IRB (1210M22281).

Anotações Manuais

Os DN foram segmentados manualmente por especialistas em cada imagem QSM para compor o conjunto de dados *ground truth*. Foram utilizados os softwares de visualização e anotação MRICron³, ITK-SNAP (108) ou FSLeys (109), dependendo da preferência de cada anotador. Os níveis de contraste foram padronizados para visualização em todas as imagens, de -0,1 a 0,2 ppm de suscetibilidade magnética, com intuito de minimizar o viés na detecção de bordas pelos anotadores e entre eles.

As segmentações incluíram as regiões tridimensionais completas de hiperintensidade, incluindo a substância cinzenta externa (*ribbon*) e a hilum central de

³<https://www.nitrc.org/projects/mricron>

substância branca do DN. Para delinear, foram realizadas segmentações nos três planos (axial, coronal e sagital), para garantir que fosse gerada uma superfície regular, excluindo demais núcleos vizinhos, como o emboliforme. Para cada imagem, o DN esquerdo (LDN) e o direito (RDN) foram segmentados individualmente.

Cada imagem QSM foi segmentada manualmente por um dos três experientes anotadores (SS, SC e IHH), cego às patologias subjacentes ao indivíduo correspondente. Avaliações intra e inter-avaliador foram conduzidas a partir de um conjunto selecionado aleatoriamente de nove imagens, estratificada por centro de aquisição. Essas imagens foram duplicadas, embaralhadas aleatoriamente, tendo seus nomes trocados por identificadores genéricos, de modo que ao avaliador fosse dificultada a identificação da ordem e da repetição. Todos os três anotadores geraram máscaras para as 18 imagens. Com esses resultados, as seguintes métricas intra e inter-avaliador foram calculadas: coeficiente Dice, distância Hausdorff e o coeficiente de correlação intraclass (do inglês, *intraclass correlation coefficient*, ICC) (110).

Validação Externa

Para avaliar o desempenho do modelo em dados não utilizados em treinamento, 21 imagens de dez participantes foram adquiridas de três centros de imagem: n=4 indivíduos saudáveis do Instituto Neurologico “Carlo Besta”, Itália; n=2 pacientes com MS de Medical University of Graz, Áustria; e n=3 indivíduos controle e n=1 paciente com FRDA da McGill University, Estados Unidos. Todos esses locais de aquisição não forneceram dados para o treinamento dos modelos, portanto são considerados um conjunto de validação externa. Os mapas QSM foram obtidos com os *pipelines* JHU/KKI QSM v3.0 (102, 35, 103) e SEPIA v1.2.2.6 (111), resultando em duas imagens por participante das amostras de Carlo Besta e McGill. Três participantes de Carlo Besta foram escaneados novamente em um protocolo anisotrópico e processado com o *pipeline* do SEPIA, apenas. Um dos anotadores experientes (SS) segmentou manualmente o DN esquerdo e direito, para fornecer o *ground truth*, seguindo as mesmas instruções para os dados de treinamento. Em seguida, as imagens foram processadas pelo modelo treinado e pelo MRICloud, a fim de comparar seus desempenhos. Para o *toolbox* SEPIA, foi utilizado o 3D *best path* (112) para *phase unwrapping*, seguido por MEDI *non-linear fit*, Laplacian Boundary Value (LBV)

(113) para *background magnetic field removal* e STreaking Artifact Reduction for QSM (Star-QSM) (114) para obtenção dos mapas de suscetibilidade quantitativos.

Os comitês de éticas de cada instituição aprovaram os estudos: Fondazione IRCCS Istituto Neurologico “Carlo Besta”: 42/2017 07/06/2017; comitê de ética local de Medical University of Graz: 31-432ex18/191264-2019.

Processamento no MRICloud

Como apresentado na seção 3.2.1, o MRICloud requer como entrada imagens QSM e T1w co-registradas no formato ANALYZE. O *skull-stripping* da imagem T1w foi feito com o Brain Extraction Tool (BET) do FSL (115) e o registro com o ANTs (SyNQuick). O FSL (115) foi utilizado para conversões entre o formato NIfTI e ANALYZE. As execuções do *pipeline* MRICloud ocorreram entre 5 e 7 de janeiro de 2023 para o conjunto de dados principal e entre 9 e 31 de março de 2024 para o *dataset* de validação externa.

4.1.5 Arquiteturas de Deep Learning

Uma extensa variedade de arquiteturas de redes neurais profundas foram propostas para a segmentação em imagens médicas. Dentre elas, neste trabalho foram analisados desde a mais popular, a U-Net, incluindo uma variação com *deep supervision* (DS) (116), até um *framework* totalmente automatizado, o nnU-Net (15), passando por arquiteturas mais recentes envolvendo Transformers (77) e *Vision Transformers* (ViT) (16).

U-Net

A U-Net (13) é um modelo cuja estrutura simétrica possui uma parte de contração (do inglês, *contraction*) ou codificação (do inglês, *encoding*) e outra de expansão (do inglês, *expansion*) ou decodificação (do inglês, *decoding*). Essa é uma arquitetura amplamente utilizada em segmentação semântica, sendo base para outras derivações que surgiram posteriormente. Originalmente, as imagens tratadas pela U-Net são em duas dimensões. A necessidade de segmentação em imagens em três dimensões culminou em algumas abordagens. Uma delas é o treinamento

de três redes 2D distintas, uma em cada plano (axial, coronal e sagital, no caso de imagens médicas 3D), concluindo em uma votação ou consenso entre os três mapas finais de probabilidade. Outra solução é o advento de redes 2.5D (48), em que um conjunto de *slices* vizinhos ao que se deseja obter a segmentação são imputados ao modelo. Desse modo, características de contexto são consideradas na segmentação, também terminada em votação das três redes treinadas. Uma evolução natural, por fim, é a U-Net 3D (14), a qual aceita diretamente matrizes 3D. Nesse caso, a demanda por recursos computacionais é mais elevada, dado que a quantidade de parâmetros nessa arquitetura é maior em comparação com as soluções anteriores.

A U-Net 3D é uma proposta baseada na arquitetura de Ronneberger et al. (13), adaptada para imagens 3D. Na parte de contração, cada nível é formado por uma sequência de duas convoluções, *batch normalization*, função de ativação ReLU e uma operação de *max pooling* que reduz a resolução. A cada nível, a quantidade de canais é dobrada. Nessa parte, os *feature maps* vão sendo reduzidos, menores a cada passo, aumentando o campo de receptores. Já na expansão, há a substituição do *max pooling* pela convolução transposta (ou deconvolução, ou *upconvolution*) que dobra a resolução, recuperando-a a cada nível. Além disso, incorporam-se *feature maps* — de mesma resolução — intermediários da parte de codificação através de *skip connections*, provendo *features* de alta resolução. Ao final, a saída desse modelo é um conjunto de mapas de probabilidades de cada classe com a mesma resolução da imagem de entrada.

O *deep supervision* (116) é uma técnica utilizada para auxiliar o processo de treinamento. Em uma situação convencional, apenas a saída final da última camada de uma rede neural é avaliada. Já na supervisão profunda, são avaliadas as saídas em várias camadas intermediárias. Essas camadas de saída são treinadas para prever o mapa de segmentação final e são utilizadas de modo a guiar o treinamento da rede, inclusive em estágios iniciais do aprendizado. Essa supervisão auxiliar é contabilizada na função de custo, fazendo com que na operação de *back-propagation*, o fluxo de gradiente propicie um treinamento mais eficiente, por exemplo, nas camadas iniciais da rede. O emprego do *deep supervision* auxilia a resolver o problema do *vanishing gradient*, especialmente em redes muito profundas, e a melhorar a robustez do aprendizado dos extratores de atributos da rede. No caso da U-Net, em

cada nível da parte de expansão é avaliada uma saída adicional do bloco convolucional. Com isso, induz-se que as camadas intermediárias também produzam mapas de segmentação significativos e condizentes com os *targets* avaliados. Para isso, os mapas de segmentação alvo são subamostrados para serem compatíveis e comparáveis com as saídas dos níveis da parte de decodificação da U-Net. Alternativamente, as saídas intermediárias podem ser reamostradas para que essas previsões atinjam a mesma dimensão do *target* do dado de entrada. Logo, a função de perda pode ser calculada agregando-se também os resultados das camadas intermediárias.

Swin UNETR

Ao passo que redes convolucionais foram amplamente empregadas em visão computacional, em processamento de linguagem natural (do inglês, *natural language processing*, NLP) a introdução de *Transformers* (77) representou uma mudança significativa. A partir disso, foram propostas adaptações ao domínio de imagens, como o Vision Transformers (ViT) (16) e o Swin Transformer (76). Isso permitiu o desenvolvimento de novas arquiteturas incorporando essas novas abordagens no contexto de imagens médicas: o UNETR (17) e o Swin UNETR (18).

O Shifted Windows UNETR Transformers (Swin UNETR) é um modelo baseado em Swin Transformer proposto para segmentação em imagens médicas. O Swin Transformer implementa um *Transformer* de visão hierárquico que computa as operações de *self-attention* localmente em janelas não sobrepostas, permitindo que o modelo capture dependências de longo alcance (do inglês, *long-range dependencies*). Na estrutura hierárquica, *patches* são agregados *layer a layer*, reduzindo a dimensão dos mapas de *features* a cada etapa. Em vez de uma abordagem de *self-attention* global, como em ViT, agora a aplicação é local. Para não restringir o *self-attention* apenas a cada janela e preservar o poder representativo, o *cyclic-shifting* implementa a abordagem de particionamento por janelas deslocadas e introduz importantes conexões entre as janelas, permitindo sobreposições com janelas anteriores, enquanto o número de *patches* permanece fixo. Essa propriedade explica a complexidade computacional linear em relação ao tamanho da imagem, ao contrário da complexidade quadrática em outros modelos baseados em *transformers*, como os Vision Transformers, que apresentariam um problema de falta de escalabilidade para segmentação

semântica, uma tarefa de predição densa ao nível do *pixel* (2D) ou *voxel* (3D). O Swin Transformer produz mapas de *features* hierárquicos para obter representações de características em vários níveis de resolução, ou escalas, tornando-o adequado na função de rede *backbone* para tarefas de visão computacional. No Swin UNETR, o Swin transformer é aplicado como o *encoder* e é integrado com uma rede convolucional como *decoder* através de *skip connections* em uma arquitetura em formato de U.

nnU-Net

O nnU-Net é um *framework* de aprendizado profundo para imagens médicas que apresenta configuração automática (15). Esse modelo compreende um *pipeline* completo de treinamento para segmentação, tendo como principal característica a adaptação automática às diferentes resoluções, categorias de exames, sequências, espaçamentos, etc. Entre os diferenciais, está a estratégia de inicialização *data-driven*, que permite uma adaptação da estratégia de treinamento aos dados de entrada, resultando em uma performance otimizada. Além disso, os autores exaltam que a nnU-Net não compreende uma arquitetura inovadora, mas alcançam resultados estados da arte utilizando uma U-Net com algumas adaptações. Entretanto, a ferramenta se beneficia da aplicação de várias técnicas de *augmentation* dos dados no treinamento e possibilita treinamento com diferentes configurações de redes, como 2D, 3D *full resolution* e 3D *cascade* (uma combinação de um modelo treinado em imagens sub-amostradas seguido de um modelo de refinamento com resolução completa). Adicionalmente, o nnU-Net pode aplicar pós-processamento, se isso refletir em melhora das métricas, por exemplo, o “non-largest component suppression”. Por fim, ajustam-se automaticamente hiperparâmetros, compatibilizando o *framework* com a disponibilidade de recursos, como a memória da GPU, para treinamento. Esse *framework* tem demonstrado desempenho no estado da arte em diversos desafios de imagens médicas (15), levando a uma adoção abrangente pela comunidade.

4.2 *Data Augmentation*

Em neuroimagem, principalmente pareados com máscaras anotadas manualmente, a quantidade de dados de qualidade disponíveis é um recurso escasso. Por outro lado, os modelos de *deep learning* dependem intensamente de dados para que ocorra o aprendizado de modo suficiente, com intuito de atingir uma capacidade de generalização suficiente para seu uso efetivo. Como forma de contornar esse problema, existem diversos métodos de aumento de dados (do inglês, *data augmentation*) que auxiliam na geração artificial de exemplos que podem ser apresentados ao modelo e contribuir para o treinamento, aumentando a variabilidade. Um cuidado é que os métodos devem gerar novos exemplos baseados nos dados disponíveis que sejam minimamente verossímeis. Assim, é necessário escolher transformações e parâmetros adequados.

Apesar da variedade ampla de métodos aplicáveis a imagens, deve-se selecionar e configurar aqueles que gerem novos exemplos plausíveis. Por exemplo, no caso de uma imagem tridimensional da região da cabeça, ao escolher o *flipping*, deve-se determinar em qual eixo a transformação será aplicada. No entanto, apenas um espelhamento no eixo sagital, isto é, no eixo X, supondo orientação RAS. Além dessa técnica são aplicáveis também a rotação, a escala, a deformação elástica, a reescala ou *shift* de intensidade dos *voxels*. Algumas outras técnicas visam simular artefatos de MRI, como a movimentação, ruído e inomogeneidade de campo. Todas essas operações são aplicadas em treinamento de modo aleatório com probabilidade definida como hiperparâmetro. No desenvolvimento dos modelos foram utilizadas as implementações dos pacotes TorchIO (117) e MONAI (118).

4.3 Pipeline de Segmentação

Para ambos os problemas abordados neste trabalho, foi adotada uma abordagem de duas etapas. Inicialmente, a imagem de entrada é pré-processada para um *voxel spacing* comum. Na primeira etapa, um modelo de aprendizado profundo realiza uma localização do órgão primário de interesse. Essa primeira rede neural tem como objetivo a simples localização, isto é, não se deseja segmentar o cerebelo ou DN de

modo preciso. Dado haver uma possibilidade de recursos — imagens e ferramentas automatizadas — para obtenção de uma segmentação do cerebelo, ainda que grosseira, em imagens T1, redes 3D U-Net foram treinadas para isso. Ainda que fosse uma tarefa de detecção de objetos, utilizou-se uma arquitetura desenvolvida para segmentação semântica, adaptando-se a saída do modelo para o objetivo desejado. Logo, a arquitetura pode ser mais rasa ou, equivalentemente, menos profunda, sendo sua execução, portanto, mais rápida.

Após essa primeira etapa, com a estrutura de interesse localizada, é realizado um corte ao redor desse volume. Para isso é definido um *bounding box* de dimensões pré-estabelecidas, que compreende a totalidade do órgão alvo, mantendo uma tolerância para que haja preservação de contexto espacial suficiente nas bordas. Com isso, espera-se reduzir dois potenciais problemas, o custo computacional em processar uma imagem maior, completa, e falsos positivos.

Finalmente, o modelo de segmentação realiza a predição da máscara tendo como entrada a imagem recortada, o *bounding box* em torno do alvo. Por último, deve-se retornar a máscara predita para as dimensões originais da imagem, realizando a operação inversa de recorte do *bounding box*. Além disso, é feita a recuperação ao *voxel spacing* original. Assim, é possível ter o correto alinhamento entre a imagem original de entrada e a máscara predita pelo modelo.

Exemplificando, no caso da segmentação do DN (Figura 9), a imagem QSM de entrada é reamostrada para um *voxel spacing* comum de 0,86 mm isotrópico. O modelo de localização foi treinado de modo a identificar o cerebelo diretamente na imagem de QSM. A partir dessa *target* binária foi obtida com uso da ferramenta ACAPULCO (33) em imagens ponderadas em T1 originárias do mesmo indivíduo. Após a predição com o ACAPULCO, as máscaras de segmentação do cerebelo foram transformadas pela mesma matriz resultante da operação de registro da imagem T1 (*moving*) na QSM (*fixed*), obtendo alinhamento espacial. Os registros foram executados com software ANTs (119). Com o cerebelo identificado, o centroide é utilizado para alinhar o *bounding box* ao redor do cerebelo. Dado que o DN é uma estrutura pequena no centro da substância cinzenta em cada hemisfério cerebelar, há baixa possibilidade de truncar ou perder parte da estrutura, preservando as referências contextuais anatômicas. Após a segmentação de cada DN, a última etapa garante o retorno da máscara

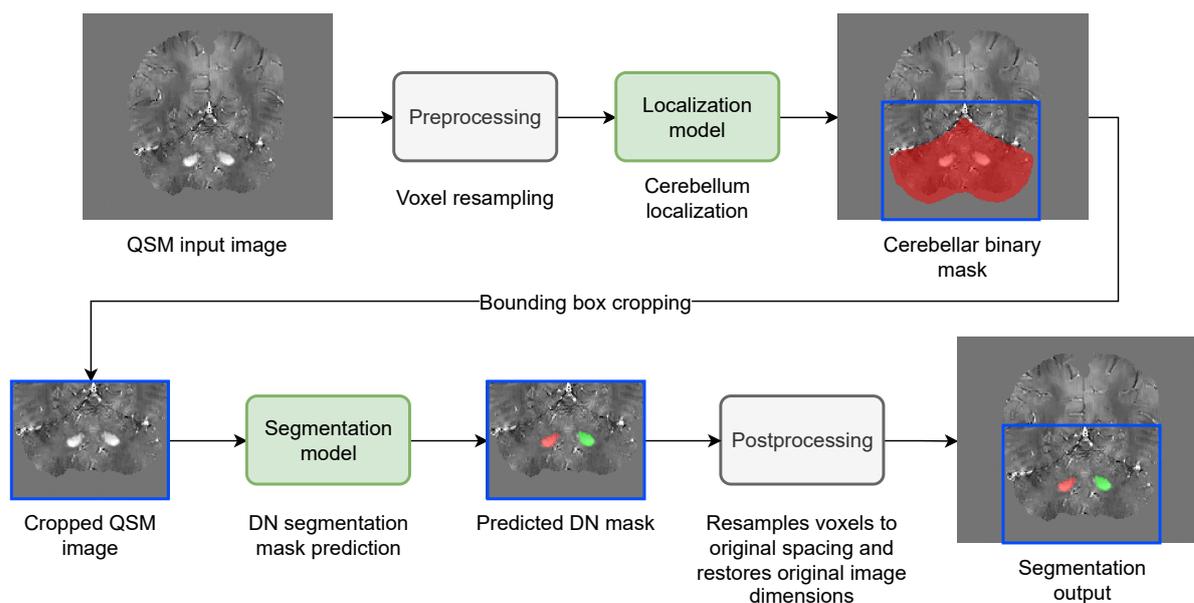


Figura 9 – Pipeline do modelo de segmentação do núcleo denteado.

predita para o espaço e *spacing* originais.

No caso da segmentação do núcleo denteado, inicialmente em experimentações utilizando as imagens em tamanho original, foram constatadas, no conjunto de teste, predições com presença de falsos positivos. Ao processar uma imagem QSM englobando a cabeça completa pela rede neural, abre-se margem para mais atribuições indesejadas de probabilidades aos *voxels* não relacionados ao denteado. Dentre esses, foram notados falsos positivos nos gânglios da base, tais como o núcleo caudado e o globo pálido. Em comum com o núcleo denteado, está o acúmulo de ferro (120) devido a fatores como idade e doenças neurodegenerativas. Assim, viu-se producente restringir espacialmente a imagem apresentada para o modelo de segmentação, de modo a evitar tais erros de classificação. Com isso, uma rede de localização foi treinada para segmentar de modo grosseiro o cerebelo em imagem QSM e, com isso, delimitar um centroide, sob o qual é definido um *bounding box* de tamanho fixo que englobe o cerebelo e mantenha contexto espacial suficiente em torno dos núcleos denteados. Além da vantagem de redução da propensão de falsos positivos, agrega-se menor custo computacional para o modelo de segmentação, uma vez que as dimensões da matriz da imagem de entrada são, agora, menores.

4.4 Métricas e Análise Estatística

Analisando as métricas de segmentação, pode-se avaliar por *voxel*. Um verdadeiro positivo (do inglês, *true positive*, TP) ocorre quando um *voxel* é corretamente classificado, ou seja, o modelo acerta a atribuição da classe de interesse. Um verdadeiro negativo (do inglês, *true negative*, TN) acontece quando é predito que o *voxel* não pertence à classe procurada. Já um falso positivo (do inglês, *false positive*, FP) se estabelece na atribuição incorreta de *label* sobre um *voxel*. Por fim, um falso negativo (do inglês, *false negative*, FN) é a situação na qual o modelo não atribui a classe em um *voxel* que pertence ao tecido de interesse.

O coeficiente de similaridade Sørensen-Dice (DSC) (121, 122), também conhecido como coeficiente Dice, é uma métrica amplamente utilizada para avaliar tarefas de segmentação. O coeficiente Dice mede a similaridade através da sobreposição entre duas máscaras de segmentação binárias e considera a concordância espacial (123). Está definido na Equação 4.1.

$$Dice(G, P) = \frac{2|G \cap P|}{|G| + |P|} = \frac{2TP}{2TP + FP + FN} \quad (4.1)$$

Onde $|A|$ é a cardinalidade do conjunto A , G e P representam a máscara de segmentação *ground truth* e a predição do modelo, respectivamente. Esta métrica varia de 0 a 1, com um valor de 1 indicando uma concordância perfeita. O coeficiente Dice foi escolhido como a principal métrica de desempenho e também fez parte da métrica de perda para o treinamento do modelo na forma de perda Dice ($Dice\ loss = 1 - Dice$).

Adicionalmente, outras métricas de segmentação foram avaliadas. A distância de Hausdorff (HD) (124) e a distância média de Hausdorff (AHD) (125) são relevantes para a avaliação da qualidade das bordas, sendo classificadas como métricas de distância espacial. Enquanto a HD corresponde ao maior das menores distâncias de um ponto em um conjunto para o ponto mais próximo no outro conjunto, a AHD calcula a distância média dos pontos de um conjunto para os pontos mais próximos no outro conjunto, em vez de considerar apenas a distância máxima como na HD. Desse modo, a última está menos sujeita à *outliers* do que a primeira. Também foi avaliada a Interseção sobre União (IoU) ou índice de Jaccard (126), que é semelhante ao Dice. Finalmente, a similaridade de volume (125) foi avaliada para medir a concordância

quantitativa de volume, ignorando informações de forma, posição e qualidade.

Para comparar conjuntos de medidas, utilizou-se o teste não-paramétrico dos postos sinalizados de Wilcoxon. Caso os dados apresentassem distribuição normal, foi aplicado teste t pareado. O teste de Kolmogorov-Smirnov foi empregado para avaliar a normalidade em variáveis contínuas. O coeficiente de correlação de Pearson foi escolhido para avaliar relações entre variáveis, ou Spearman em caso de não normalidade. Mensurou-se concordância das medidas de vários avaliadores usando o coeficiente de correlação intraclassa (ICC) (110) e os resultados foram avaliados (ICC < 0.5: pobre, $0.5 \leq \text{ICC} < 0.75$: moderado, $0.75 \leq \text{ICC} < 0.9$: bom, e $\text{ICC} \geq 0.9$: excelente confiabilidade) (127). Os intervalos de confiança do ICC foram calculados para fornecer um intervalo dentro do qual o ICC provavelmente se encontra, tipicamente com 95% de confiança. A análise estatística foi conduzida utilizando Python 3.10.8, R 4.1.1 e SPSS 27.0.1.0. Por fim, p-valores inferiores a 0,05 foram considerados estatisticamente significativos.

4.5 Disponibilização dos Modelos

A disponibilização dos modelos em repositórios abertos é um primeiro passo para um amplo uso. A fim de minimizar requisitos e maximizar a compatibilidade, fez-se a opção pela disponibilização do *pipeline* em imagens Docker⁴ e Singularity⁵. Além disso, com intuito de gerar uma interface ao usuário facilitada, houve a escolha do *framework open-source* Streamlit⁶. Este pacote em Python permite a criação de aplicativos web interativos. Logo, sendo um *container* um pacote de software que inclui todo o necessário para executar a aplicação, ao usuário é requerido apenas o Docker ou Singularity instalado e toda interação será pela interface web. Dessa forma, o usuário final tem uma experiência na qual está, em geral, familiarizado e com todo processamento podendo ser realizado localmente.

⁴<https://www.docker.com/>

⁵<https://sylabs.io/singularity/>

⁶<https://streamlit.io/>

5 Resultados

Neste capítulo são apresentados os resultados para os dois objetivos de segmentação. Para a vertente do cerebelo, são dispostos os resultados iniciais, dado que se encontra em andamento e será finalizado sob o escopo do projeto de pesquisa original e na vigência do *grant* da Friedreich's Ataxia Research Alliance (FARA). Já para a segmentação do núcleo denteado, os resultados são apresentados na forma de artigo submetido.

5.1 Segmentação do Cerebelo

O modelo U-Net *baseline* treinado com poucos dados alcançou uma métrica Dice de 0,921 no conjunto de validação. Já no conjunto de teste de 14 imagens não utilizadas para treinamento, o modelo treinado apresentou métrica Dice de 0.928 ± 0.030 , enquanto o CERES e o ACAPULCO obtiveram 0.918 ± 0.039 e 0.909 ± 0.043 , respectivamente (Figura 10). Em pacientes com grau elevado de dano cerebelar, essa diferença em performance é sustentada com Dice de 0.924 ± 0.033 para o modelo proposto, 0.917 ± 0.044 para o CERES e 0.904 ± 0.047 para o ACAPULCO. Em seguida, foram conduzidos testes t pareados em um conjunto de imagens *out-of-sample*, resultando em diferenças estatisticamente significativas entre o modelo treinado e o CERES ($p < 0,01$) e o ACAPULCO ($p < 0,001$) em amostras de indivíduos controle e, novamente, em imagens de pacientes com SCA3 e FRDA ($p < 0,001$ em ambos os casos) para o volume mensurado pelas segmentações (Figura 11). Em inspeção visual qualitativa, pode-se notar uma melhor precisão das máscaras produzidas pela U-Net treinada, especialmente no correto delineamento de fissuras cerebelares (Figura 12).

Uma avaliação no conjunto de dados de teste-reteste Kirby-21 foi conduzida

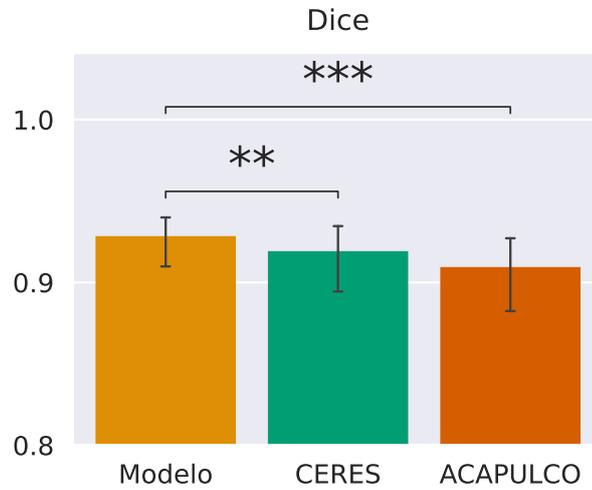


Figura 10 – Métrica Dice no conjunto de teste. As barras de erro indicam intervalo de confiança de 95%. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

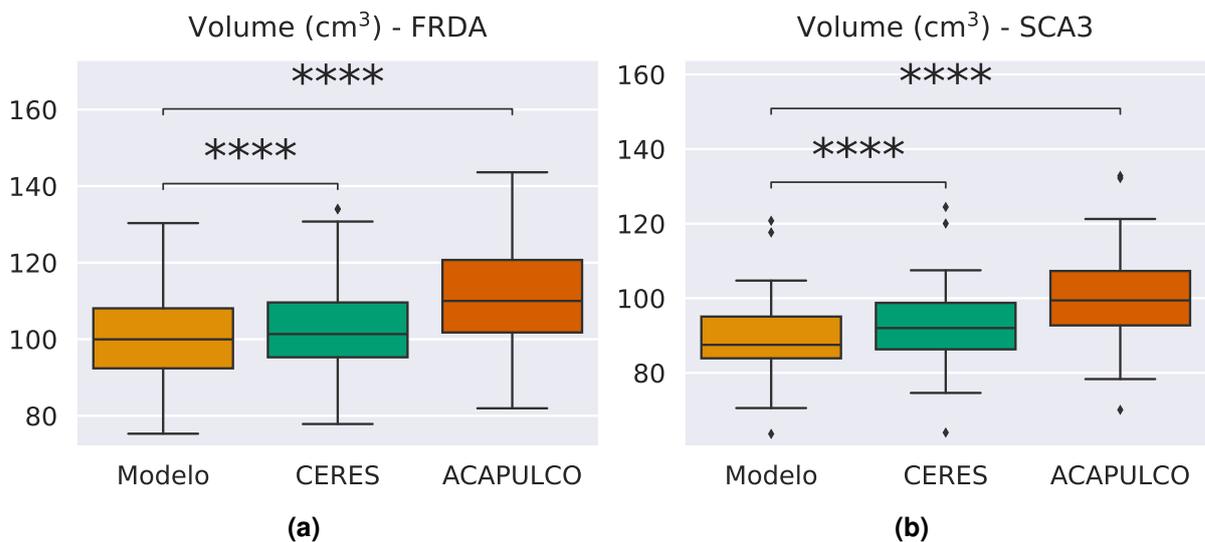


Figura 11 – Volume em cm^3 para amostras *out-of-sample* em pacientes com FRDA (a) e SCA3 (b). As barras de erro indicam intervalo de confiança de 95%. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

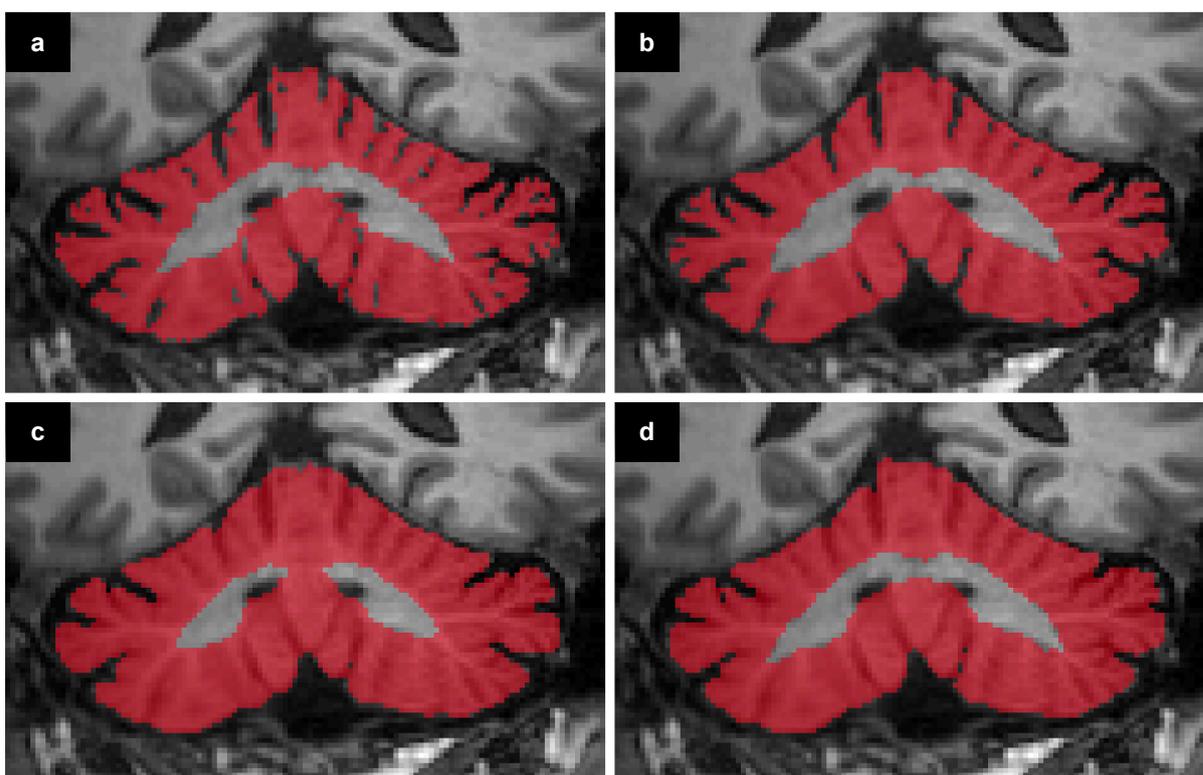


Figura 12 – Comparação de segmentações de um slice coronal para um paciente com ataxia espinocerebelar tipo 1. (a) é a máscara de segmentação produzida por um anotador experiente, o *ground truth*. (b) é o resultado do modelo treinado. Em (c) e (d) têm-se os resultados do ACAPULCO e CERES, respectivamente. Nota-se uma maior similaridade qualitativa entre o *ground truth* e a máscara predita pelo modelo U-Net. Já ACAPULCO e CERES apresentam problemas de sobre-segmentação. Essa imagem de ressonância magnética não fez parte do treinamento.

e os resultados estão na Tabela 4. Note que para essa avaliação foi adicionado um novo modelo, o nnU-Net, resultado de um teste pontual para análise de performance frente ao modelo U-Net padrão. Ambos os modelos treinados neste trabalho obtiveram métricas de reprodutibilidade e de desvio percentual de volume superiores às soluções estado da arte.

Tabela 4 – Métricas de reprodutibilidade em conjunto de dados de teste-reteste Kirby-21.

Modelo	ICC	Diferença em volume em % (média±SD)
CERES	0,998	0,685±0,525
ACAPULCO	0,991	1,060±1,704
U-Net	0,998	0,639±0,422
nnU-Net	0,999	0,451±0,338

ICC: coeficiente de correlação intraclasse; SD: desvio padrão.

Os resultados iniciais dos modelos treinados para segmentação mostraram-se promissores, ainda que a disponibilidade de anotações de alta qualidade fosse reduzida. Devido às necessidades do grupo de pesquisa do TRACK-FA e alinhado aos objetivos do *grant* recebido da FARA, os esforços foram direcionados para o desenvolvimento do *pipeline* para a segmentação do núcleo denteado. No ínterim, foi priorizado para o contexto da segmentação do cerebelo a organização de um esforço de anotação manual de máscaras por um grupo treinado para essa tarefa. Assim, nessa atividade que está em progresso, já foram completadas mais de 50 máscaras por anotadores experientes. As mesmas metodologias empregadas à segmentação do núcleo denteado são aplicáveis para a segmentação cerebelar, sejam as técnicas de *augmentation* ou as arquiteturas testadas.

5.2 Segmentação do Núcleo Denteado

Os resultados do trabalho relacionados à segmentação do núcleo denteado são apresentados na forma de artigo submetido. O modelo desenvolvido será disponibilizado no repositório <https://github.com/art2mri/DentateSeg>. O manuscrito é apresentado a seguir.

Title: Automated Dentate Nucleus Segmentation from QSM Images Using Deep Learning

Author names and affiliations: Diogo H Shiraishi^{1*}, Susmita Saha^{2,3*}, Isaac M Adanyeguh⁴, Sirio Coccozza⁵, Louise A Corben^{6,7,2}, Andreas Deistung^{8,9}, Martin B Delatycki⁶, Imis Dogan^{10,11}, William Gaetz¹², Nellie Georgiou-Karistianis², Simon Graf⁸, Marina Grisoli¹³, Pierre-Gilles Henry⁴, Gustavo M Jarola¹, James M Joers⁴, Christian Langkammer¹⁴, Christophe Lenglet⁴, Jiakun Li¹⁵, Camila C Lobo¹, Eric F Lock¹⁵, David R Lynch¹⁶, Thomas H Mareci¹⁷, Alberto R M Martinez¹, Serena Monti¹⁸, Anna Nigri¹³, Massimo Pandolfo¹⁹, Kathrin Reetz^{10,11}, Timothy P Roberts¹², Sandro Romanzetti¹⁰, David A Rudko^{19,20,21}, Alessandra Scaravilli⁵, Jörg B Schulz^{10,11}, S H Subramony²², Dagmar Timmann²³, Marcondes C França¹, Ian H Harding^{24,25†}, Thiago J R Rezende^{1†}, TRACK-FA Neuroimaging Consortium[‡]

1 Department of Neurology, School of Medical Sciences, University of Campinas (Unicamp), Campinas, Brazil

2 School of Psychological Sciences, The Turner Institute for Brain and Mental Health, Monash University, Clayton, Victoria, Australia

3 Department of Neuroscience, School of Translational Medicine, Monash University, Melbourne, Australia

4 Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, Minnesota, United States of America

5 Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy

6 Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, Parkville, Victoria, Australia

7 Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

8 University Clinic and Outpatient Clinic for Radiology, Department for Radiation Medicine, University Hospital Halle (Saale), University Medicine Halle, Halle (Saale), Germany

9 Halle MR Imaging Core Facility, Medical Faculty, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

10 Department of Neurology, RWTH Aachen University, Aachen, Germany

11 JARA-BRAIN Institute Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich GmbH and RWTH Aachen University, Aachen, Germany

12 Department of Radiology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America

13 Department of Neuroradiology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

14 Department of Neurology, Medical University of Graz, Graz, Austria

15 Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, Minneapolis, Minnesota, United States of America

16 Department of Neurology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America

17 Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida, United States of America

18 Institute of Biostructures and Bioimaging, Italian National Research Council, Naples, Italy

19 Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada

20 McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, Montreal, Quebec, Canada

21 Department of Biomedical Engineering, McGill University, Montreal, Quebec, Canada

22 Department of Neurology and the Fixel Institute for Neurological Diseases, University of Florida, Gainesville, Florida, United States of America

23 Department of Neurology and Center for Translational and Behavioral Neuroscience (C-TNBS), Essen University Hospital, University of Duisburg-Essen, Essen, Germany

24 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

25 School of Translational Medicine, Monash University, Melbourne, Australia

* *Contributed equally*

† *Senior principal investigator*

‡ *see Acknowledgments for collaborators*

Running title: Automated segmentation of the dentate nuclei

Keywords: Dentate nuclei, QSM, MRI, deep learning, segmentation, TRACK-FA

Characters in the title/running head: 74/44, **Words in the abstract:** 233, **Words in the body of the manuscript:** 6,558, **Words in the introduction:** 609, **Words in the discussion:** 1274, **Figures:** 8; **Tables:** 8, **References:** 66

Corresponding author:

Dr. Thiago J R Rezende

Department of Neurology, School of Medical Sciences, University of Campinas (Unicamp)
Neuroimaging Laboratory

Rua Vital Brasil, 89-99

Cidade Universitária "Zeferino Vaz"

Campinas, São Paulo, Brazil

Zip code: 13083-888

E-mail: thiago.jrezende@gmail.com, thiagorz@unicamp.br

Phone: +55 19 3521-8242

Author contributions: Cohort principal investigator (PI): TJRR, IHH, MCFJ, SC, MBD, ID, NGK, PGH, CL, MP, KR. Imaging data analysis: DHS, SS, TJRR, IHH, IMA, JMJ, CCL, SM, AS. Imaging data collection: TJRR, MCFJ, SC, LAC, ID, MG, PGH, JMJ, CLK, THM, AN, MP, KR, TPR, SR, DAR. Clinical data collection: TJRR, MCFJ, LAC, ID, CLK, ARMM, MP, KR, SHS. Intellectual contribution to study design & conduct at your site: DHS, TJRR, IHH, SC, ID, WG, NGK, PGH, GMJ, ARMM, MP, TPR, DAR, JBS, SHS. Methods development: DHS, TJRR, IHH, NGK, SG. Statistical analysis: DHS, SS, TJRR, EFL. Manuscript review: DHS, SS, TJRR, IHH, MCFJ, SC, LAC, AD, ID, WG, NGK, SG, MG, PGH, CLK, CL, JL, EFL, THM, ARMM, AN, KR, TPR, SR, DAR, JBS, SHS, DT.

Funding: This work was supported by the Friedreich's Ataxia Research Alliance (FARA General Research Grant) and grants from the Australian National Health and Medical Research Council (NHMRC Ideas Grant 1184403). The funding agencies did not influence the study design, data collection, or manuscript drafting.

Competing interests: TJRR is employed by Biogen, receives a salary, and is a grant recipient from Friedreich's Ataxia Research Alliance (FARA). DHS is an Itaú Unibanco SA employee and a grant recipient from Friedreich's Ataxia Research Alliance (FARA). SC received fees from Amicus for the advisory board. LAC is funded by Friedreich's Ataxia Research Alliance (FARA) and is a consultant for Biogen Pharmaceuticals. ID is funded by Friedreich's Ataxia Research Alliance (FARA). NGK is funded by Friedreich's Ataxia Research Alliance (FARA) and CHDI Foundation Inc (New York, USA). PGH is a grant recipient from Friedreich's Ataxia Research Alliance (FARA) and the National Institute of Health (NIH) P41EB027061, P30NS076408, S10OD017974. JMJ is funded by Friedreich's Ataxia Research Alliance (FARA) and National Institute of Health (NIH). CLK reports grant FWF P35887. CL received research grants from Minoryx Therapeutics, research support from Biogen Inc and is funded by Friedreich's Ataxia Research Alliance (FARA) and NIH P41 EB027061. JL is funded by Friedreich's Ataxia Research Alliance (FARA). TM reports funding from National Institute of Health (NIH) U01 NS104326. SM is funded by the Italian

MUR for the project "SEE LIFE - StrEngthEning the Italian InFrastructure of Euro-bioimaging." AN is funded by the Italian Ministry of Health (RRC). KR is funded by Friedreich's Ataxia Research Alliance (FARA). SHS is industry support research for Reata, Biogen, Biohaven, Avidity Biosciences, Fulcrum therapeutics, Vertex, Arthex, PTC, Reneo, Larimar, has a consulting work for Reata, Biogen, Fulcrum, participated in speaking engagements for Biogen, MDA, Bionews, Medscape, and is funded by NIH, FDA, Muscular Dystrophy Association, Wyck Foundation, Friedreich's Ataxia Research Alliance, and National Ataxia Foundation. DT reports grants DFG, DE 2516/1-1 and TI 239/17-1. MCFJ reports funding from FAPESP (São Paulo Research Foundation). IHH reports Friedreich's Ataxia Research Alliance general research grant and NHMRC Ideas and Investigator Grants (2026191, 1184403). All other authors declare no financial or non-financial competing interests.

1. Abstract

Purpose

To develop a dentate nucleus (DN) segmentation tool using deep learning (DL) applied to brain quantitative susceptibility mapping (QSM) images.

Materials and Methods

Brain QSM images from 132 healthy controls and 170 individuals with cerebellar ataxia or multiple sclerosis were collected from nine different datasets worldwide for this retrospective study. Manual delineation of the DN (gray matter and white matter hilus) was first undertaken by experienced raters with a robust quality control process. Performance of automated segmentation was compared following training using several DL architectures. A two-step approach was implemented, composed of a localization model followed by DN segmentation.

Results

The manual tracing protocol produced ground-truth data with high intra-rater (average ICC 0.906) and inter-rater reliability (average ICC 0.776). Initial DL architecture exploration indicated that the nnU-Net framework performed best. The two-step localization plus segmentation pipeline achieved a Dice score of 0.898 ± 0.031 and 0.894 ± 0.036 for left and right DN, respectively. In external validation, our algorithm outperformed the leading existing automated tool (left/right DN Dice $0.863 \pm 0.038/0.843 \pm 0.066$ vs. $0.568 \pm 0.222/0.582 \pm 0.239$). The model demonstrated generalizability across unseen datasets during the training step. The measures showed a superior correlation index with manual annotations and performed well in both isotropic and anisotropic QSM datasets.

Conclusion

We provide a model that accurately and efficiently segments the DN from brain QSM images. The model can be readily deployed for use in observational, natural history, and treatment trials for biomarker discovery.

2. Introduction

The dentate nuclei (DN) are the largest of the deep cerebellar nuclei and are the primary efferent stations of the human cerebellum. The DN are primarily innervated by Purkinje cells of the lateral hemispheres of the cerebellar cortex and give rise to the major ascending cerebellar output pathways, including the dentato-rubral and dentato-thalamo-cortical tracts¹. The DN are divided into motor and non-motor functional subregions with distinct profiles of disynaptic axonal innervation (*via* the thalamus) to cerebral cortices¹. These cerebral regions, in turn, innervate corresponding subregions of the cerebellar cortex *via* descending disynaptic projections that pass through the pontine nuclei, forming the cerebro-cerebellar circuitry².

Cerebro-cerebellar systems, with the DN as a key hub, are implicated in a broad array of motor, cognitive, social, and language functions³. Abnormalities in the DN may therefore contribute to disruption in large-scale brain networks involved in a broad array of behavioral deficits. Abnormal functional connectivity between the DN and brain cortical areas has been reported in patients with Alzheimer's disease, autism, and schizophrenia⁴⁻⁶. Furthermore, neuropathological studies have provided evidence that the DN plays a central role in the pathogenesis of several cerebellar diseases, especially in the inherited cerebellar ataxias⁷.

Despite the established importance of the DN in cerebro-cerebellar loops, and growing evidence of involvement in brain diseases, direct *in vivo* quantitative investigations of the structure of the DN and other deep cerebellar nuclei using neuroimaging in humans are scarce⁸⁻¹¹. Such investigations have been particularly challenging due to the tissue properties of these nuclei that make them invisible or poorly defined using standard MRI sequences, such as T1-weighted and T2-weighted images. Susceptibility-weighted MRI (SWI) offers a limited solution, allowing for visualization of the DN due to their high iron content¹²⁻¹⁴. However, SWI is a qualitative technique that is primarily used for the clinical detection of vascular abnormalities and microangiopathies¹⁵. Although useful, SWI has several limitations, including its non-quantitative nature and distortion of tissue boundaries due to blooming effects and non-local phase contributions of the iron deposits on the tissues. These limitations have largely been overcome through the development of quantitative susceptibility mapping (QSM)¹⁶. QSM allows for more precise mapping of the anatomy and a more direct link to the underlying iron concentration¹⁷, providing opportunities for direct, quantitative evaluation of DN structure and microstructure in clinical populations^{8,11}.

QSM has been employed to demonstrate robust and/or progressive changes in the DN in people with inherited cerebellar ataxias, including Friedreich's ataxia (FRDA)^{11,18} and spinocerebellar ataxias (SCA)⁸. Changes in the structure and susceptibility of DN related to healthy aging and other movement disorders, including Parkinson's disease and essential tremor, have also recently been examined^{9,19,20}. Taken together, these studies demonstrate the utility of QSM in the neuroimaging toolkit for examining the DN in health and disease, and motivate investigation of DN changes in other neurological, developmental, and psychiatric diseases that impact cerebellar circuitry.

Although QSM holds great promise for quantifying and tracking DN changes in disease, a major roadblock to-date in undertaking large-scale and reliable QSM

investigations of the DN in patient cohorts has been the reliance on manual, hand-drawn segmentations. In order to overcome this limitation, fully automated tools are necessary. The MRICloud toolkit^{21,22} provided the first (and currently, to our knowledge, only) publicly available automated DN segmentation tool. However, MRICloud has not been trained on data with DN pathology. In this work, we address these limitations by utilizing data from healthy subjects and cerebellar ataxia or multiple sclerosis (MS) patient cohorts acquired at multiple imaging centers using different acquisition protocols to develop an optimized and generalizable deep-learning analytical tool for segmenting the DN using QSM images. This tool can be readily deployed in observational, natural history, and treatment trials.

3. Materials and Methods

3.1. Data

Multi-echo gradient-recalled echo MRI data was acquired using four different MRI protocols implemented across ten imaging centers around the world using 3 Tesla Philips or Siemens MRI scanners (Tables 1 and 2). The collected research data was de-identified at each source, ensuring adherence to a data pipeline free from personal health information. The dataset included a total of 132 healthy controls, 154 individuals with FRDA, and 16 people with MS, as described in Table 2. Follow-up scans at 12 months were available from 55 of these participants. Images of FRDA and MS patients were selected to account for anatomical variability in DN caused by neurodegeneration throughout the course of the diseases. The corresponding MRI acquisition protocols have been previously published^{11,23,24} and are summarized in Table 1.

Table 1. Acquisition protocols for each dataset.

Dataset	Scanner	Sequence	TR (ms)	TE1 (ms)	Δ TE (ms)	# of echoes	FoV (mm)	Image matrix (voxels)	Voxel size (mm)	Acquisition time
Aachen Campinas CHoP Melbourne Florida Minnesota	3T Siemens Skyra 3T Siemens Prisma 3T Philips Ingenia	GRE	27	3.7	6	4	220 x 220 x 176	208 x 256 x 176	0.86 iso	7' 22"
Naples	3T Siemens Magnetom Trio	GRE	32	7.38	14.76	2	230 x 194 x 160	378 x 448 x 160	0.51 x 0.51 x 1.0	9' 53"
IMAGE-FRDA	3T Siemens Skyra	GRE	30	7.38	14.76	2	230 x 230 160 axial slices	232 x 256 x 160	0.90 iso	11' 30"
INFLAM-FRDA	3T Siemens Biograph	GRE	31	5.70	5.27	5	230 x 230 104 axial slices	384 x 384 x 104	0.60 x 0.60 x 1.2	6' 50"

GRE: gradient recalled echo; iso: isotropic; TR: repetition time; TE: echo time; TE1: first echo time; FoV: field of view; iso: isotropic.

Table 2. Subject demographics for each dataset.

	Aachen, Campinas, CHoP, Melbourne, Florida, Minnesota		Naples		IMAGE-FRDA		INFLAM-FRDA		
	Controls	FRDA	Controls	FRDA	MS	Controls	FRDA	Controls	FRDA
Subjects	44	98	48	12	16	31	30	9	14
Children (<18 years)	14	31	0	3	0	0	0	0	0
Age (mean\pmSD)	23.6 \pm 7.7	23.9 \pm 8.7	37.8 \pm 12.7	32.6 \pm 15.8	46.8 \pm 5.6	37.6 \pm 13.1	35.7 \pm 12.2	28.6 \pm 5.8	27.8 \pm 7.6
Sex (M/F)	24/20	49/49	19/29	4/8	5/11	16/15	17/13	4/5	9/5

FRDA: Friedreich's ataxia; MS: multiple sclerosis.

A sample QSM image for each dataset is presented in Figure 1, and Figure 2 provides an overview of the data workflow. The QSM images were reconstructed using the JHU/KKI QSM^{21,25} and STI Suite (<https://people.eecs.berkeley.edu/~chunlei.liu/software.html>) toolboxes, using Laplacian unwrapping to overcome phase aliasing, V-SHARP for background field removal²⁶ and either MEDI^{27,28} or iLSQR²⁹ for field-to-susceptibility reconstruction.

To train the deep learning (DL) models, we divided the available data into 70/10/20 % proportions for training, validation, and test sets, respectively. We applied stratified sampling, creating groups that maintain the relative distributions by MRI acquisition center, while also keeping images belonging to an individual in just one set, minimizing the possibility of data leakage.

The ethics committee or institutional review board (IRB) respective to each project data source or site approved the use or ethics waiver for this retrospective study, and all participants provided written informed consent prior to original data collection. The TRACK-FA steering committee approved the data use, and IRB reference numbers were previously published²³ (Monash Health Human Research Ethics Committee: RES-20-0000-139A; Children's Hospital of Philadelphia: IRB 20-017611; University of Minnesota: IRB STUDY00009047; University of Florida: IRB202000399; RWTH Aachen University: EK195/20; University of Campinas (CAAE NO): 83241318.3.1001.5404; McGill IRB Approved Project Number: 2022-8676). Ethics approval was obtained independently for the remaining studies, respectively: Ethical Committee "Carlo Romano" of the University of Naples "Federico II" (Naples A: 209/13, Naples B: 47/15), Monash University Human Research Ethics Committee (IMAGE-FRDA: 13201B, INFLAM-FRDA: 7810), and University of Minnesota IRB (1210M22281).

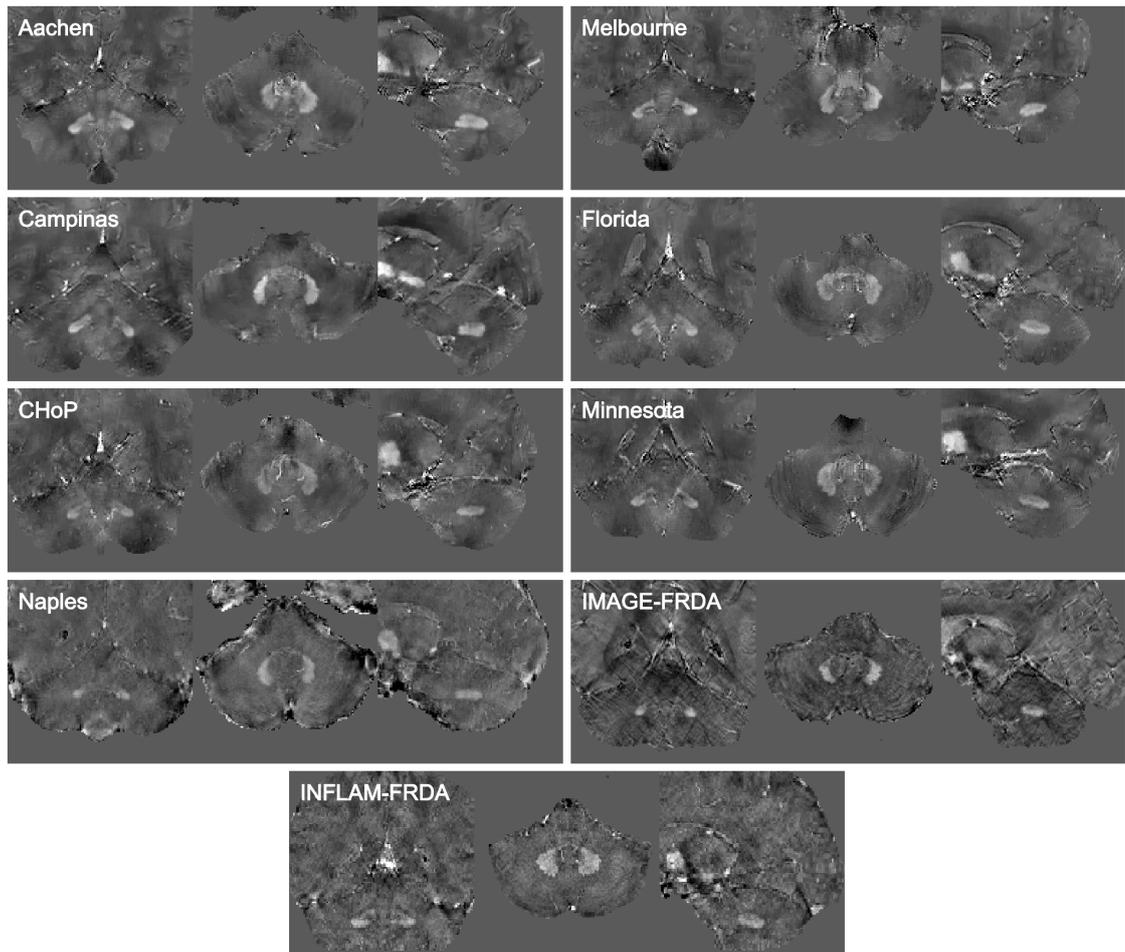


Figure 1. Image samples for each dataset. Randomly selected. Images were cropped focusing on the cerebellum, voxel intensities were normalized using z-score to improve visualization.

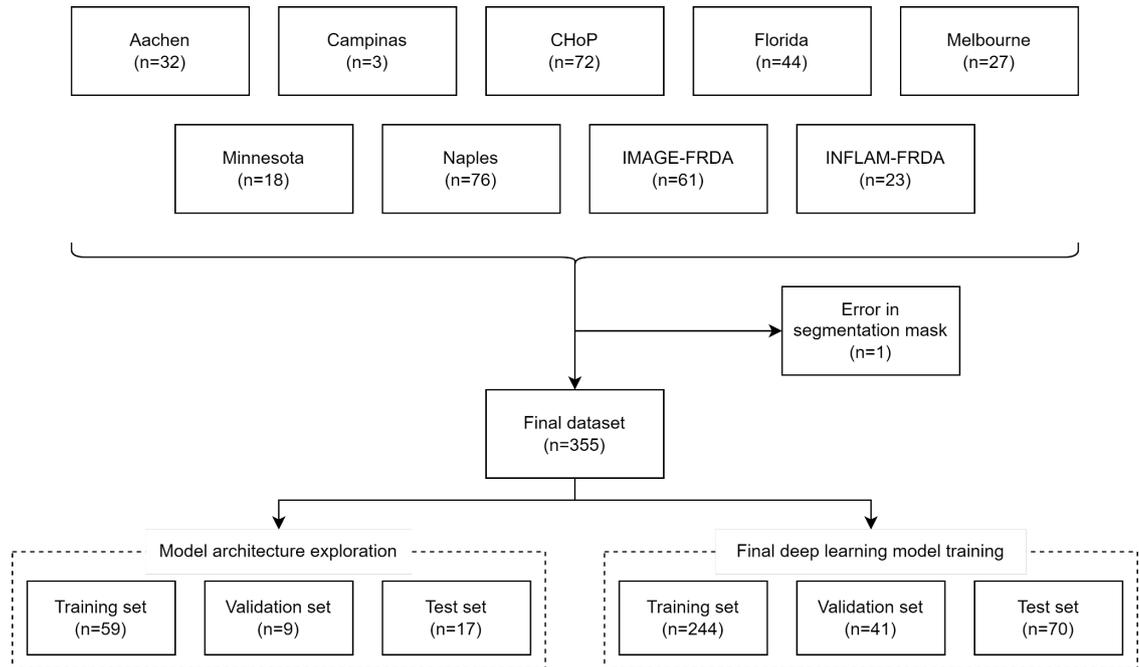


Figure 2. Diagram details the inclusion of image datasets into the study.

3.2. Ground Truth Segmentation

The DN were manually traced on each QSM image to establish the ground truth dataset using MRICron (<https://www.nitrc.org/projects/mricron>), ITK-SNAP³⁰, or FSLeys³¹ annotation software. The contrast threshold was set at a constant range (-0.1 to 0.2 ppm) for all images to minimize bias in edge detection within and across raters.

The segmentations included the full 3D region of hyperintensity, including both the outer gray matter ribbon and the central white matter hilum of the DN (Figure 3). Manual tracing was performed in all three planes (axial, coronal, and sagittal) to ensure a smooth surface was generated and neighboring nuclei (e.g., emboliform) were not included. The left DN (LDN) and right DN (RDN) were uniquely segmented on each image (Figure 3).

Each image was manually segmented by one of three experienced raters (SS, SC, or IHH), blinded to the disease pathology. Inter-rater and intra-rater reliability was assessed using a sample of 9 randomly selected images, stratified by acquisition center. These images were duplicated and randomly shuffled so that the rater was unaware of the order or the repetition. All three raters segmented this set of 18 images. Intra-rater and inter-rater variability were calculated using Dice score, Hausdorff Distance (HD), and intraclass correlation coefficient (ICC) metrics.

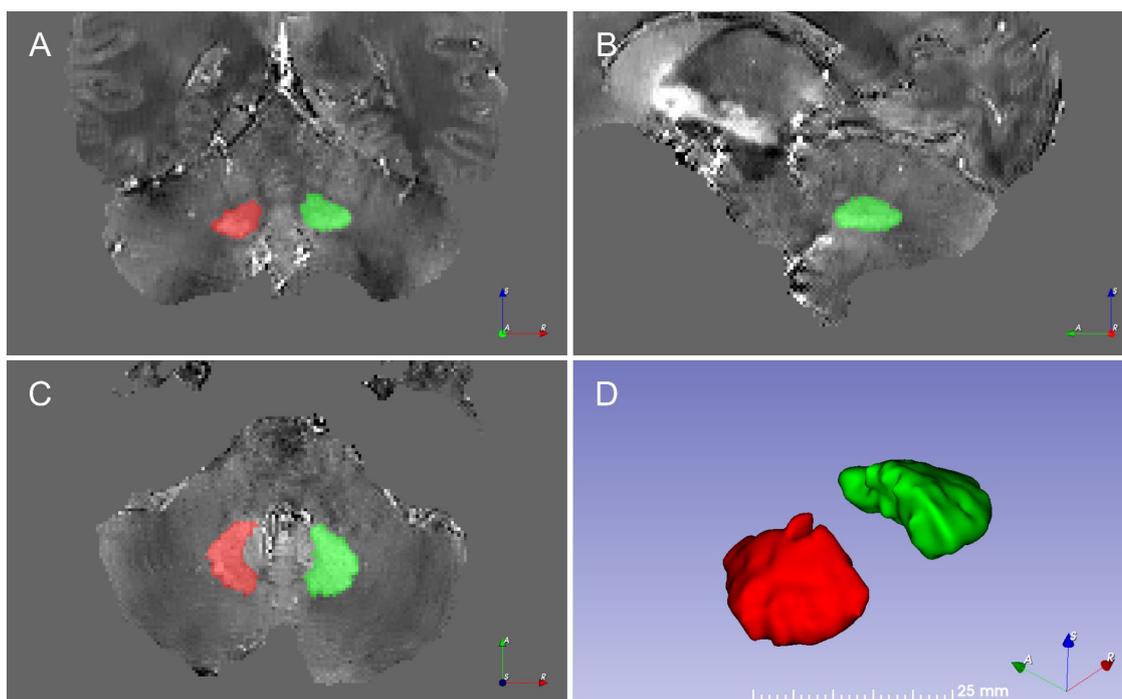


Figure 3. Manually segmented mask delineating the left dentate nucleus (LDN) in red and the right dentate nucleus (RDN) in green from a sample QSM dataset, healthy participant. Coronal, sagittal, and axial views as well as shaded-surface display are shown from A - D, respectively.

3.3. Deep Learning Architectures

A number of artificial intelligence approaches for image segmentation are available. In the DL domain, convolutional neural network (CNN) architectures are typically used to process image data. For image segmentation tasks, a great number of studies rely on U-Nets³² and their 3D variation, 3D U-Net³³, both based on CNNs. Numerous variations of these architectures have been reported in the literature. The no-new-Net (nnU-Net) is considered one of the state-of-the-art medical imaging segmentation frameworks, capable of dealing with image sets comprising different domains³⁴. The nnU-Net does not implement a rigorous complex architecture, instead relying on a conventional U-Net with deep supervision³⁵. More recently, the scientific community introduced new architectures based on transformers³⁶, which arose in the language sequence models domain. These include TransUNet³⁷, UNETR³⁸ and Swin UNETR³⁹. In this work, we applied and contrasted the performance of three DL architectures, described below, to develop an optimal approach for automated DN parcellation.

3.4. 3D U-Net with Deep Supervision

The 3D U-Net³³ is an extension of the original U-Net architecture³² to handle three-dimensional data. Designed for image segmentation tasks, the architecture is composed of a contracting path, which gradually downsamples the input image, and an expanding path, which upsamples the feature maps back to the original image size. Also, the network makes use of skip connections, allowing information to flow directly from the contracting path to the

corresponding layer in the expanding path. This helps to preserve spatial information and improve the accuracy of the segmentation. Additionally, the 3D U-Net can leverage the context information in the whole 3D image, learning intricate and complex spatial relationships between the voxels in the image.

Deep supervision (DS)³⁵ is a technique that adds extra output layers at different levels of the network. These output layers are trained to predict the final segmentation and guide the network in the earlier stages of training, providing additional guidance that helps improve segmentation accuracy. This allows gradients to be injected deeper into the network and facilitates the training of all layers in the network³⁴. This architecture is also adopted by the nnU-Net framework.

3.5. Swin UNETR

The Shifted WINDows UNET TRansformers (Swin UNETR) is a Swin transformer-based deep learning model designed for medical image segmentation³⁹. Swin transformer⁴⁰ implements a hierarchical vision transformer that computes self-attention locally in non-overlapping windows, enabling the model to capture long-range dependencies. In the hierarchical structure, patches are merged layer by layer, decreasing the dimension of feature maps at each stage. The cyclic-shifting implements the shifted window partitioning approach and introduces important cross-windows connections, overlapping with previous windows, while the number of patches remains fixed. These properties explain the linear computational complexity to image size, unlike to the quadratic in other transformed-based models, such as Vision Transformers⁴¹ (ViT), which would impose an issue of scalability for semantic segmentation, a dense prediction task. Swin transformers build hierarchical feature maps to obtain multi-resolution feature representations, making them suitable as backbone networks for various computer vision tasks. In Swin UNETR, the Swin transformer serves as the encoder and is integrated with a CNN-based decoder through skip connections in a U-shaped architecture.

3.6. nnU-Net

The nnU-Net is a deep learning framework for medical image segmentation that features automated configuration³⁴. The core architecture of nnU-Net is a U-Net with DS, which is enhanced with a data-driven initialization strategy. This approach adapts the training to the specific characteristics of the input data, resulting in improved performance. The pipeline includes data augmentation and training across different configurations such as 2D U-Net, 3D U-Net at full resolution, and 3D U-Net cascade (a combination of a model trained on downsampled images followed by a refinement model at full resolution). Additionally, the nnU-Net has the capability to apply post-processing steps, such as “non-largest component suppression”, if it improves the results. It can also self-adjust hyperparameters to fit the GPU memory available during training.

This framework has been demonstrated to achieve state-of-the-art performance on various medical image segmentation tasks, leading to widespread adoption in the medical imaging community.

3.7. Augmentation Techniques

In order to improve model generalization capacity, we applied data augmentation techniques such as flipping, rotation, scaling, elastic deformation, and intensity scale and shift operations. These operations were randomly applied to each image during the training phase. Utilizing these transformations is an important strategy to artificially increase data variability, thereby preventing overfitting and boosting generalizability. To accomplish this, we selected MONAI⁴² augmentation implementations.

3.8. Deep Learning Segmentation Pipeline

The DL segmentation pipeline is presented in Figure 4. First, the QSM image is resampled to a common isotropic voxel spacing of 0.86 mm, consistent with the original resolution of the majority of the datasets.

The second step in our pipeline is the application of the localization model, which is a deep learning network designed to identify the location of the cerebellum within a 3D MRI dataset. The localization model was trained with cerebellar masks obtained with ACAPULCO⁴³ based on T1-weighted MRI of each subject. To ensure spatial alignment of the cerebellum masks with the QSM images, the T1w data were registered to the corresponding QSM data using ANTs toolkit⁴⁴, and the resulting transformation matrix was applied to the cerebellum mask. The localization model will therefore output a cerebellar mask. The centroid of the mask is determined and used to place a bounding box cropping around the cerebellum, ensuring the preservation of the entire structure without any loss or truncation, thereby constraining and reducing the spatial extent of the region of interest (RoI). This approach allows the retention of contextual anatomical references.

The third step was the segmentation stage. The output of this model was the binary mask for each dentate nucleus label (left and right). Subsequently, another resampling process is performed to provide the predicted DN segmentation mask in the original voxel spacing and in the same dimensions as the input image.

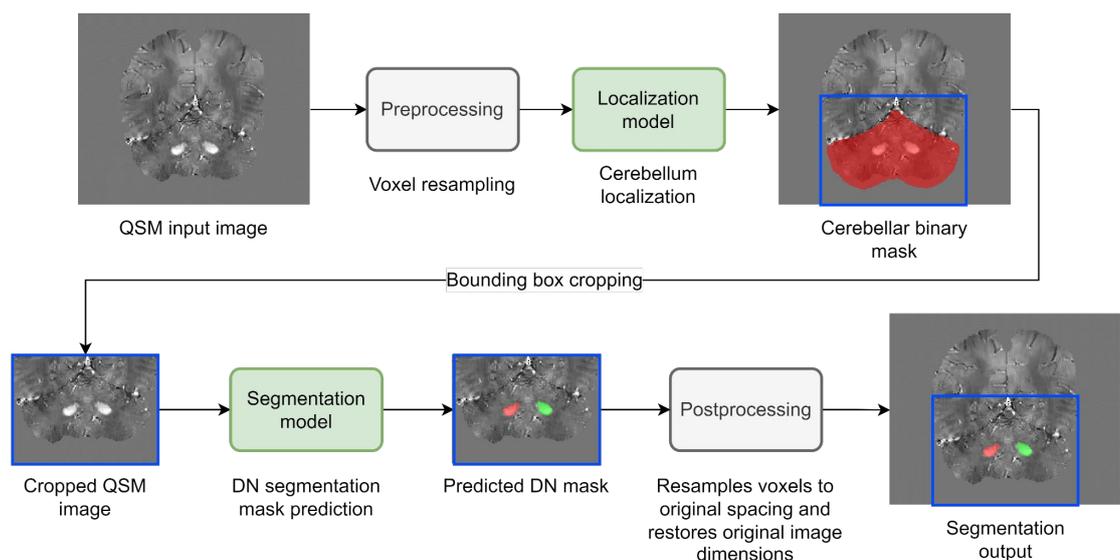


Figure 4. Deep learning QSM segmentation pipeline.

3.9. Statistical Analysis

The Sørensen-Dice similarity coefficient (DSC), also known as the Dice score coefficient^{45,46}, is a widely used metric for assessing segmentation tasks. It is noteworthy that the Dice score, which measures the overlap between two binary segmentation masks, considers both size and localization agreement⁴⁷ that is, both volumetric and anatomic characteristics. The Dice score is defined as follows:

$$Dice = 2 \frac{|G \cap P|}{|G| + |P|}$$

where $|A|$ is the cardinality of set A . G and P represent the ground truth segmentation mask and the model prediction segmentation, respectively. This metric ranges from 0 to 1, with a value of 1 indicating a perfect agreement. The Dice score was our main performance metric, and it was also part of the loss metric for model training in the form of Dice loss:

$$Dice\ loss = 1 - Dice$$

Additionally, other segmentation metrics were evaluated. The Hausdorff Distance⁴⁸ and Average Hausdorff Distance (AHD)⁴⁹ are relevant for boundary quality assessment. AHD calculates the average distance between points in both masks instead of considering only the maximum distance as in HD. Also, Intersection over Union (IoU) or Jaccard index⁵⁰, which is similar to Dice, was evaluated. Finally, volume similarity⁴⁹ was assessed to measure the quantitative volume agreement, ignoring shape, position, and quality information. To compare metrics between models, we used the Wilcoxon signed-rank test. Kolmogorov-Smirnov test was employed to assess normality in continuous variables. Pearson correlation coefficients were chosen to evaluate relationships within variables. We measured the reliability of measurements from multiple raters using the intraclass correlation coefficient (ICC) and evaluated the results (ICC<0.5 poor, 0.5≤ICC<0.75 moderate, 0.75≤ICC<0.9 good, and ICC≥0.9 excellent reliability)⁵¹. ICC confidence intervals were calculated to provide a range within which the ICC is likely to fall, typically with 95% confidence. Statistical analysis was conducted using Python 3.10.8, R 4.1.1, and SPSS 27.0.1.0. P-values < 0.05 were considered statistically significant.

3.10. Comparison with other available methods

In order to compare our results with the leading available automated DN segmentation solution, we processed all our test images through the MRICloud web-based service^{21,22} using its susceptibility multi-atlas tool for segmentation of QSM images.

The MRICloud pipeline requires both coregistered QSM and skull-stripped T1 images in ANALYZE file format. We used ANTs (SyNQuick) for co-registration, FSL⁵² Brain Extraction Tool (BET) for skull-stripping, and FSL for file conversion between NIFTI and ANALYZE formats. MRICloud jobs for the segmentation model architecture comparison section were executed between January 5 and January 7, 2023, and then again from March 9 to March 31, 2024, for the external validation datasets.

3.11. External Validation

In order to assess the performance of our DN segmentation model on unseen data, 21 external datasets of 10 participants were acquired from three new imaging sites: n=4 healthy participants from Istituto Neurologico “Carlo Besta”, Italy; n=2 MS patients from the Medical University of Graz, Austria; and n=3 healthy participants and n=1 FRDA patient from McGill University. QSM maps were derived for all participants using both the JHU/KKI QSM v3.0^{21,53} and SEPIA v1.2.2.6⁵⁴ toolboxes, resulting in two images per participant for Carlo Besta and McGill datasets. Three reacquired Carlo Besta anisotropic QSM images processed with SEPIA were added. An experienced rater (SS) manually segmented the right and left DN, and the data was processed through both our models and MRICloud to compare their performances. For SEPIA processing, 3D best path⁵⁵ was used for phase unwrapping and MEDI non-linear fit for echo phase combination, Laplacian Boundary Value (LBV)⁵⁶ approach for background magnetic field removal, and quantitative susceptibility maps were obtained using STreaking Artifact Reduction for QSM (Star-QSM)⁵⁷.

The institutional ethics committee respective to each project approved their study (Fondazione IRCCS Istituto Neurologico “Carlo Besta”: 42/2017 07/06/2017; Medical University of Graz local ethics-committee: 31-432ex18/191264-2019).

4. Results

4.1. Intra-Rater and Inter-Rater Reproducibility Results

All raters demonstrated good to excellent intra-rater reliability ($ICC > 0.8$; Table 3), and moderate ($ICC > 0.5$) to good inter-rater agreement ($ICC > 0.75$) when evaluated in pairs⁵¹ (Table 4). Analyzing all three raters at once, the ICC is 0.763, 95% CI [0.541, 0.897] for LDN and 0.675, 95% CI [0.173, 0.883] for RDN. Dice and HD additionally support strong reliability, evaluating segmentation overlapping and annotation deviations. Volume similarity metric, comparing the segmented volumes (mm^3), indicates that absolute differences within and between raters are marginal. The results are shown in Tables 3 and 4.

Table 3. Intra-rater reproducibility results (mean±SD).

Metric	Rater 1		Rater 2		Rater 3	
	LDN	RDN	LDN	RDN	LDN	RDN
Dice	0.884±0.025	0.888±0.027	0.904±0.023	0.918±0.024	0.892±0.014	0.912±0.010
HD	3.276±1.174	3.491±1.169	2.397±0.837	2.109±0.879	2.324±0.394	1.745±0.256
Volume similarity	0.943±0.037	0.961±0.029	0.964±0.023	0.963±0.029	0.969±0.016	0.980±0.015
ICC	0.839	0.849	0.939	0.902	0.930	0.977
[95% CI]	[0.258, 0.965]	[0.440, 0.964]	[0.306, 0.989]	[0.640, 0.977]	[0.735, 0.984]	[0.905, 0.995]

LDN: left dentate nucleus, RDN: right dentate nucleus, HD: Hausdorff distance, ICC: intraclass correlation coefficient, CI: confidence interval.

Table 4. Inter-rater reproducibility results (mean±SD).

Metric	Raters					
	1 and 2		2 and 3		1 and 3	
	LDN	RDN	LDN	RDN	LDN	RDN
Dice	0.852±0.040	0.864±0.027	0.852±0.032	0.855±0.025	0.872±0.020	0.870±0.028
HD	3.361±1.325	3.569±1.021	3.416±1.154	3.573±1.230	2.946±1.266	3.867±1.451
Volume similarity	0.931±0.055	0.944±0.041	0.918±0.040	0.897±0.041	0.950±0.026	0.933±0.039
ICC	0.751	0.740	0.709	0.599	0.845	0.731
[95% CI]	[0.451, 0.899]	[0.298, 0.905]	[0.268, 0.890]	[-0.079, 0.886]	[0.376, 0.951]	[-0.048, 0.925]

LDN: left dentate nucleus, RDN: right dentate nucleus, HD: Hausdorff distance, ICC: intraclass correlation coefficient, CI: confidence interval.

4.2. Localization Model

The model for the cerebellum localization features a 3D U-Net model with feature channels ranging from 16 to 256, doubling at each subsequent level. Levels combine two 3x3x3 convolutions, followed by instance normalization⁵⁸ and Parametric Rectified Linear Unit (PReLU)⁵⁹ activations. An AdamW optimizer⁶⁰ with learning rate = 3×10^{-4} , weight decay = 1×10^{-5} , $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. A learning rate scheduler reduced the learning rate by a factor of 0.5 if no improvement in loss was observed for 20 epochs. The batch size was configured to 1, and training was limited to 400 epochs. After 105 epochs, the model training was interrupted by the early stopping scheduler, after 30 epochs without a decrease in Dice loss. Consequently, the model checkpoint at epoch 75, which performed best on the validation set, was saved. The trained model showed a high Dice score, 0.918 ± 0.030 , when compared to the ground truth. After running inference for all available QSM images, they were cropped to 128x96x96, considering the centroid of the mask as the crop center for the enclosing region of interest. The localization model inference runs in less than five seconds when executed on CPU-only hardware and provides the model with as

much spatial context as possible, while taking into account computational resources constraints.

4.3. Segmentation Model Architectures Comparison

All the results presented below refer to metrics calculated on the test set (17 images) from the model experimentation set (Figure 2). The performance of three model architectures (3D U-Net with DS, Swin UNETR architectures, and nnU-Net framework) were compared with MRICloud (Table 5). The nnU-Net architecture provided the best performance across most comparison metrics (Table 5), and all DL architectures were substantially superior to MRICloud (Dice score comparison, $p < 0.0001$ in all cases; Figure 5). Jaccard index and AHD also indicate an advantage for nnU-Net, considering overlap and mask similarity. While the Hausdorff distance is lower in the U-Net with DS model, the average values for the two dentate nuclei are close to those of nnU-Net. Finally, despite U-Net with DS having the best average volume similarity, the nnU-Net standard deviation is lower, which is preferable.

Table 5. Trained models and MRICloud performance metrics (mean±SD) for left DN (LDN) and right DN (RDN) on the experimentation set.

Model	Dice		Jaccard		HD		AHD		Volume similarity	
	LDN	RDN								
U-Net with DS	0.889±0.021	0.895±0.023	0.802±0.033	0.810±0.037	2.514±0.749	2.361±0.654	0.117±0.024	0.112±0.027	0.965±0.021	0.963±0.026
Swin UNETR	0.890±0.022	0.896±0.021	0.803±0.035	0.812±0.034	2.887±0.885	2.495±0.550	0.120±0.032	0.111±0.024	0.962±0.023	0.953±0.029
nnU-Net	0.896±0.020	0.900±0.021	0.811±0.032	0.819±0.034	2.518±0.915	2.198±0.534	0.112±0.025	0.106±0.025	0.964±0.018	0.958±0.023
MRICloud	0.725±0.074	0.760±0.053	0.573±0.090	0.616±0.068	6.302±2.611	5.736±1.725	0.445±0.187	0.367±0.125	0.850±0.086	0.877±0.074

DN: dentate nucleus; LDN: left DN; RDN: right DN; DS: deep supervision; HD: Hausdorff distance; AHD: average Hausdorff distance.

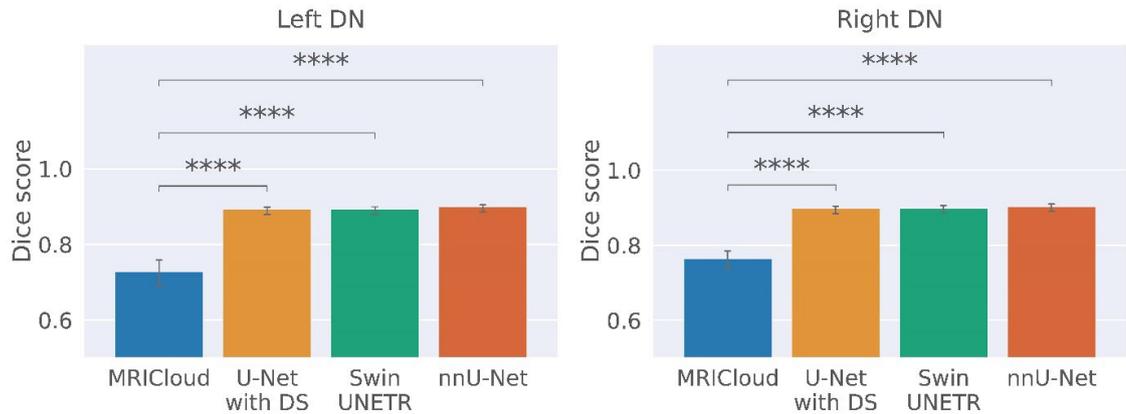


Figure 5. Bar plots of the trained segmentation models and MRICloud. All metrics of the segmentation model variants are statistically significantly higher than the MRICloud results. ns: non-significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. DN: dentate nucleus, DS: deep supervision.

After the extensive exploration of model architectures, the nnU-Net framework was selected for our study.

4.4. Final Segmentation Model Outputs

The final trained model (nnU-Net) with the entire dataset resulted in metrics on the test set presented in Table 6, including those for the control and patient groups individually. Samples of predicted segmentation masks for each center are shown in Figure 6. In agreement with that, the predicted DN volumes highly correlate with the volumes of the ground truth annotations (Figure 7).

Table 6. Trained nnU-Net final model performance metrics (mean±SD) for left DN (LDN) and right DN (RDN) on the complete test set.

Group	Dice		Jaccard		HD		AHD		Volume similarity	
	LDN	RDN	LDN	RDN	LDN	RDN	LDN	RDN	LDN	RDN
All participants	0.898±0.031	0.894±0.036	0.816±0.050	0.810±0.057	3.124±1.825	4.025±4.569	0.118±0.060	0.140±0.144	0.943±0.044	0.944±0.050
Controls	0.893±0.033	0.891±0.044	0.808±0.053	0.806±0.067	3.408±1.916	3.850±2.046	0.127±0.068	0.134±0.073	0.949±0.046	0.951±0.058
Patients*	0.901±0.029	0.896±0.032	0.821±0.048	0.813±0.051	2.932±1.783	4.144±5.747	0.112±0.055	0.144±0.179	0.940±0.043	0.939±0.044

* FRDA and MS patients. FRDA: Friedreich's ataxia; MS: multiple sclerosis; DN: dentate nucleus; LDN: left DN; RDN: right RN; HD: Hausdorff distance; AHD: average Hausdorff distance.

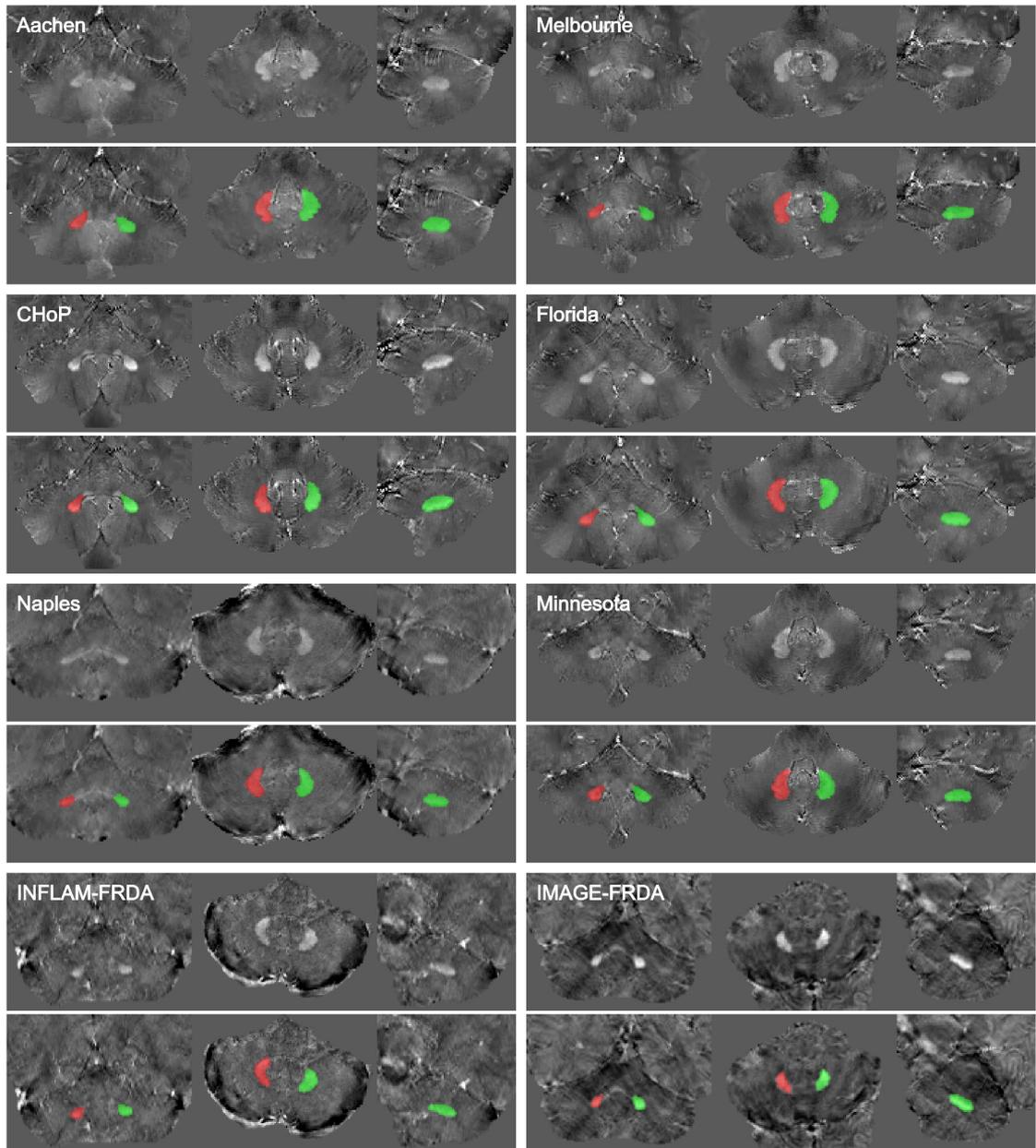


Figure 6. Overview of pipeline prediction examples for datasets of each center. QSM images without (top) and with the predicted segmentation mask as overlay (bottom) are presented in the coronal, axial, and sagittal planes, respectively. Left dentate nucleus is depicted in red and the right one in green.

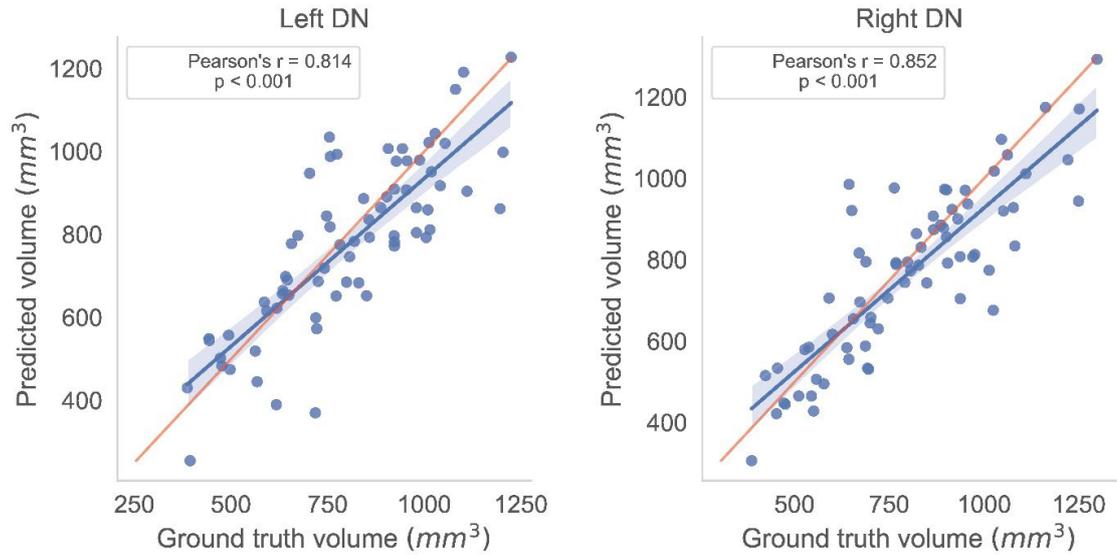


Figure 7. Volume scatter plots for the final segmentation model. Ground truth measures versus model predicted values. Each dot represents a prediction sample from the test set. Pearson correlation coefficients and p-values are indicated in each plot. The red line represents the line of identity, i.e., the expected perfect fit. DN: dentate nucleus.

4.5. External Validation

Our model performed satisfactorily on an external dataset of 21 images from three different sites, achieving Dice scores of 0.863 ± 0.038 for LDN and 0.843 ± 0.066 for RDN. In contrast, MRICloud demonstrated significantly lower performance for both LDN (Dice score: 0.568 ± 0.222) and RDN (Dice score: 0.582 ± 0.239). The observed performance difference in Dice metric was found to be statistically significant according to the Wilcoxon test for both LDN and RDN ($p < 0.05$). Additionally, Pearson correlation between volumes predicted by the DL model (LDN: $r = 0.740$, $p < 0.001$; RDN: $r = 0.484$, $p = 0.026$) and those manually traced was significant for both DN sides ($p < 0.05$), a result not observed with MRICloud outcomes (LDN: $r = 0.420$, $p = 0.058$; RDN: $r = 0.187$, $p = 0.417$). Interestingly, our model showed consistency across three external datasets when considering the individual QSM subsets per acquisition center. Furthermore, our segmentation pipeline demonstrated robust performance when evaluating susceptibility maps generated by the alternative QSM reconstruction method (Star-QSM, see section 3.11), yielding high Dice scores both overall and individually for each site (Tables 7 and 8).

Table 7. External validation Dice scores (mean \pm SD) for both DN, comparing results from MRICloud and our proposed DL model.

Dataset	Voxel	QSM reconstruction	n	MRICloud		Our model	
				LDN	RDN	LDN	RDN
Carlo Besta	Anisotropic	Star-QSM	3	0.513 \pm 0.125	0.494 \pm 0.214	0.830 \pm 0.007	0.741 \pm 0.067
	Isotropic	MEDI	4	0.537 \pm 0.305	0.538 \pm 0.307	0.900 \pm 0.017	0.892 \pm 0.025
		Star-QSM	4	0.478 \pm 0.114	0.519 \pm 0.223	0.856 \pm 0.039	0.834 \pm 0.048
Graz	Anisotropic	Star-QSM	2	0.247 \pm 0.349	0.298 \pm 0.422	0.841 \pm 0.001	0.788 \pm 0.098
TRACK-FA McGill	Isotropic	MEDI	4	0.698 \pm 0.121	0.710 \pm 0.073	0.895 \pm 0.008	0.880 \pm 0.023
		Star-QSM	4	0.760 \pm 0.029	0.771 \pm 0.034	0.837 \pm 0.046	0.872 \pm 0.028

DN: dentate nucleus; QSM: quantitative susceptibility mapping; LDN: left DN; RDN: right RN.

Table 8. Quantitative validation (mean \pm SD) of QSM reconstruction methods on external data.

QSM reconstruction	Voxel	MRICloud		Our model	
		LDN	RDN	LDN	RDN
MEDI	Isotropic	0.618 \pm 0.231	0.624 \pm 0.226	0.898 \pm 0.012	0.886 \pm 0.023
Star-QSM	Anisotropic	0.407 \pm 0.244	0.415 \pm 0.281	0.835 \pm 0.008	0.760 \pm 0.073
	Isotropic	0.619 \pm 0.169	0.645 \pm 0.200	0.847 \pm 0.041	0.853 \pm 0.042

DN: dentate nucleus; QSM: quantitative susceptibility mapping; LDN: left DN; RDN: right RN.

4.6. Biological Findings

4.6.1. DN Volume versus Mean Magnetic Susceptibility

Significant positive correlations between DN volume and magnetic susceptibility were observed in the healthy control group ($r > 0.3$, Figure 8, Figure S2). These findings were consistently observed in manual tracings (Figure S2) and model-predicted segmentations (Figure 8), with correction for age and head size (total intracranial volume). Moreover, the correlation is stronger in children (<18 years) than in adults. This effect replicates previous reports^{19,20}.

This correlation suggests that the susceptibility/intensity level may impact border detection due to partial volume effects, possibly introducing a bias in volume estimation. However, the volume versus susceptibility correlation persists, although it is partially attenuated, even after z-score intensity normalization (in a subset of $n=20$ images, before normalization: LDN ($r=0.288$), RDN ($r=0.197$); after normalization: LDN ($r=0.087$), RDN ($r=0.347$)).

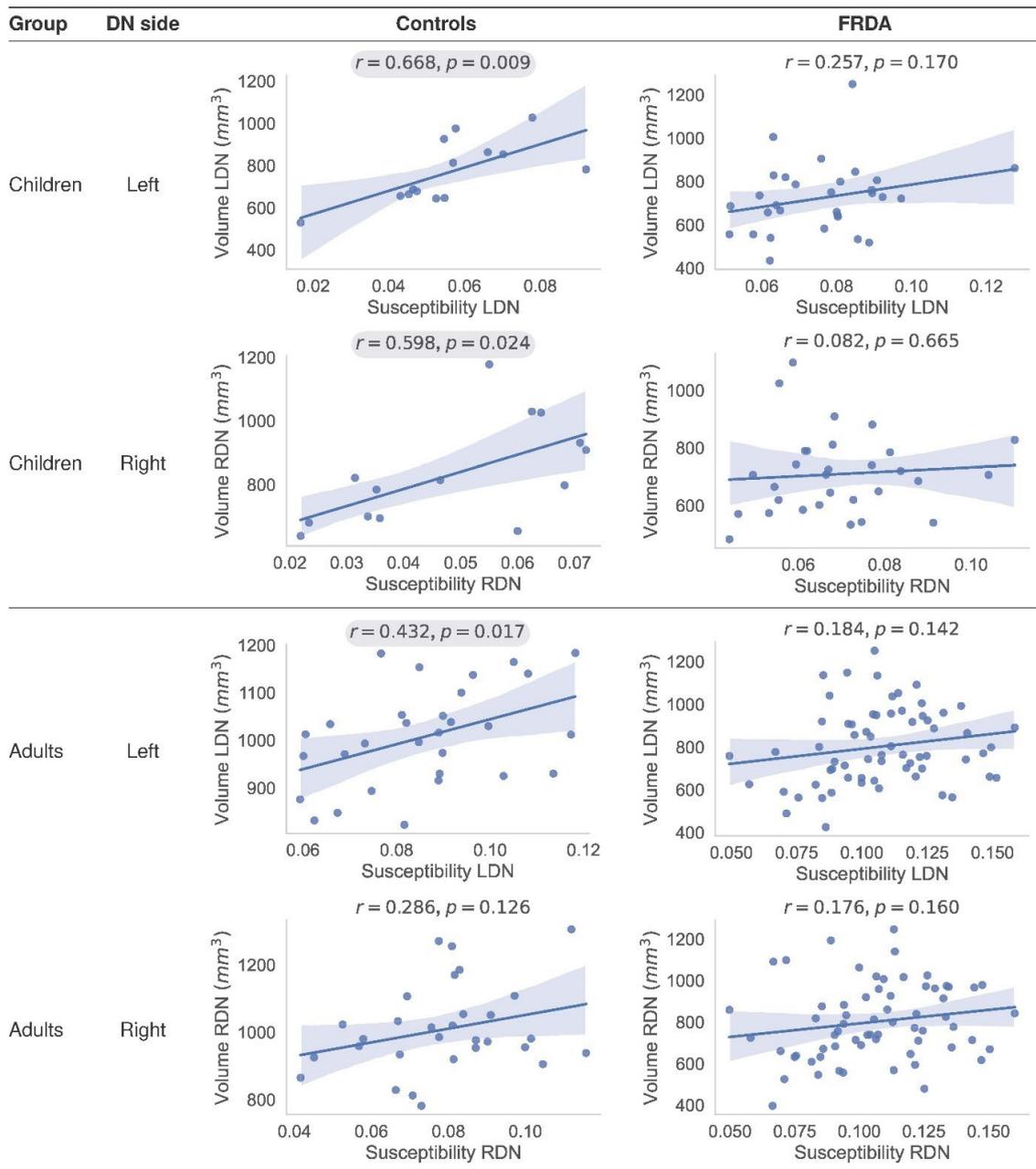


Figure 8. Model prediction results. Pearson's correlation coefficients and p-values for each group of individuals and DN side. Correlation coefficients with $p < 0.05$ are emphasized with a gray background. Children: subjects under 18 years of age. The volume estimations were corrected for age and head size (eTIV). FRDA: Friedreich's ataxia; DN: dentate nucleus; LDN: left DN; RDN: right DN.

4.6.2. Statistical Correction for Susceptibility-Driven Volume Bias

The impact of susceptibility variability on volume estimation may be estimated, and thus corrected for, using susceptibility and site (if applicable) as independent variables in multiple linear regression. A correction factor can be estimated from healthy control samples, and subsequently applied to patient data, to avoid confounds between true disease effects and quantification artifacts.

We have estimated a representation correction factor ($CF = 3636.84$) in a large cohort of demographically-diverse healthy individuals (age range 11-64 years, male to female ratio of 0.93:1) and several imaging protocols (including different imaging resolutions), which can be used to adjust data in future research studies using the following equation:

$$y_{adj} = y - CF(x - x_{median})$$

where y is the individual DN volume predicted by the segmentation model, x is the individual median DN susceptibility in ppm, x_{median} is the population estimate of median DN susceptibility across a dataset, and CF is the regression/correction factor. Note that this formula must be applied uniformly across both disease and control groups in a study to ensure accurate comparative analyses.

5. Discussion

Here we propose and develop an automated DN segmentation solution consisting of a two-step deep learning pipeline based on convolutional neural networks. Starting from a comprehensive dataset of QSM images sourced from several MRI centers and high-quality DN manual tracings by experienced annotators, we trained and tested multiple potential model architectures. All tested models delivered statistically superior results compared to MRICloud, the current state-of-the-art automated technique. Our final trained pipeline using the nnU-Net framework performed strongly during external validation, and we have introduced guidance regarding appropriately controlling for the influence of magnetic susceptibility on volume measurements. This work provides a robust, new, open-source tool (<https://github.com/art2mri/QSM-Cereb>) to the neuroscience community.

In the field of artificial intelligence, validation with external datasets is crucial to assess the generalization capacity of the proposed models⁶¹, although rarely reported⁶². This step is important to challenge the model across different acquisition protocols, populations, MRI sites, and scanner vendors, providing a valuable way to perform a real-world scenario analysis by evaluating model robustness and performance on unseen data⁶³. Our external validation provided strong support for the accuracy and robustness of the model, indicated a low risk of overfitting, and exceeded the accuracy of currently available approaches across a range of experimental conditions (different imaging acquisitions and QSM reconstruction protocols). When focusing on the reconstruction pipeline metrics, after aggregating the results, consistent Dice performance metrics were observed regardless of which QSM pipeline and reconstruction method was used (Table 8). A greater score for MEDI reconstructed images may derive from the fact that experienced raters were exposed to a majority of QSM images processed using this pipeline. We therefore provide a generalizable, fast and open-source solution for accurate and automated DN segmentation.

A combination of several factors contributes to the performance of the proposed model. First, we have gathered an extensive dataset that effectively captures the diversity found in real-world data, enabling good generalizability of the trained model. Additionally, we directly compared multiple DL frameworks to select the nnU-Net³⁴ as the optimal architecture for our medical imaging segmentation task among the architectures we tested. The nnU-Net incorporates a well-known architecture and a thorough set of data augmentation techniques. Moreover, our findings are aligned with analogous studies, such as the TotalSegmentator⁶⁴, wherein the nnU-Net stands out as a recognized solution for intricate tasks in the medical imaging domain. Also, our proposed pipeline is composed of two-step models, comprising localization followed by segmentation. In this way, we drive the focus of the second model into a restricted region of the brain, mitigating false positives. Certain voxels in the basal ganglia were predicted as DN in some outputs when the entire image was provided as input, potentially due to the presence of iron in these structures and slight shape similarity. The entire pipeline runs in less than 60 seconds using only the CPU and at most 15 seconds in GPU-based hardware (Intel Core i7 and Nvidia RTX 4070 12 GB).

This work provides a reliable automated solution to a task that has often been undertaken, to-date, by manual delineation (i.e., human hand-drawn segmentation). As such, research outcomes have traditionally relied on the availability of human expertise and are prone to errors resulting from task fatigue and inherent human variability. These considerations have limited the efficiency and accuracy of research outcomes, such as volumetric studies of DN volume in populations with neurological disorders that impact the cerebellum. The availability of a fast, accurate, and scalable automated segmentation tool opens new avenues of research into the DN structure in health and disease, including longitudinal natural history studies and clinical trials in disorders with known DN involvement, such as FRDA and SCA3⁶⁵.

In terms of biological findings, we replicated previously reported positive significant correlations between dentate nucleus volume and the magnetic susceptibility measures, in the healthy population^{19,20}. These effects were present both in the manual tracings and the prediction masks obtained using the trained models and could be explained by partial volume effects along the edges of the structure, whereby voxels that include a mix of tissue both inside and outside the DN would be more likely to be identified as being inside the DN in individuals with higher mean susceptibility intensity. The result persisted (albeit with a lesser magnitude) even after intensity normalization. Moreover, work by Li and colleagues¹⁹ also found that this positive association (with manual tracing) was present in the DN, but not in other midbrain or basal ganglia nuclei. If intensity-induced variability were indeed driving partial volume effects, this effect would be expected to be more severe in small structures such as the DN and the red nucleus¹⁹.

Regardless of the mechanism (biological or methodological) underlying this dependency between susceptibility and volume, it could be seen as an artifact that may be accounted for in DN volume assessments. This is further supported by the observation that, because susceptibility increases naturally with age, failure to correct for this dependency results in an apparent positive relationship between DN volume and age in the adult population (i.e., apparent DN growth over time)^{19,20}, which is biologically implausible. Care must also be taken when assessing populations with brain pathology, as the relationship between susceptibility and volume in neurological conditions may be a mix of true disease effects and the artifact described above. Neurological diseases, such as FRDA, SCA1, and

multiple system atrophy are characterized by increased DN susceptibility^{8,66}, which may lead to underestimation of DN volume loss. We, therefore, recommend statistically controlling for susceptibility levels when inferring volume effects to account for this artifact. In studies of normative populations, this can be achieved through simple regression (e.g., including a predictor of non-interest coding for susceptibility in regression models); for studies in pathological conditions, a correction factor estimated from normative data can be applied. We have reported a correction factor of 3636.84 for adjusting the total volume of DN, which represents the average of LDN and RDN volumes. However, it is important to note that slightly different correction factors were observed for LDN (3553.7188) and RDN (3422.2106). Consequently, future studies may select the appropriate correction factor based on their specific data and objective. Additionally, adjustments for other confounding variables, such as the impact of aging on DN volumes, susceptibility, or other measures, can be made either in sequential steps or by integrating all the confounding variables into a multiple regression model.

A limitation of this study lies in the fact that only three QSM reconstruction pipelines were used for generating the training images. Several QSM pipelines are available, including the possibility of combining a wide range of phase unwrapping, background field removal, and susceptibility mapping reconstruction methods and parameters. However, collecting such a diverse dataset would impose another layer of complexity on the study. As the QSM field continues to mature, it is likely that a narrower range of acquisition and reconstruction protocols will become the norm, reducing variability in this dimension. Notably, recent consensus papers have provided recommendations focused on clinical research in this area^{67,68}, which might be usefully explored in future investigations. Future work should also explore histological validation of the imaging segmentations to provide further improved ground truth. Similarly, higher resolution imaging protocols combined with increasingly accurate QSM reconstruction approaches offer the potential to isolate the gray matter ribbon of the DN from the central white matter core, which will enable tissue-type specific inferences.

In conclusion, our work provides state-of-the-art performance in automated DN segmentation from *in vivo* MRI based on extensive training and evaluation of a diverse dataset and methods. This outcome provides an important tool for characterizing cerebellar neuroanatomy in health and disease and biomarker discovery relevant for tracking disease progression and treatment efficacy in cerebellar disorders.

6. Data and Code Availability Statement

The patient MRI data are not publicly available due to privacy regulations. Access can be provided upon reasonable request to scientists in accordance with our Data Use and Access Policy. TRACK-FA and Enigma data might be provided upon request directed to Helena Bujalka (helena.bujalka@monash.edu) or Ian Harding (ian.harding@qimrberghofer.edu.au).

The source code of QSM deep-learning segmentation pipeline will be made publicly available on GitHub (<https://github.com/art2mri/QSM-Cereb>) upon acceptance.

7. Acknowledgments

The authors gratefully thank all of the individuals for their participation in the study and the Friedreich's Ataxia Research Alliance (FARA, USA) for their help with FRDA participant recruitment.

The authors thank and acknowledge the FARA, Takeda Pharmaceuticals Company, Novartis Gene Therapies, IXICO plc and PTC Therapeutics for support in funding and for their participation in the TRACK-FA Neuroimaging Consortium, including providing advice and feedback on study design, implementation and analysis of outcomes.

Study data were collected and managed using REDCap electronic data capture tools hosted and managed by Helix (Monash University, Australia).

Study data were collected, stored, and managed using XNAT, a software framework for managing neuroimaging laboratory data hosted by Monash (MXNAT) and Monash Biomedical Imaging (MBIXNAT) (Monash University, Australia).

The authors thank the research coordinators across all TRACK-FA sites for their assistance in participant recruitment and testing.

The following TRACK-FA collaborators opted to have their name acknowledged: Helena Bujalka, School of Psychological Sciences, The Turner Institute for Brain and Mental Health, Monash University, Clayton, Victoria, Australia; Manuela Corti, Powell Gene Therapy Centre, University of Florida, Gainesville, Florida, United States of America; Jennifer Farmer, Friedreich's Ataxia Research Alliance (FARA), Downingtown, Pennsylvania, United States of America; Myriam Rai, Friedreich's Ataxia Research Alliance (FARA), Downingtown, Pennsylvania, United States of America.

8. References

- 1 Dum, R. P. & Strick, P. L. An unfolded map of the cerebellar dentate nucleus and its projections to the cerebral cortex. *J Neurophysiol* **89**, 634-639 (2003). <https://doi.org/10.1152/jn.00626.2002>
- 2 Henschke, J. U. & Pakan, J. M. Disynaptic cerebrocerebellar pathways originating from multiple functionally distinct cortical areas. *Elife* **9** (2020). <https://doi.org/10.7554/eLife.59148>
- 3 Stoodley, C. J. & Schmahmann, J. D. Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex* **46**, 831-844 (2010). <https://doi.org/10.1016/j.cortex.2009.11.008>
- 4 Anteraper, S. A. *et al.* Intrinsic Functional Connectivity of Dentate Nuclei in Autism Spectrum Disorder. *Brain Connect* **9**, 692-702 (2019). <https://doi.org/10.1089/brain.2019.0692>
- 5 Olivito, G. *et al.* Cerebellar dentate nucleus functional connectivity with cerebral cortex in Alzheimer's disease and memory: a seed-based approach. *Neurobiol Aging* **89**, 32-40 (2020). <https://doi.org/10.1016/j.neurobiolaging.2019.10.026>
- 6 Xie, Y. J. *et al.* Functional connectivity of cerebellar dentate nucleus and cognitive impairments in patients with drug-naive and first-episode schizophrenia. *Psychiatry Res* **300**, 113937 (2021). <https://doi.org/10.1016/j.psychres.2021.113937>
- 7 Koeppen, A. H. The pathogenesis of spinocerebellar ataxia. *Cerebellum* **4**, 62-73 (2005). <https://doi.org/10.1080/14734220510007950>
- 8 Deistung, A. *et al.* Quantitative susceptibility mapping reveals alterations of dentate nuclei in common types of degenerative cerebellar ataxias. *Brain Commun* **4**, fcab306 (2022). <https://doi.org/10.1093/braincomms/fcab306>
- 9 Jaschke, D. *et al.* Age-related differences of cerebellar cortex and nuclei: MRI findings in healthy controls and its application to spinocerebellar ataxia (SCA6) patients. *Neuroimage* **270**, 119950 (2023). <https://doi.org/10.1016/j.neuroimage.2023.119950>
- 10 Stefanescu, M. R. *et al.* Structural and functional MRI abnormalities of cerebellar cortex and nuclei in SCA3, SCA6 and Friedreich's ataxia. *Brain* **138**, 1182-1197 (2015). <https://doi.org/10.1093/brain/awv064>
- 11 Ward, P. G. D. *et al.* Longitudinal evaluation of iron concentration and atrophy in the dentate nuclei in friedreich ataxia. *Mov Disord* **34**, 335-343 (2019). <https://doi.org/10.1002/mds.27606>
- 12 Diedrichsen, J. *et al.* Imaging the deep cerebellar nuclei: a probabilistic atlas and normalization procedure. *Neuroimage* **54**, 1786-1794 (2011). <https://doi.org/10.1016/j.neuroimage.2010.10.035>
- 13 Haacke, E. M., Xu, Y., Cheng, Y. C. & Reichenbach, J. R. Susceptibility weighted imaging (SWI). *Magn Reson Med* **52**, 612-618 (2004). <https://doi.org/10.1002/mrm.20198>
- 14 Ramos, P. *et al.* Iron levels in the human brain: a post-mortem study of anatomical region differences and age-related changes. *J Trace Elem Med Biol* **28**, 13-17 (2014). <https://doi.org/10.1016/j.jtemb.2013.08.001>
- 15 Halefoglou, A. M. & Yousem, D. M. Susceptibility weighted imaging: Clinical applications and future directions. *World J Radiol* **10**, 30-45 (2018). <https://doi.org/10.4329/wjr.v10.i4.30>
- 16 Liu, C., Li, W., Tong, K. A., Yeom, K. W. & Kuzminski, S. Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain. *J Magn Reson Imaging* **42**, 23-41 (2015). <https://doi.org/10.1002/jmri.24768>
- 17 Langkammer, C. *et al.* Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. *Neuroimage* **62**, 1593-1599 (2012). <https://doi.org/10.1016/j.neuroimage.2012.05.049>

- 18 Bonilha da Silva, C. *et al.* Dentate nuclei T2 relaxometry is a reliable neuroimaging marker in Friedreich's ataxia. *Eur J Neurol* **21**, 1131-1136 (2014). <https://doi.org/10.1111/ene.12448>
- 19 Li, G. *et al.* Age-dependent changes in brain iron deposition and volume in deep gray matter nuclei using quantitative susceptibility mapping. *Neuroimage* **269**, 119923 (2023). <https://doi.org/10.1016/j.neuroimage.2023.119923>
- 20 Zhang, Y. *et al.* Visualizing the deep cerebellar nuclei using quantitative susceptibility mapping: An application in healthy controls, Parkinson's disease patients and essential tremor patients. *Hum Brain Mapp* **44**, 1810-1824 (2023). <https://doi.org/10.1002/hbm.26178>
- 21 Li, X. *et al.* Multi-atlas tool for automated segmentation of brain gray matter nuclei and quantification of their magnetic susceptibility. *Neuroimage* **191**, 337-349 (2019). <https://doi.org/10.1016/j.neuroimage.2019.02.016>
- 22 Mori, S. *et al.* MRICloud: delivering high-throughput MRI neuroinformatics as cloud-based software as a service. *Computing in Science & Engineering* **18**, 21-35 (2016).
- 23 Georgiou-Karistianis, N. *et al.* A natural history study to track brain and spinal cord changes in individuals with Friedreich's ataxia: TRACK-FA study protocol. *PLoS One* **17**, e0269649 (2022). <https://doi.org/10.1371/journal.pone.0269649>
- 24 Khan, W. *et al.* Neuroinflammation in the Cerebellum and Brainstem in Friedreich Ataxia: An [18F]-FEMPA PET Study. *Mov Disord* **37**, 218-224 (2022). <https://doi.org/10.1002/mds.28825>
- 25 Chen, L. *et al.* Quantitative Susceptibility Mapping of Brain Iron and beta-Amyloid in MRI and PET Relating to Cognitive Performance in Cognitively Normal Older Adults. *Radiology* **298**, 353-362 (2021). <https://doi.org/10.1148/radiol.2020201603>
- 26 Fang, J., Bao, L., Li, X., van Zijl, P. C. M. & Chen, Z. Background field removal using a region adaptive kernel for quantitative susceptibility mapping of human brain. *J Magn Reson* **281**, 130-140 (2017). <https://doi.org/10.1016/j.jmr.2017.05.004>
- 27 Liu, T. *et al.* Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: comparison with COSMOS in human brain imaging. *Magn Reson Med* **66**, 777-783 (2011). <https://doi.org/10.1002/mrm.22816>
- 28 Liu, J. *et al.* Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map. *Neuroimage* **59**, 2560-2568 (2012). <https://doi.org/10.1016/j.neuroimage.2011.08.082>
- 29 Li, W., Wu, B. & Liu, C. Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. *Neuroimage* **55**, 1645-1656 (2011). <https://doi.org/10.1016/j.neuroimage.2010.11.088>
- 30 Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116-1128 (2006). <https://doi.org/10.1016/j.neuroimage.2006.01.015>
- 31 FSLeves v. 1.11.0 (Zenodo, 2024).
- 32 Ronneberger, O., Fischer, P. & Brox, T. in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. 234-241 (Springer).
- 33 Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. 424-432 (Springer).
- 34 Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**, 203-211 (2021). <https://doi.org/10.1038/s41592-020-01008-z>
- 35 Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z. & Tu, Z. in *Artificial intelligence and statistics*. 562-570 (Pmlr).
- 36 Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).

- 37 Chen, J. *et al.* Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306* (2021).
- 38 Hatamizadeh, A. *et al.* in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 574-584.
- 39 Hatamizadeh, A. *et al.* in *International MICCAI Brainlesion Workshop*. 272-284 (Springer).
- 40 Liu, Z. *et al.* in *Proceedings of the IEEE/CVF international conference on computer vision*. 10012-10022.
- 41 Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
- 42 Cardoso, M. J. *et al.* Monai: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701* (2022).
- 43 Han, S., Carass, A., He, Y. & Prince, J. L. Automatic cerebellum anatomical parcellation using U-Net with locally constrained optimization. *Neuroimage* **218**, 116819 (2020). <https://doi.org/10.1016/j.neuroimage.2020.116819>
- 44 Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033-2044 (2011). <https://doi.org/10.1016/j.neuroimage.2010.09.025>
- 45 Dice, L. R. Measures of the Amount of Ecologic Association between Species. *Ecology* **26**, 297-302 (1945). <https://doi.org/Doi.10.2307/1932409>
- 46 Sorenson, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analysis of vegetation on Danish commons. *Kong Dan Vidensk Selsk Biol Skr* **5**, 1-5 (1948).
- 47 Bertels, J. *et al.* Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. *Lect Notes Comput Sc* **11765**, 92-100 (2019). https://doi.org/10.1007/978-3-030-32245-8_11
- 48 Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing Images Using the Hausdorff Distance. *Ieee T Pattern Anal* **15**, 850-863 (1993). <https://doi.org/Doi.10.1109/34.232073>
- 49 Taha, A. A. & Hanbury, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* **15**, 29 (2015). <https://doi.org/10.1186/s12880-015-0068-x>
- 50 Jaccard, P. The distribution of the flora in the alpine zone. 1. *New phytologist* **11**, 37-50 (1912).
- 51 Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **15**, 155-163 (2016). <https://doi.org/10.1016/j.jcm.2016.02.012>
- 52 Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. Fsl. *Neuroimage* **62**, 782-790 (2012). <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- 53 van Bergen, J. M. *et al.* Quantitative Susceptibility Mapping Suggests Altered Brain Iron in Premanifest Huntington Disease. *AJNR Am J Neuroradiol* **37**, 789-796 (2016). <https://doi.org/10.3174/ajnr.A4617>
- 54 Chan, K. S. & Marques, J. P. SEPIA-Susceptibility mapping pipeline tool for phase images. *Neuroimage* **227**, 117611 (2021). <https://doi.org/10.1016/j.neuroimage.2020.117611>
- 55 Abdul-Rahman, H. S. *et al.* Fast and robust three-dimensional best path phase unwrapping algorithm. *Appl Opt* **46**, 6623-6635 (2007). <https://doi.org/10.1364/ao.46.006623>
- 56 Zhou, D., Liu, T., Spincemaille, P. & Wang, Y. Background field removal by solving the Laplacian boundary value problem. *NMR Biomed* **27**, 312-319 (2014). <https://doi.org/10.1002/nbm.3064>
- 57 Wei, H. *et al.* Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range. *NMR Biomed* **28**, 1294-1303 (2015). <https://doi.org/10.1002/nbm.3383>

- 58 Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022* (2016).
- 59 He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE international conference on computer vision*. 1026-1034.
- 60 Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
- 61 Ramspek, C. L., Jager, K. J., Dekker, F. W., Zoccali, C. & van Diepen, M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* **14**, 49-58 (2021). <https://doi.org/10.1093/ckj/sfaa188>
- 62 Collins, G. S. & Moons, K. G. M. Reporting of artificial intelligence prediction models. *Lancet* **393**, 1577-1579 (2019). [https://doi.org/10.1016/S0140-6736\(19\)30037-6](https://doi.org/10.1016/S0140-6736(19)30037-6)
- 63 de Hond, A. A. H. *et al.* Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med* **5**, 2 (2022). <https://doi.org/10.1038/s41746-021-00549-7>
- 64 Wasserthal, J. *et al.* TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. *Radiol Artif Intell* **5**, e230024 (2023). <https://doi.org/10.1148/ryai.230024>
- 65 Koeppen, A. H. The Neuropathology of Spinocerebellar Ataxia Type 3/Machado-Joseph Disease. *Adv Exp Med Biol* **1049**, 233-241 (2018). https://doi.org/10.1007/978-3-319-71779-1_11
- 66 Harding, I. H. *et al.* Tissue atrophy and elevated iron concentration in the extrapyramidal motor system in Friedreich ataxia: the IMAGE-FRDA study. *J Neurol Neurosurg Psychiatry* **87**, 1261-1263 (2016). <https://doi.org/10.1136/jnnp-2015-312665>
- 67 Committee, Q. S. M. C. O. *et al.* Recommended implementation of quantitative susceptibility mapping for clinical research in the brain: A consensus of the ISMRM electro-magnetic tissue properties study group. *Magn Reson Med* **91**, 1834-1862 (2024). <https://doi.org/10.1002/mrm.30006>
- 68 Oz, G. *et al.* MR Imaging in Ataxias: Consensus Recommendations by the Ataxia Global Initiative Working Group on MRI Biomarkers. *Cerebellum* (2023). <https://doi.org/10.1007/s12311-023-01572-y>

Supplementary Material

Title: Automated Dentate Nucleus Segmentation from QSM Images Using Deep Learning

Author names and affiliations: Diogo H Shiraishi^{1*}, Susmita Saha^{2,3*}, Isaac M Adanyeguh⁴, Sirio Coccozza⁵, Louise A Corben^{6,7,2}, Andreas Deistung^{8,9}, Martin B Delatycki⁶, Imis Dogan^{10,11}, William Gaetz¹², Nellie Georgiou-Karistianis², Simon Graf⁸, Marina Grisoli¹³, Pierre-Gilles Henry⁴, Gustavo M Jarola¹, James M Joers⁴, Christian Langkammer¹⁴, Christophe Lenglet⁴, Jiakun Li¹⁵, Camila C Lobo¹, Eric F Lock¹⁵, David R Lynch¹⁶, Thomas H Mareci¹⁷, Alberto R M Martinez¹, Serena Monti¹⁸, Anna Nigri¹³, Massimo Pandolfo¹⁹, Kathrin Reetz^{10,11}, Timothy P Roberts¹², Sandro Romanzetti¹⁰, David A Rudko^{19,20,21}, Alessandra Scaravilli⁵, Jörg B Schulz^{10,11}, S H Subramony²², Dagmar Timmann²³, Marcondes C França¹, Ian H Harding^{24,25†}, Thiago J R Rezende^{1†}, TRACK-FA Neuroimaging Consortium[‡]

1 Department of Neurology, School of Medical Sciences, University of Campinas (Unicamp), Campinas, Brazil

2 School of Psychological Sciences, The Turner Institute for Brain and Mental Health, Monash University, Clayton, Victoria, Australia

3 Department of Neuroscience, School of Translational Medicine, Monash University, Melbourne, Australia

4 Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, Minnesota, United States of America

5 Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy

6 Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, Parkville, Victoria, Australia

7 Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

8 University Clinic and Outpatient Clinic for Radiology, Department for Radiation Medicine, University Hospital Halle (Saale), University Medicine Halle, Halle (Saale), Germany

9 Halle MR Imaging Core Facility, Medical Faculty, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany

10 Department of Neurology, RWTH Aachen University, Aachen, Germany

11 JARA-BRAIN Institute Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich GmbH and RWTH Aachen University, Aachen, Germany

12 Department of Radiology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America

13 Department of Neuroradiology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

14 Department of Neurology, Medical University of Graz, Graz, Austria

15 Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, Minneapolis, Minnesota, United States of America

16 Department of Neurology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States of America

17 Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida, United States of America

18 Institute of Biostructures and Bioimaging, Italian National Research Council, Naples, Italy

19 Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada

20 McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, Montreal, Quebec, Canada

21 Department of Biomedical Engineering, McGill University, Montreal, Quebec, Canada

22 Department of Neurology and the Fixel Institute for Neurological Diseases, University of Florida, Gainesville, Florida, United States of America

23 Department of Neurology and Center for Translational and Behavioral Neuroscience (C-TNBS), Essen University Hospital, University of Duisburg-Essen, Essen, Germany

24 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

25 School of Translational Medicine, Monash University, Melbourne, Australia

* *Contributed equally*

† *Senior principal investigator*

‡ *see Acknowledgments for collaborators*

Deep learning model training specifications

The models explored in the architecture experimentation task included U-Net¹ with deep supervision (DS)², Swin UNETR³, and the nnU-Net framework⁴. U-Net with DS started with feature channels at 32, doubling up to a value of 512. Each level combined two 3x3x3 convolutions, followed by instance normalization⁵ and Leaky Rectified Linear Unit (Leaky ReLU)⁶ activation with a slope of 0.01. A Stochastic Gradient Descent (SGD) optimizer with a learning rate of 3×10^{-2} and a Nesterov momentum^{7,8} of 0.99 was used, combined with a Dice loss function for training. The learning rate was decayed using a polynomial function scheduler. Training lasted 600 epochs with a batch size of 1. The original Swin UNETR implementation was configured with a feature size of 24. An AdamW⁹ optimizer was selected with a learning rate of 1×10^{-4} , weight decay of 1×10^{-5} , $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. For the loss function, we used the Dice loss. A cosine annealing learning rate scheduler was set. The network was trained for 1000 epochs and a batch size of 1. Finally, the nnU-Net framework was trained using the default settings for the 3D full resolution trainer. For both U-Net with DS and Swin UNETR, a same set of augmentation techniques were selected. To replicate the nnU-Net augmentation pipeline, we incorporated random rotation, shear, scaling, translation, flipping, elastic deformation, and gamma contrast methods.

External Validation

Table S1. External validation datasets acquisition parameters.

Dataset	Scanner	Sequence	TR (ms)	TE1 (ms)	Δ TE (ms)	# of echoes	FoV (mm)	Image matrix (voxels)	Voxel size (mm)	Acquisition time
TRACK-FA McGill	3T Siemens Prisma	GRE	27	3.7	6	4	220 x 220 x 176	208 x 256 x 176	0.86 iso	7' 22"
Carlo Besta	3T Philips Achieva	GRE	40	4.5	5	7	240 x 180 140 axial slices	480 x 480 x 140	0.50 x 0.50 x 1.0	4' 23"
	3T Philips Achieva	GRE	40	5.4	5.2	7	224 x 224 140 axial slices	224 x 224 x 140	1.0 iso	8' 12"
Graz	3T Siemens Magnetom Trio	GRE	68	4.92	4.92	12	188 x 230 x 128	208 x 256 x 64	0.90 x 0.90 x 2.0	4' 51"

GRE: gradient recalled echo; iso: isotropic; TR: repetition time; TE: echo time; TE1: first echo time; FoV: field of view; iso: isotropic.

Table S2. External validation datasets demographics.

	McGill			Carlo Besta		Graz
	Controls	FRDA	Controls	Controls	MS	MS
Subjects	3	1	4	4	2	2
Age	29.8±4.4	41	64.5±10.0	43.5±17.7		
Sex (M/F)	2/1	0/1	1/3	1/1		

FRDA: Friedreich's ataxia; MS: multiple sclerosis.

Ethics approval was obtained for the CMRR study: IRB 1210M22281 (University of Minnesota).

Table S3. Acquisition protocols for CMRR dataset.

Dataset	Scanner	Sequence	TR (ms)	TE (ms)	Δ TE (ms)	# of echoes	FoV (mm)	Image matrix (voxels)	Voxel size (mm)	Acquisition time
CMRR	3T Siemens Magnetom Prisma Fit	GRE	86	20.48/30/45	-	3	205 x 186 x 32	232 x 256 x 32	0.80 iso	6' 59"
			65	7.26 (TE1)	5.00	8	205 x 185 x (30-36)	232 x 256 x (30-36)	0.80 iso	4' 40" - 5' 50"
			65	7.26 (TE1)	5.00	8	174 x 192 x 36	174 x 192 x 36	1.0 iso	4' 38"
			54	9.84 (TE1)	9.84	5	230 x 230 x 32	256 x 256 x 32	0.90 iso	4' 34"

MRI: magnetic resonance imaging; GRE: gradient recalled echo; iso: isotropic; TR: repetition time; TE: echo time; TE1: first echo time; FoV: field of view; iso: isotropic.

Table S4. Subject demographics for CMRR dataset.

	CMRR	
	Controls	FRDA
Subjects	3	16
Age	25.0±11.1	23.6±9.0
Sex (M/F)	2/1	6/10

FRDA: Friedreich's ataxia.

Segmentation Model Architectures Comparison

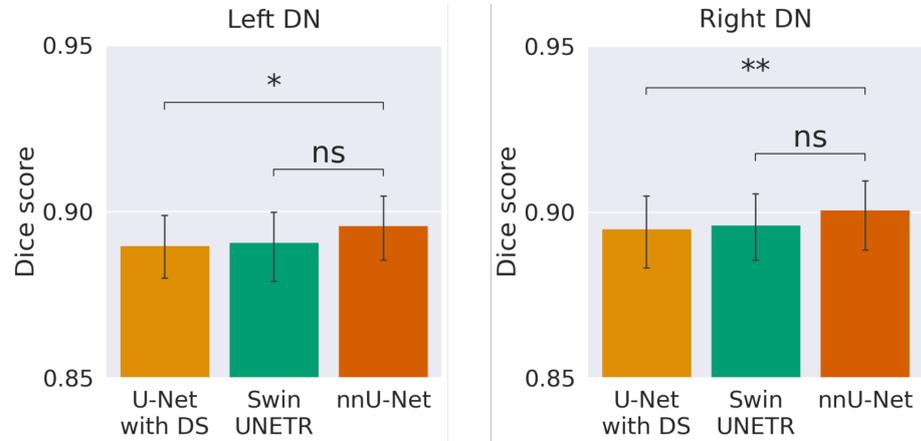


Figure S1. Bar plots of trained segmentation models. nnU-Net Dice score is statistically significantly higher than U-Net, and no significance was found when compared to Swin UNETR. ns: non-significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

DN Volume versus Mean Magnetic Susceptibility

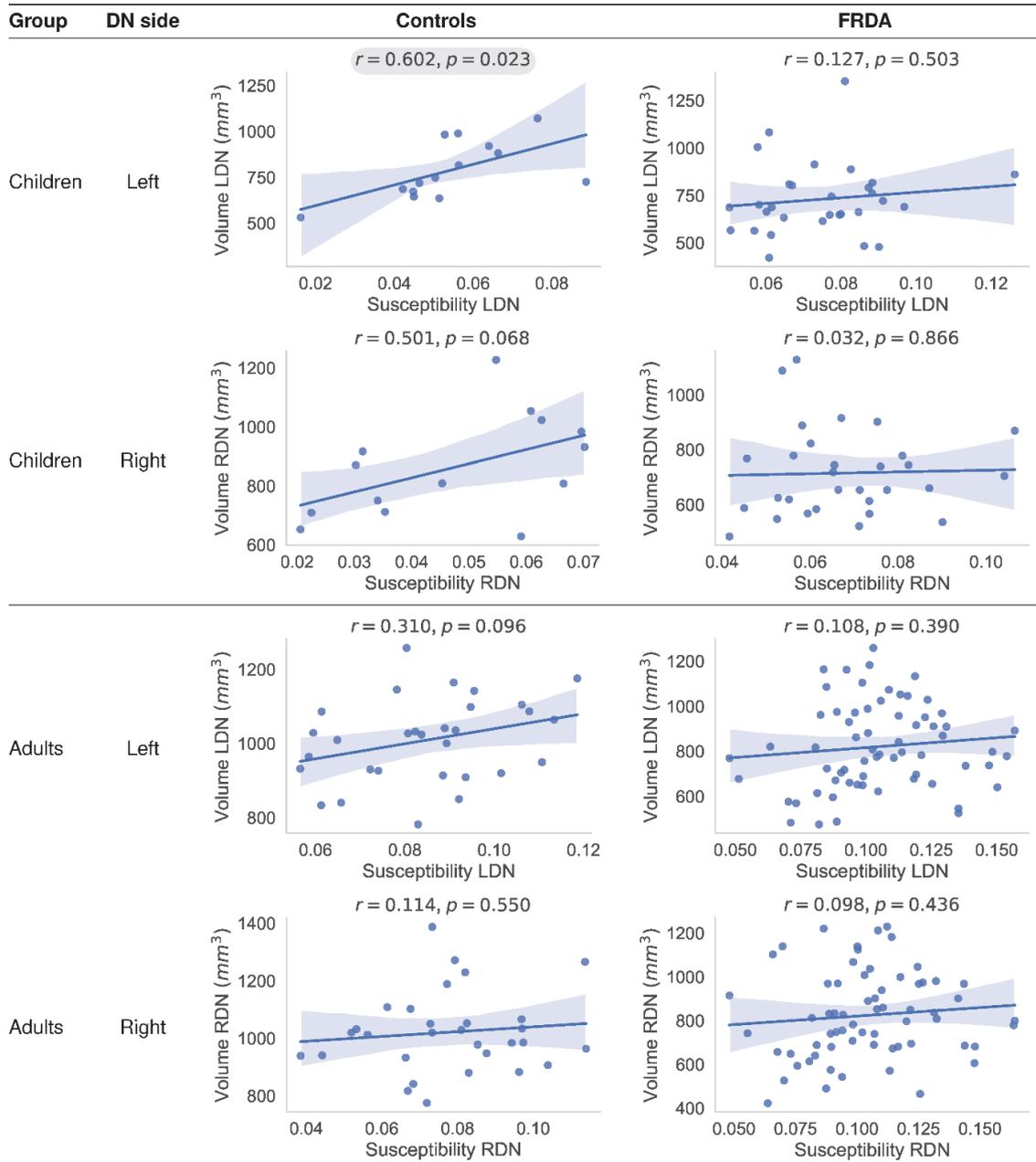


Figure S2. Manual segmentation results. Pearson's correlation coefficients and p-values for each group of individuals and DN side. Correlation coefficients with $p < 0.05$ are emphasized with a gray background. Children: subjects under 18 years of age. The volume estimations were corrected for age and head size (eTIV). FRDA: Friedreich's ataxia; DN: dentate nucleus; LDN: left DN; RDN: right DN; Vol: volume.

References

- 1 Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* 19. 424-432 (Springer).
- 2 Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z. & Tu, Z. in *Artificial intelligence and statistics*. 562-570 (Pmlr).
- 3 Hatamizadeh, A. *et al.* in *International MICCAI Brainlesion Workshop*. 272-284 (Springer).
- 4 Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**, 203-211 (2021). <https://doi.org/10.1038/s41592-020-01008-z>
- 5 Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022* (2016).
- 6 Maas, A. L., Hannun, A. Y. & Ng, A. Y. in *Proc. icml*. 3 (Atlanta, GA).
- 7 Sutskever, I., Martens, J., Dahl, G. & Hinton, G. in *International conference on machine learning*. 1139-1147 (PMLR).
- 8 Nesterov, Y. A method of solving a convex programming problem with convergence rate $O(1/k^{** 2})$. *Doklady Akademii Nauk SSSR* **269**, 543 (1983).
- 9 Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

6 Discussão

Os modelos de segmentação desenvolvidos mostraram resultados positivos em comparação com métodos existentes. Não somente isso é resultado da aplicação de modelos de redes neurais profundas adequadas para o aprendizado da tarefa de segmentação, mas também das técnicas de pré e pós-processamento e, principalmente, dos dados. A abordagem adotada neste trabalho versou, essencialmente, na coleta de dados de qualidade que representem minimamente uma variedade similar ao existente em mundo real. Isto é, supondo uma aplicação de uma solução proposta em produção ou mesmo em usos pontuais por grupos de pesquisa, não há como garantir uma mesma disponibilidade de equipamentos e uma similaridade de protocolos de aquisição, por exemplo. Portanto, é necessário que os dados reflitam essa diversidade, considerando também aspectos demográficos e aquisições tanto de indivíduos saudáveis quanto de portadores de alguma condição que afete estruturalmente os órgãos de interesse. Aliado aos dados, suas anotações de qualidade, seguindo um protocolo definido e realizado por anotadores experientes ou resultado de consenso, completa a tríade do insumo necessário para que os modelos gerados por treinamento desempenhem bem em diferentes contextos. A essa estratégia pode-se definir como centrada em dados, sendo este o cerne deste estudo em ambas as frentes.

Os resultados iniciais dos modelos treinados para segmentação do córtex cerebelar em imagens T1w foram positivos. Quando comparados com modelos existentes na literatura, nota-se um refinamento principalmente da precisão sobre as bordas externas do cerebelo. Fissuras e invaginações foram melhor delineadas e se aproximaram dos padrões de referência estabelecidos manualmente, ainda que em um conjunto restrito. Além disso, os resultados obtidos com o modelo treinado foram melhores do que aqueles gerados pelo CERES e pelo ACAPULCO e essa diferença encontrada foi estatisticamente significativa (ver resultados na seção 5.1). Durante o desenvolvimento dessa vertente do projeto de pesquisa, foi publicado o artigo do Ce-

rebNet (34) que, dentre os principais resultados, apresenta uma melhor performance em imagens com atrofia cerebelar. Além disso, os autores adotaram uma abordagem similar de dois estágios, com uma primeira localização para restrição espacial, o que acaba corroborando as escolhas deste projeto. Porém, ainda há oportunidade de melhoria, seja nas fissuras e na interface entre a substância branca e cinzenta. Isso reforça o interesse e justifica a manutenção dos esforços de incremento e refinamento dos conjuntos de dados anotados.

Ainda durante o estudo sobre a segmentação cerebelar um novo desafio foi proposto. Uma demanda para desenvolver um *pipeline* de segmentação automatizada do núcleo denteado para o projeto do TRACK-FA foi requerida, devido ao aumento das aquisições desse estudo longitudinal associado à custosa anotação manual do DN. Estando essa solicitação dentro do contexto do *grant* da FARA, ou seja, alinhado aos objetivos do projeto de pesquisa, sendo o TRACK-FA um projeto financiado pela FARA e havendo uma prévia disponibilidade de um conjunto de dados com algumas anotações manuais, o desafio foi aceito. Concomitantemente, no decorrer do desenvolvimento do modelo para o cerebelo, foi identificada a necessidade de reestruturar o processo de anotação. Foi identificada uma não conformidade com relação à qualidade das segmentações manuais dos colaboradores. Sendo o direcionador desse projeto a abordagem centrada em dados, um esforço adicional dedicado à melhoria das máscaras anotadas manualmente pelos especialistas foi possibilitado pelo suporte financeiro da FARA. Atribuindo-se um retorno financeiro por unidade com a qualidade validada, está sendo possível reunir anotações que atendam a requisitos mínimos dado o protocolo de segmentação estabelecido. Ao ser atingido uma volumetria significativa, o desenvolvimento do estudo será retomado.

Para a segmentação do DN, uma solução automatizada foi proposta e desenvolvida, composta por um *pipeline* de aprendizado profundo em duas etapas, baseado em redes neurais convolucionais. A partir de um conjunto abrangente de imagens QSM obtidas de vários centros de ressonância magnética e anotações manuais de alta qualidade do DN feitas por especialistas, diversas arquiteturas de modelos potenciais foram treinadas e avaliadas. Todos os modelos testados apresentaram resultados estatisticamente superiores comparados ao MRICloud, a técnica automatizada de última geração atualmente disponível. O *pipeline* final treinado usando o *framework* nnU-

Net apresentou um desempenho robusto em validação externa, e foram introduzidas orientações sobre como controlar adequadamente a influência da suscetibilidade magnética nas medidas de volume. Este trabalho fornece uma nova ferramenta robusta e de código aberto¹ para a comunidade científica.

Uma combinação de vários fatores contribui para o desempenho do modelo proposto. Primeiro, reunimos um conjunto de dados extenso que captura efetivamente a diversidade encontrada nos dados do mundo real, permitindo que os modelos treinados atingissem boa capacidade de generalização. Além disso, foram comparadas diretamente uma variedade de arquiteturas de aprendizado profundo até a seleção do nnU-Net (15) como a arquitetura ideal. O nnU-Net incorpora uma arquitetura bem conhecida e um conjunto completo de técnicas de *data augmentation*. Além disso, a escolha está alinhada com estudos análogos, como o TotalSegmentator (83), nos quais o nnU-Net se destaca como uma solução reconhecida para tarefas complexas no domínio de imagens médicas. Também, nosso *pipeline* proposto é composto por uma abordagem de dois modelos, compreendendo localização seguida de segmentação. Desta forma, o foco do segundo modelo é dedicado para uma região restrita do cérebro, mitigando falsos positivos. Antes da implementação dessa restrição espacial, alguns *voxels* nos gânglios da base foram previstos como DN em algumas saídas quando a imagem inteira foi fornecida como entrada, possivelmente devido à presença de ferro nessas estruturas e à ligeira semelhança de forma. Todo o *pipeline* é executado em menos de 60 segundos usando apenas a CPU e no máximo 15 segundos quando acelerado por GPU (Intel Core i7 e Nvidia RTX 4070 12 GB).

No campo da inteligência artificial, a validação com conjuntos de dados externos é crucial para avaliar a capacidade de generalização dos modelos propostos (128), embora seja reportada com frequência em estudos (129). Esta etapa é importante para desafiar o modelo através de diferentes protocolos de aquisição, populações, centros de aquisição de ressonância magnética e fabricantes e modelos de *scanners* MRI, proporcionando um modo de realizar uma análise de cenário real ao avaliar a robustez e o desempenho do modelo em dados não vistos em treinamento (130). A validação externa forneceu evidências da precisão e robustez do modelo proposto, indicando um baixo risco de *overfitting* e superando o MRICloud em uma

¹ <https://github.com/art2mri/DentateSeg>

variedade de condições analisadas (diferentes aquisições de imagem e protocolos de reconstrução QSM). Após agregar os resultados, foram observadas métricas de desempenho de Dice consistentes, independentemente do *pipeline* QSM e método de reconstrução utilizados. Um melhor desempenho para imagens reconstruídas pelo MEDI pode ser consequência do fato de que avaliadores experientes foram expostos à maioria das imagens QSM processadas usando este *pipeline*. Portanto, fornecemos uma solução com capacidade de generalização razoável, execução rápida e de código aberto para segmentação precisa e automatizada do DN.

Este trabalho fornece uma solução automatizada confiável para uma tarefa que, até o momento, tem sido frequentemente realizada por esforço manual (ou seja, segmentação desenhada à mão, frente à necessidade). Como tal, os resultados das pesquisas têm tradicionalmente dependido da disponibilidade de expertise humana e são propensos a erros resultantes da fadiga dessa tarefa e da variabilidade inerente ao anotador. Essas considerações têm limitado a eficiência e a precisão dos resultados de pesquisa, como estudos volumétricos do DN em populações com doenças neurodegenerativas que afetam o cerebelo. A disponibilidade de uma ferramenta de segmentação automatizada rápida, precisa e escalável abre novas possibilidades de pesquisa sobre a estrutura do DN, incluindo estudos longitudinais da história natural e ensaios clínicos, como, por exemplo, FRDA e SCA3 (131).

Em termos de achados biológicos, foram encontradas correlações positivas significativas previamente relatadas entre o volume do núcleo denteado e as medidas de suscetibilidade magnética nos indivíduos saudáveis (60, 61). Esses efeitos estavam presentes tanto nos resultados de segmentação manual quanto nas máscaras preditas obtidas usando os modelos treinados e podem ser explicados pelos efeitos de volume parcial principalmente sobre as bordas da estrutura, onde *voxels* que incluem um entrelaçamento de tecido tanto dentro quanto fora do DN teriam mais probabilidade de serem identificados como estando dentro do DN em indivíduos com maior intensidade média de suscetibilidade. Esse resultado persistiu (embora com menor magnitude) mesmo após a normalização da intensidade dos *voxels*. Além disso, o trabalho de Li e colegas (60) também encontrou que essa associação positiva (com a anotação manual) estava presente no DN, além do núcleo rubro, mas não em outros núcleos do mesencéfalo ou dos demais gânglios da base. Se a variabilidade induzida

pela intensidade estivesse realmente impulsionando os efeitos de volume parcial, esse efeito seria, de fato, esperado ser mais severo em estruturas pequenas, como o DN e o núcleo rubro (60).

Independentemente do mecanismo, biológico ou metodológico, subjacente a essa relação entre suscetibilidade e volume, isso poderia ser visto como um artefato que deve ser levado em consideração nas avaliações de volume do DN. Isso é corroborado pela observação de que, como a suscetibilidade aumenta naturalmente com a idade, a falha em corrigir essa dependência resulta em uma aparente relação positiva entre volume do DN e idade na população adulta (ou seja, aparente crescimento do DN ao longo do tempo) (60, 61), o que é biologicamente implausível. Deve-se também ter cautela ao avaliar populações com patologias cerebrais, pois a relação entre suscetibilidade e volume em condições neurológicas pode ser uma combinação de efeitos relativos à doença e ao artefato descrito acima. Doenças neurológicas, como FRDA, SCA1 e MSA, são caracterizadas pelo aumento da suscetibilidade do DN (50, 132), o que pode levar à subestimação da perda de volume do DN. Portanto, recomenda-se o controle estatístico dos níveis de suscetibilidade ao inferir efeitos de volume para levar em conta esse artefato. Em estudos populacionais saudáveis, isso pode ser realizado através de uma regressão simples. Já para estudos em condições patológicas, um fator de correção estimado a partir de dados normativos pode ser aplicado. Foi calculado um fator de correção para ajustar o volume total do DN, que representa a média dos volumes do LDN e RDN deste estudo. No entanto, é importante salientar que foram observados fatores de correção ligeiramente diferentes para LDN e RDN (ver resultados na seção 5.2). Consequentemente, estudos futuros podem selecionar o fator de correção apropriado, derivados de seus objetivos e dados específicos. Além disso, ajustes para outras variáveis confundidoras, como o impacto da idade nos volumes do DN, suscetibilidade ou outras medidas, podem ser feitos em etapas sequenciais ou integrando todas as variáveis confundidoras em um modelo de regressão múltipla.

Há um *trade-off* na escolha entre ter um *pipeline* completo ou simplificado. Alguns modelos de segmentação como o CerebNet citam como vantagem a não necessidade de pré-processar o dado antes de submetê-lo a processamento. Ou seja, uma correção de inomogeneidade de campo ou um simples registro para um *template*

como o MNI apenas para padronizar a orientação seriam desnecessárias, uma vez que o modelo releva isso pela natureza do aprendizado das redes neurais treinadas com uma variabilidade suficiente. Por outro lado, se há a disponibilidade de métodos que contribuem para redução de ruídos ou melhoria em geral dos dados apresentados aos modelos, há de se avaliar o custo dessa inclusão. De fato, ferramentas de segmentação de primeira geração, como os baseados em atlas, complexos em operações matriciais, demandam muitos minutos ou horas para um único processamento. Nas soluções baseadas em redes neurais, como o ACAPULCO e o CerebNet, tempos inferiores a minutos são mencionados, ainda mais quando aceleradas com poder computacional disponibilizado por GPUs. Ou seja, uma inclusão de passos de pré-processamento compatíveis com essa ordem de grandeza de tempo de processamento são razoavelmente adequados, sem impacto relevante de custo temporal e com potencial de melhoria da qualidade dos dados apresentados aos modelos.

No contexto de segmentação de imagens médicas, diversas arquiteturas e *frameworks* foram propostas no decorrer das últimas décadas. É evidente que a rede U-Net detém contínua relevância, dada sua performance. Porém, novas propostas têm surgido, como as derivadas de *Transformers* (ver seção 4.1.5). Mais recentemente, o advento do mecanismo de atenção tem revolucionado o campo de processamento de linguagem natural (do inglês, *natural language processing*, NLP), dado seu fundamental emprego em *Transformers* em *large language models* (LLMs), por exemplo. Esses avanços foram acompanhados pelo surgimento de modelos fundacionais, que são treinados em extensas massas de dados e podem ser adaptados para outras tarefas. Não restrito a texto, essa evolução produziu aplicações em imagens, sendo o Segment Anything Model (SAM) um exemplo com resultados notáveis na tarefa de segmentação de diversos domínios. Um passo seguinte é o modelo MedSAM que é uma aplicação similar ao SAM, porém para o contexto de imagens médicas. As arquiteturas baseadas em modelos fundacionais não foram objeto de estudo deste trabalho, porém devem ser considerados na investigação posterior na continuidade dos modelos para o cerebelo.

A qualidade da imagem QSM é diretamente impactada pela metodologia adotada na sua reconstrução. Isto é, a partir das aquisições de magnitude e fase, uma sequência de passos para obtenção final do mapa de suscetibilidade magnética

deve ser aplicada. Para cada etapa de reconstrução existem diversos métodos disponíveis na literatura e cada um deles possui um conjunto de parâmetros configuráveis. Ou seja, não somente tem-se a variabilidade dada pelo equipamento de ressonância, pelo protocolo de aquisição, é necessário considerar esse passo final. Neste trabalho, três métodos de reconstrução foram adotados, seja na etapa de treinamento dos modelos ou na validação externa. Porém, os três métodos aplicados compreendem uma pequena parcela dos algoritmos existentes, o que é uma limitação deste estudo. De fato, é possível uma grande variedade de combinação de métodos para cada etapa de *phase unwrapping*, *background field removal* e reconstrução do mapeamento de suscetibilidade. No entanto, coletar e processar um conjunto de dados tão vasto imporia uma camada de complexidade adicional ao projeto. Assim, faz-se adequada uma análise aprofundada do impacto de demais metodologias de reconstrução. Contudo, o campo de estudo de QSM continua a progredir e é provável que haja um estreitamento de opções otimizadas de protocolos de aquisição e reconstrução tornando-se padrões, reduzindo a variabilidade dessa etapa. A exemplo disso, recentemente artigos de consenso têm proposto recomendações direcionadas à pesquisa clínica (133, 134). A partir disso, pode ser proposto um *pipeline* estendido. Nesse caso, não mais a entrada da ferramenta deve ser uma imagem QSM, mas, eventualmente, o par de magnitude e fase. Assim, à sequência de pré-processamento pode ser incluída a reconstrução QSM otimizada com parâmetros ajustados a uma melhor qualidade que propicie mais controle sobre a variabilidade de imagens.

Por fim, há a possibilidade de exploração de estudo histológico. Assim, haveria chance de validar as segmentações comparando as imagens aos tecidos diretamente e, assim, estabelecer anotações *ground truth* ainda mais precisas. De modo similar, protocolos de aquisição de alta resolução combinados com os *pipelines* otimizados para QSM oferecem um potencial de isolar o *ribbon* de substância cinzenta do DN de seu núcleo central de substância branca, permitindo inferências específicas por tipo de tecido.

7 Conclusão

Este trabalho apresentou duas aplicações de aprendizado profundo para a segmentação em imagens de ressonância magnética *in vivo*, focando, inicialmente, no córtex cerebelar em imagens T1w e, posteriormente, nos núcleos denteados em imagens QSM. Os resultados preliminares do modelo para o córtex do cerebelo foram promissores, especialmente em relação ao delineamento das fissuras, das interfaces entre substância branca e cinzenta e das pequenas invaginações do líquido nas porções externas da estrutura foliar. O desenvolvimento será retomado conforme o refinamento das máscaras manuais por anotadores experientes, que visa incrementar a qualidade dos dados e garantir a variabilidade necessária para generalização dos modelos. Para os núcleos denteados, foi desenvolvido um *pipeline* validado com dados externos, permitindo a segmentação precisa a partir de imagens QSM. Essa oportunidade surgiu no decorrer das atividades do projeto inicial e resulta de uma demanda do TRACK-FA Neuroimaging Consortium e, assim, já está sendo utilizada nas análises longitudinais desse estudo.

8 Referências Bibliográficas

- [1] Schmahmann JD. The cerebellum and cognition. *Neuroscience Letters*. 2019 Jan;688:62-75.
- [2] Schmahmann JD, Caplan D. Cognition, emotion and the cerebellum. *Brain*. 2006 Feb;129(2):290-2.
- [3] Klaes A, Reckziegel E, Franca MC, Rezende TJR, Vedolin LM, Jardim LB, et al. MR Imaging in Spinocerebellar Ataxias: A Systematic Review. *American Journal of Neuroradiology*. 2016 Aug;37(8):1405-12.
- [4] Olivito G, Serra L, Marra C, Di Domenico C, Caltagirone C, Toniolo S, et al. Cerebellar dentate nucleus functional connectivity with cerebral cortex in Alzheimer's disease and memory: a seed-based approach. *Neurobiology of Aging*. 2020 May;89:32-40.
- [5] Streng ML, Krook-Magnuson E. The cerebellum and epilepsy. *Epilepsy & Behavior*. 2021 Aug;121:106909.
- [6] Xie YJ, Xi YB, Cui LB, Guan MZ, Li C, Wang ZH, et al. Functional connectivity of cerebellar dentate nucleus and cognitive impairments in patients with drug-naive and first-episode schizophrenia. *Psychiatry Research*. 2021 Jun;300:113937.
- [7] Parmar K, Fonov VS, Naegelin Y, Amann M, Wuerfel J, Collins DL, et al. Regional Cerebellar Volume Loss Predicts Future Disability in Multiple Sclerosis Patients. *Cerebellum*. 2022 Aug;21(4):632-46.
- [8] Reetz K, Rodríguez-Labrada R, Dogan I, Mirzazade S, Romanzetti S, Schulz JB, et al. Brain atrophy measures in preclinical and manifest spinocerebellar ataxia type 2. *Ann Clin Transl Neurol*. 2018 Feb;5(2):128-37.
- [9] Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. *NeuroImage*. 2006 Oct;33(1):127-38.
- [10] Romero JE, Coupé P, Giraud R, Ta VT, Fonov V, Park MTM, et al. CERES: A new cerebellum lobule segmentation method. *NeuroImage*. 2017 Feb;147:916-24.
- [11] Baldaçara L, Currie S, Hadjivassiliou M, Hoggard N, Jack A, Jackowski AP, et al. Consensus Paper: Radiological Biomarkers of Cerebellar Diseases. *Cerebellum*. 2015 Apr;14(2):175-96.

- [12] Dogan I, Romanzetti S, Didszun C, Mirzazade S, Timmann D, Saft C, et al. Structural characteristics of the central nervous system in Friedreich ataxia: an in vivo spinal cord and brain MRI study. *J Neurol Neurosurg Psychiatry*. 2019 May;90(5):615-7.
- [13] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. vol. 9351. Cham: Springer International Publishing; 2015. p. 234-41.
- [14] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv*; 2016.
- [15] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021 Feb;18(2):203-11.
- [16] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al.. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv*; 2021.
- [17] Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. UNETR: Transformers for 3D Medical Image Segmentation. In: *2022 IEEE CVF Winter Conf. Appl. Comput. Vis. WACV*. Waikoloa, HI, USA: IEEE; 2022. p. 1748-58.
- [18] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth H, Xu D. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. *arXiv*; 2022.
- [19] Leiner HC, Leiner AL, Dow RS. Cognitive and language functions of the human cerebellum. *Trends in Neurosciences*. 1993 Nov;16(11):444-7.
- [20] Stoodley CJ, Schmahmann JD. Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex*. 2010 Jul;46(7):831-44.
- [21] Strick PL, Dum RP, Fiez JA. Cerebellum and Nonmotor Function. *Annu Rev Neurosci*. 2009 Jun;32(1):413-34.
- [22] Rezende TJR, Silva CB, Yassuda CL, Campos BM, D'Abreu A, Cendes F, et al. Longitudinal magnetic resonance imaging study shows progressive pyramidal and callosal damage in Friedreich's ataxia. *Movement Disorders*. 2016 Jan;31(1):70-8.
- [23] Ward PGD, Harding IH, Close TG, Corben LA, Delatycki MB, Storey E, et al. Longitudinal evaluation of iron concentration and atrophy in the dentate nuclei in friedreich ataxia. *Movement Disorders*. 2019 Mar;34(3):335-43.

- [24] Bonilha Da Silva C, Bergo FPG, D'Abreu A, Cendes F, Lopes-Cendes I, França MC. Dentate nuclei T2 relaxometry is a reliable neuroimaging marker in Friedreich's ataxia. *Euro J of Neurology*. 2014 Aug;21(8):1131-6.
- [25] Harding IH, Chopra S, Arrigoni F, Boesch S, Brunetti A, Coccozza S, et al. Brain Structure and Degeneration Staging in Friedreich Ataxia: MAGNETIC RESONANCE IMAGING Volumetrics from the ENIGMA-ATAXIA Working Group. *Annals of Neurology*. 2021 Oct;90(4):570-83.
- [26] Kerestes R, Perry A, Vivash L, O'Brien TJ, Alvim MKM, Arienzo D, et al. Patterns of subregional cerebellar atrophy across epilepsy syndromes: An ENIGMA-Epilepsy study. *Epilepsia*. 2024 Apr;65(4):1072-91.
- [27] Jacobs HIL, Hopkins DA, Mayrhofer HC, Bruner E, Van Leeuwen FW, Raaijmakers W, et al. The cerebellum in Alzheimer's disease: evaluating its role in cognitive decline. *Brain*. 2018 Jan;141(1):37-47.
- [28] Laidi C, d'Albis MA, Wessa M, Linke J, Phillips ML, Delavest M, et al. Cerebellar volume in schizophrenia and bipolar I disorder with and without psychotic features. *Acta Psychiatr Scand*. 2015 Mar;131(3):223-33.
- [29] Anteraper SA, Guell X, Taylor HP, D'Mello A, Whitfield-Gabrieli S, Joshi G. Intrinsic Functional Connectivity of Dentate Nuclei in Autism Spectrum Disorder. *Brain Connectivity*. 2019 Nov;9(9):692-702.
- [30] Koeppen A. The pathogenesis of spinocerebellar ataxia. *MCER*. 2005 Apr;4(1):62-73.
- [31] Dum RP, Strick PL. An Unfolded Map of the Cerebellar Dentate Nucleus and its Projections to the Cerebral Cortex. *Journal of Neurophysiology*. 2003 Jan;89(1):634-9.
- [32] Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, et al. Three-Dimensional MRI Atlas of the Human Cerebellum in Proportional Stereotaxic Space. *NeuroImage*. 1999 Sep;10(3):233-60.
- [33] Han S, Carass A, He Y, Prince JL. Automatic cerebellum anatomical parcellation using U-Net with locally constrained optimization. *NeuroImage*. 2020 Sep;218:116819.
- [34] Faber J, Kügler D, Bahrami E, Heinz LS, Timmann D, Ernst TM, et al. CerebNet: A fast and reliable deep-learning pipeline for detailed cerebellum subsegmentation. *NeuroImage*. 2022 Dec;264:119703.
- [35] Li X, Chen L, Kutten K, Ceritoglu C, Li Y, Kang N, et al. Multi-atlas tool for automated segmentation of brain gray matter nuclei and quantification of their magnetic susceptibility. *NeuroImage*. 2019 May;191:337-49.

- [36] Mori S, Wu D, Ceritoglu C, Li Y, Kolasny A, Vaillant MA, et al. MRICloud: Delivering High-Throughput MRI Neuroinformatics as Cloud-Based Software as a Service. *Comput Sci Eng*. 2016 Sep;18(5):21-35.
- [37] Diedrichsen J, Zotow E. Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. *PLoS ONE*. 2015 Jul;10(7):e0133402.
- [38] Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR Images Using Registration for Signal Averaging:. *Journal of Computer Assisted Tomography*. 1998 Mar;22(2):324-33.
- [39] Ta VT, Giraud R, Collins DL, Coupé P. Optimized PatchMatch for Near Real Time and Accurate Label Fusion. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. vol. 8675. Cham: Springer International Publishing; 2014. p. 105-12.
- [40] Carass A, Cuzzocreo JL, Han S, Hernandez-Castillo CR, Rasser PE, Ganz M, et al. Comparing fully automated state-of-the-art cerebellum parcellation from magnetic resonance images. *NeuroImage*. 2018 Dec;183:150-72.
- [41] Tustison NJ, Avants BB, Cook PA, Yuanjie Zheng, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging*. 2010 Jun;29(6):1310-20.
- [42] Coupé P, Manjón JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*. 2011 Jan;54(2):940-54.
- [43] Kansal K, Yang Z, Fishman AM, Sair HI, Ying SH, Jodynak BM, et al. Structural cerebellar correlates of cognitive and motor dysfunctions in cerebellar degeneration. *Brain*. 2017 Jan:aww327.
- [44] Sörös P, Wölk L, Bantel C, Bräuer A, Klawonn F, Witt K. Replicability, Repeatability, and Long-term Reproducibility of Cerebellar Morphometry. *Cerebellum*. 2021 Jun;20(3):439-53.
- [45] Iglesias JE, Cheng-Yi Liu, Thompson PM, Zhuowen Tu. Robust Brain Extraction Across Datasets and Comparison With Publicly Available Methods. *IEEE Trans Med Imaging*. 2011 Sep;30(9):1617-34.
- [46] Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*. 2011 Feb;54(3):2033-44.
- [47] Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*. 2014 Oct;99:166-79.

- [48] Henschel L, Conjeti S, Estrada S, Diers K, Fischl B, Reuter M. FastSurfer - A fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*. 2020 Oct;219:117012.
- [49] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole Brain Segmentation. *Neuron*. 2002 Jan;33(3):341-55.
- [50] Deistung A, Jäschke D, Draganova R, Pfaffenrot V, Hulst T, Steiner KM, et al. Quantitative susceptibility mapping reveals alterations of dentate nuclei in common types of degenerative cerebellar ataxias. *Brain Commun*. 2022 Jan;4(1):fcab306.
- [51] Jäschke D, Steiner KM, Chang DI, Claaßen J, Uslar E, Thieme A, et al. Age-related differences of cerebellar cortex and nuclei: MRI findings in healthy controls and its application to spinocerebellar ataxia (SCA6) patients. *NeuroImage*. 2023 Apr;270:119950.
- [52] Stefanescu MR, Dohnalek M, Maderwald S, Thürling M, Minnerop M, Beck A, et al. Structural and functional MRI abnormalities of cerebellar cortex and nuclei in SCA3, SCA6 and Friedreich's ataxia. *Brain*. 2015 May;138(5):1182-97.
- [53] Kanda T, Ishii K, Kawaguchi H, Kitajima K, Takenaka D. High Signal Intensity in the Dentate Nucleus and Globus Pallidus on Unenhanced T1-weighted MR Images: Relationship with Increasing Cumulative Dose of a Gadolinium-based Contrast Material. *Radiology*. 2014 Mar;270(3):834-41.
- [54] Diedrichsen J, Maderwald S, Küper M, Thürling M, Rabe K, Gizewski ER, et al. Imaging the deep cerebellar nuclei: A probabilistic atlas and normalization procedure. *NeuroImage*. 2011 Feb;54(3):1786-94.
- [55] Haacke EM, Xu Y, Cheng YCN, Reichenbach JR. Susceptibility weighted imaging (SWI). *Magnetic Resonance in Med*. 2004 Sep;52(3):612-8.
- [56] Ramos P, Santos A, Pinto NR, Mendes R, Magalhães T, Almeida A. Iron levels in the human brain: A post-mortem study of anatomical region differences and age-related changes. *Journal of Trace Elements in Medicine and Biology*. 2014 Jan;28(1):13-7.
- [57] Liu C, Li W, Tong KA, Yeom KW, Kuzminski S. Susceptibility-weighted imaging and quantitative susceptibility mapping in the brain: Brain Susceptibility Imaging and Mapping. *J Magn Reson Imaging*. 2015 Jul;42(1):23-41.
- [58] Langkammer C, Schweser F, Krebs N, Deistung A, Goessler W, Scheurer E, et al. Quantitative susceptibility mapping (QSM) as a means to measure brain iron? A post mortem validation study. *NeuroImage*. 2012 Sep;62(3):1593-9.
- [59] He N, Langley J, Huddlestone DE, Ling H, Xu H, Liu C, et al. Improved Neuroimaging Atlas of the Dentate Nucleus. *Cerebellum*. 2017 Dec;16(5-6):951-6.

- [60] Li G, Tong R, Zhang M, Gillen KM, Jiang W, Du Y, et al. Age-dependent changes in brain iron deposition and volume in deep gray matter nuclei using quantitative susceptibility mapping. *NeuroImage*. 2023 Apr;269:119923.
- [61] Zhang Y, Huang P, Wang X, Xu Q, Liu Y, Jin Z, et al. Visualizing the deep cerebellar nuclei using quantitative susceptibility mapping: An application in healthy controls, Parkinson's disease patients and essential tremor patients. *Human Brain Mapping*. 2023 Mar;44(4):1810-24.
- [62] Tang X, Oishi K, Faria AV, Hillis AE, Albert MS, Mori S, et al. Bayesian Parameter Estimation and Segmentation in the Multi-Atlas Random Orbit Model. *PLoS ONE*. 2013 Jun;8(6):e65591.
- [63] Tang X, Yoshida S, Hsu J, Huisman TAGM, Faria AV, Oishi K, et al. Multi-Contrast Multi-Atlas Parcellation of Diffusion Tensor Imaging of the Human Brain. *PLoS ONE*. 2014 May;9(5):e96985.
- [64] Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*. 2015 Aug;24(1):205-19.
- [65] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015 May;521(7553):436-44.
- [66] Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. *npj Digit Med*. 2021 Jan;4(1):5.
- [67] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv*; 2017.
- [68] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. *J Digit Imaging*. 2017 Aug;30(4):449-59.
- [69] Cui Z, Yang J, Qiao Y. Brain MRI segmentation with patch-based CNN approach. In: 2016 35th Chin. Control Conf. CCC. Chengdu, China: IEEE; 2016. p. 7026-31.
- [70] Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. *Magnetic Resonance Imaging*. 2020 Jan;51(1):175-82.
- [71] Milletari F, Ahmadi SA, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *Computer Vision and Image Understanding*. 2017 Nov;164:92-102.
- [72] Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018 Nov;19(6):1236-46.
- [73] Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*. 1959 Oct;148(3):574-91.

- [74] Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth Int. Conf. 3D Vis. 3DV. Stanford, CA, USA: IEEE; 2016. p. 565-71.
- [75] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR. Las Vegas, NV, USA: IEEE; 2016. p. 770-8.
- [76] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al.. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. arXiv; 2021.
- [77] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al.. Attention Is All You Need. arXiv; 2023.
- [78] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al.. Segment Anything. arXiv; 2023.
- [79] Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nat Commun. 2024 Jan;15(1):654.
- [80] Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, et al. The Medical Segmentation Decathlon. Nat Commun. 2022 Jul;13(1):4128.
- [81] Kazerooni AF, Khalili N, Liu X, Haldar D, Jiang Z, Anwar SM, et al.. The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). arXiv; 2024.
- [82] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans Med Imaging. 2015 Oct;34(10):1993-2024.
- [83] Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, et al. Total-Segmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiology: Artificial Intelligence. 2023 Sep;5(5):e230024.
- [84] DeepLearningAI. A Chat with Andrew on MLOps: From Model-centric to Data-centric AI. Youtube; 2021. <https://www.youtube.com/watch?v=06-AZXmWHjo>.
- [85] Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. Journal of Neuroscience Methods. 2016 May;264:47-56.
- [86] Fonov V, Evans A, McKinstry R, Almlí C, Collins D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 2009 Jul;47:S102.
- [87] Burk K. Cognition in hereditary ataxia. Cerebellum. 2007 Sep;6(3):280.

- [88] Dürr A, Cossee M, Agid Y, Campuzano V, Mignard C, Penet C, et al. Clinical and Genetic Abnormalities in Patients with Friedreich's Ataxia. *N Engl J Med*. 1996 Oct;335(16):1169-75.
- [89] Abrahão A, Pedroso JL, Braga-Neto P, Bor-Seng-Shu E, De Carvalho Aguiar P, Barsottini OGP. Milestones in Friedreich ataxia: more than a century and still learning. *Neurogenetics*. 2015 Jul;16(3):151-60.
- [90] Schmitz-Hübsch T, Du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, et al. Scale for the assessment and rating of ataxia: Development of a new clinical scale. *Neurology*. 2006 Jun;66(11):1717-20.
- [91] Schmitz-Hübsch T, Fimmers R, Rakowicz M, Rola R, Zdzienicka E, Fancellu R, et al. Responsiveness of different rating instruments in spinocerebellar ataxia patients. *Neurology*. 2010 Feb;74(8):678-84.
- [92] Koeppe AH, Mazurkiewicz JE. Friedreich Ataxia: Neuropathology Revised. *J Neuropathol Exp Neurol*. 2013 Feb;72(2):78-90.
- [93] Adanyeguh IM, Perlberg V, Henry PG, Rinaldi D, Petit E, Valabregue R, et al. Autosomal dominant cerebellar ataxias: Imaging biomarkers with high effect sizes. *NeuroImage: Clinical*. 2018;19:858-67.
- [94] Piccinin CC, Rezende TJR, De Paiva JLR, Moysés PC, Martinez ARM, Cendes F, et al. A 5-YEAR Longitudinal Clinical and Magnetic Resonance Imaging Study in Spinocerebellar Ataxia Type 3. *Movement Disorders*. 2020 Sep;35(9):1679-84.
- [95] Albert M, Barrantes-Freer A, Lohrberg M, Antel JP, Prineas JW, Palkovits M, et al. Synaptic pathology in the cerebellar dentate nucleus in chronic multiple sclerosis. *Brain Pathology*. 2017 Nov;27(6):737-47.
- [96] Cocozza S, Petracca M, Mormina E, Buyukturkoglu K, Podranski K, Heinig MM, et al. Cerebellar lobule atrophy and disability in progressive MS. *J Neurol Neurosurg Psychiatry*. 2017 Dec;88(12):1065-72.
- [97] Faber J, Giordano I, Jiang X, Kindler C, Spottke A, Acosta-Cabronero J, et al. Prominent White Matter Involvement in Multiple System Atrophy of Cerebellar Type. *Movement Disorders*. 2020 May;35(5):816-24.
- [98] Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JAD, et al. Multi-parametric neuroimaging reproducibility: A 3-T resource study. *NeuroImage*. 2011 Feb;54(4):2854-66.
- [99] Bollmann S. ISMRM 2021 Educational Talk on Quantitative Susceptibility Mapping; 2021. <https://mri.sbollmann.net/index.php/2021/04/30/ismrm-2021-educational-talk-on-quantitative-susceptibility-mapping/>.

- [100] Georgiou-Karistianis N, Corben LA, Reetz K, Adanyeguh IM, Corti M, Deelchand DK, et al. A natural history study to track brain and spinal cord changes in individuals with Friedreich's ataxia: TRACK-FA study protocol. *PLoS ONE*. 2022 Nov;17(11):e0269649.
- [101] Khan W, Corben LA, Bilal H, Vivash L, Delatycki MB, Egan GF, et al. Neuroinflammation in the Cerebellum and Brainstem in Friedreich Ataxia: An [^{18}F]-FEMPA PET Study. *Movement Disorders*. 2022 Jan;37(1):218-24.
- [102] Chen L, Soldan A, Oishi K, Faria A, Zhu Y, Albert M, et al. Quantitative Susceptibility Mapping of Brain Iron and β -Amyloid in MRI and PET Relating to Cognitive Performance in Cognitively Normal Older Adults. *Radiology*. 2021 Feb;298(2):353-62.
- [103] Van Bergen JMG, Hua J, Unschuld PG, Lim IAL, Jones CK, Margolis RL, et al. Quantitative Susceptibility Mapping Suggests Altered Brain Iron in Premanifest Huntington Disease. *American Journal of Neuroradiology*. 2016 May;37(5):789-96.
- [104] Fang J, Bao L, Li X, Van Zijl PCM, Chen Z. Background field removal using a region adaptive kernel for quantitative susceptibility mapping of human brain. *Journal of Magnetic Resonance*. 2017 Aug;281:130-40.
- [105] Liu J, Liu T, De Rochefort L, Ledoux J, Khalidov I, Chen W, et al. Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map. *NeuroImage*. 2012 Feb;59(3):2560-8.
- [106] Liu T, Liu J, De Rochefort L, Spincemaille P, Khalidov I, Ledoux JR, et al. Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: Comparison with COSMOS in human brain imaging. *Magnetic Resonance in Med*. 2011 Sep;66(3):777-83.
- [107] Li W, Wu B, Liu C. Quantitative susceptibility mapping of human brain reflects spatial variation in tissue composition. *NeuroImage*. 2011 Apr;55(4):1645-56.
- [108] Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*. 2006 Jul;31(3):1116-28.
- [109] Paul McCarthy. FSLeys; 2024. Zenodo.
- [110] Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979;86(2):420-8.
- [111] Chan KS, Marques JP. SEPIA—Susceptibility mapping pipeline tool for phase images. *NeuroImage*. 2021 Feb;227:117611.

- [112] Abdul-Rahman HS, Gdeisat MA, Burton DR, Lalor MJ, Lilley F, Moore CJ. Fast and robust three-dimensional best path phase unwrapping algorithm. *Appl Opt*. 2007 Sep;46(26):6623.
- [113] Zhou D, Liu T, Spincemaille P, Wang Y. Background field removal by solving the Laplacian boundary value problem. *NMR Biomed*. 2014 Mar;27(3):312-9.
- [114] Wei H, Dibb R, Zhou Y, Sun Y, Xu J, Wang N, et al. Streaking artifact reduction for quantitative susceptibility mapping of sources with large dynamic range: Streaking Artifact Reduction for QSM. *NMR Biomed*. 2015 Oct;28(10):1294-303.
- [115] Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012 Aug;62(2):782-90.
- [116] Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-Supervised Nets. *arXiv*; 2014.
- [117] Pérez-García F, Sparks R, Ourselin S. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*. 2021 Sep;208:106236.
- [118] Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al.. MONAI: An open-source framework for deep learning in healthcare. *arXiv*; 2022.
- [119] Avants B, Epstein C, Grossman M, Gee J. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. 2008 Feb;12(1):26-41.
- [120] Péran P, Cherubini A, Luccichenti G, Hagberg G, Démonet JF, Rascol O, et al. Volume and iron content in basal ganglia and thalamus. *Human Brain Mapping*. 2009 Aug;30(8):2667-75.
- [121] Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945 Jul;26(3):297-302.
- [122] Sørensen TJ. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *K Dan Vidensk Selsk*. 1948;5:1-34.
- [123] Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. vol. 11765. Cham: Springer International Publishing; 2019. p. 92-100.
- [124] Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Machine Intell*. Sept/1993;15(9):850-63.

- [125] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015 Dec;15(1):29.
- [126] Jaccard P. The distribution of the flora in the alpine zone. *New Phytologist*. 1912 Feb;11(2):37-50.
- [127] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*. 2016 Jun;15(2):155-63.
- [128] Ramspek CL, Jager KJ, Dekker FW, Zoccali C, Van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J*. 2021 Feb;14(1):49-58.
- [129] Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *The Lancet*. 2019 Apr;393(10181):1577-9.
- [130] De Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, Van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *npj Digit Med*. 2022 Jan;5(1):2.
- [131] Koeppen AH. The Neuropathology of Spinocerebellar Ataxia Type 3/Machado-Joseph Disease. In: Nóbrega C, Pereira De Almeida L, editors. *Polyglutamine Disorders*. vol. 1049. Cham: Springer International Publishing; 2018. p. 233-41.
- [132] Harding IH, Raniga P, Delatycki MB, Stagnitti MR, Corben LA, Storey E, et al. Tissue atrophy and elevated iron concentration in the extrapyramidal motor system in Friedreich ataxia: the IMAGE-FRDA study. *J Neurol Neurosurg Psychiatry*. 2016 Nov;87(11):1261-3.
- [133] Öz G, Coccozza S, Henry PG, Lenglet C, Deistung A, Faber J, et al. MR Imaging in Ataxias: Consensus Recommendations by the Ataxia Global Initiative Working Group on MRI Biomarkers. *Cerebellum*. 2023 Jun.
- [134] QSM Consensus Organization Committee, Bilgic B, Costagli M, Chan KS, Duyn J, Langkammer C, et al. Recommended implementation of quantitative susceptibility mapping for clinical research in the brain: A consensus of the ISMRM electro-magnetic tissue properties study group. *Magnetic Resonance in Med*. 2024 May;91(5):1834-62.

9 Anexos

9.1 Parecer do Comitê de Ética

Documento de aprovação na forma de parecer consubstanciado do Comitê de Ética em Pesquisa, CAAE 83241318.3.1001.5404.



PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: ESTUDO LONGITUDINAL CLÍNICO E DE NEUROIMAGEM EM PACIENTES E INDIVÍDUOS SOB RISCO DE DOENÇAS HEREDODEGENERATIVAS TARDIAS

Pesquisador: Marcondes Cavalcante França Junior

Área Temática: Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP;);

Versão: 11

CAAE: 83241318.3.1001.5404

Instituição Proponente: Hospital de Clínicas - UNICAMP

Patrocinador Principal: FUNDACAO DE AMPARO A PESQUISA DO ESTADO DE SAO PAULO

DADOS DO PARECER

Número do Parecer: 4.916.667

Apresentação do Projeto:

As informações contidas nos campos "Apresentação do Projeto", "Objetivo da Pesquisa" e "Avaliação dos Riscos e Benefícios" foram obtidas dos documentos apresentados para apreciação ética e das informações inseridas pelo Pesquisador Responsável do estudo na Plataforma Brasil.

RESUMO

Doenças neurodegenerativas são doenças neurológicas heterogêneas que podem envolver demência, doenças do neurônio motor e distúrbios do movimento. Nessas doenças há destruição progressiva e irreversível de neurônios e gradativamente o paciente perde suas funções motoras, fisiológicas e/ou sua capacidade cognitiva. Essas doenças podem cursar com síndrome demencial como manifestação clínica principal, com síndrome demencial associada a distúrbios do movimento ou exclusivamente com distúrbios do movimento. Entre elas existem a Ataxia, Doença de Huntington, Parkinson, Esclerose Lateral Amiotrófica e outras. Essas doenças constituem um grupo amplo e prevalente de enfermidades crônicas caracterizadas: 1. Pela morte prematura e acelerada de populações neuronais tanto no sistema nervoso central quanto periférico, e 2. Pelo substrato genético. Podem se manifestar tanto na infância quanto na vida adulta, mas invariavelmente tem curso progressivo culminando com grave debilidade motora e/ou cognitiva

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: 4.916.667

após alguns anos de doença. O foco deste projeto recairá sobre 3 das principais formas de doenças hereditárias motoras, que serão discutidas a seguir: a Esclerose lateral amiotrófica familiar (FALS), as paraparesias espásticas hereditárias (PEH) e as ataxias espinocerebelares (SCA). Todas de herança autossômica dominante e início tardio. estudo amplo voltado para a caracterização clínica e de imagem na FALS, PEH e SCA. Em especial, focado na identificação de possíveis biomarcadores e para a determinação da história natural de cada uma das enfermidades, pontos fundamentais para o desenho de ensaios clínicos. No estudo realizado com os pacientes (indivíduos sintomáticos), o foco recairá na identificação das áreas/estruturas danificadas em cada um dos subtipos de FALS, PEH e SCA; no padrão de progressão deste dano ao longo do tempo; e na correlação com a deterioração clínica. No estudo realizado com os indivíduos sob risco, procuraremos esclarecer quanto dura e como progride o período pre-sintomático em cada uma destas enfermidades. O objetivo é caracterizar do ponto de vista clínico e de neuroimagem os pacientes com FALS, PEH e SCA assim como os indivíduos sob risco. Haverão dois braços do estudo. O primeiro com pacientes sintomáticos que serão submetidos a avaliação clínica e por RM a cada 12 meses por um período total de 5 anos. O segundo será formado por indivíduos sob risco para FALS, PEH e SCA. Aqui serão convidados filhos e irmãos assintomáticos de pacientes portadores de formas autossômicas dominantes de FALS, PEH e SCA. Neste braço do estudo, teremos tanto indivíduos portadores da mutação quanto indivíduos não portadores da mutação). Ambos (pré-sintomáticos e normais) passarão por avaliações clínica e de RM anual ao longo de 5 anos. O acompanhamento dos indivíduos normais será realizado para permitir a comparação do padrão evolutivo com os indivíduos pré-sintomáticos.

Objetivo da Pesquisa:

Objetivo Primário:

Objetivo geral- Caracterizar do ponto de vista clínico e de neuroimagem os pacientes com FALS, PEH e SCA assim como os indivíduos sob risco.

Objetivo Secundário:

- Determinar a gravidade da disfunção clínica nos pacientes com FALS, PEH e SCA através do uso de escalas padronizadas.
- Caracterizar o padrão e gravidade do comprometimento do sistema nervoso central nos pacientes com FALS, PEH e SCA através de estudo de ressonância magnética.
- Caracterizar o padrão evolutivo dos parâmetros clínicos e de neuroimagem nos pacientes com FALS, PEH e SCA.
- Caracterizar o padrão e gravidade do comprometimento do sistema nervoso central nos

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: 4.916.667

indivíduos pré-sintomáticos com FALS, PEH e SCACaracterizar o padrão evolutivo dos parâmetros clínicos e de neuroimagem nos indivíduos pré-sintomáticos com FALS, PEH e SCA.

- Desenvolver uma análise de Segmentação Cerebelar Automatizada e Otimizada Usando Deep-Learning.

Avaliação dos Riscos e Benefícios:

Riscos: O único relacionado ao exame de ressonância magnética é o ruído intermitente durante os primeiros 15 minutos. Depois disso o ruído será muito menor. O pessoal técnico providenciará tapa-ouvidos para me deixar mais confortável. Pessoas claustrofóbicas podem se sentir desconfortáveis durante o exame. Não existem efeitos nocivos associados com a ressonância magnética dentro das condições utilizadas atualmente. Os riscos associados a coleta de material biológico são dor e manchas roxas (equimoses) no local da coleta do sangue. Trata-se de uma coleta de sangue geralmente da veia do braço que será realizada por profissional treinado e habilitado para realizar esse procedimento.

Benefícios: Sua participação no estudo poderá permitir uma avaliação mais detalhada das alterações no cérebro e na medula espinhal de pacientes com essa doença. Os resultados desse estudo podem, a longo prazo, oferecer vantagens para os pacientes, possibilitando um melhor entendimento da doença e diagnóstico mais precoce. Entretanto não há nenhum benefício ao senhor em participar.

Comentários e Considerações sobre a Pesquisa:

Essa versão é a 8ª solicitação de emenda ao projeto aprovado pelo Parecer Consubstanciado CEP n.o 2.534.828 de 9 de março de 2018.

Apresentou a seguinte justificativa: "A presente emenda solicita inclusão de um novo pesquisador à equipe de pesquisa que irá desenvolver um subprojeto, com o objetivo secundário de desenvolver uma análise de Segmentação Cerebelar Automatizada e Otimizada Usando Deep-Learning. Não há alteração de procedimentos, TCLE, riscos ou objetivos do estudo."

A equipe de pesquisa é composta por:

Marcondes Cavalcante França Junior

Thiago Junqueira Ribeiro de Rezende

Maria Fernanda Ribeiro Bittar

Guilherme Soares de Oliveira Wertheimer

Gabriel Da Silva Schmitt

Jean Levi Ribeiro de Paiva

Tauana Bernardes Leoni

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: 4.916.667

Diogo Hideki Shiraishi
 Joyce Macedo da Silva
 Katiane Raisa Servelhere
 Camila Callegari Piccinin
 Ueslei Andrade do Nascimento
 Carelis Del Valle Gonzalez Salazar
 Fabrício Castro de Borba
 Yves Glauber Silva dos Santos

Considerações sobre os Termos de apresentação obrigatória:

Foram anexados os seguintes documentos:

- 1- PB_INFORMAÇÕES_BÁSICAS_1807154_E8.pdf
- 2- Sub_projeto_DIOGO.pdf
- 3- Projeto_MRI_doencas_neurodegenerativas_emenda8.doc
- 4- Carta_CEP.pdf

Recomendações:

A Comissão Nacional de Ética em Pesquisa (Conep), do Conselho Nacional de Saúde (CNS) orienta a adoção das diretrizes do Ministério da Saúde (MS) decorrentes da pandemia causada pelo Coronavírus SARS-CoV-2 (Covid-19), com o objetivo de minimizar os potenciais riscos à saúde e a integridade dos participantes de pesquisas e pesquisadores.

De acordo com carta circular da CONEP intitulada "ORIENTAÇÕES PARA CONDUÇÃO DE PESQUISAS E ATIVIDADE DOS CEP DURANTE A PANDEMIA PROVOCADA PELO CORONAVÍRUS SARS-COV-2 (COVID-19)" publicada em 09/05/2020, referente ao item II. "Orientações para Pesquisadores":

- Aconselha-se a adoção de medidas para a prevenção e gerenciamento de todas as atividades de pesquisa, garantindo-se as ações primordiais à saúde, minimizando prejuízos e potenciais riscos, além de prover cuidado e preservar a integridade e assistência dos participantes e da equipe de pesquisa.
- Em observância às dificuldades operacionais decorrentes de todas as medidas impostas pela pandemia do SARS-CoV-2 (COVID- 19), é necessário zelar pelo melhor interesse do participante da pesquisa, mantendo-o informado sobre as modificações do protocolo de pesquisa que possam afetá-lo, principalmente se houver ajuste na condução do estudo, cronograma ou plano de trabalho.
- Caso sejam necessários a suspensão, interrupção ou o cancelamento da pesquisa, em decorrência dos riscos imprevisíveis aos participantes da pesquisa, por causas diretas ou

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: 4.916.667

indiretas, caberá aos investigadores a submissão de notificação para apreciação do Sistema CEP/Conep.

- Nos casos de ensaios clínicos, é permitida, excepcionalmente, a tramitação de emendas concomitantes à implementação de modificações/alterações no protocolo de pesquisa, visando à segurança do participante da pesquisa, assim como dos demais envolvidos no contexto da pesquisa, evitando-se, ainda, quando aplicável, a interrupção no tratamento dos participantes da pesquisa. Eventualmente, na necessidade de modificar o Termo de Consentimento Livre e Esclarecido (TCLE), o pesquisador deverá proceder com o novo consentimento, o mais breve possível.

Conclusões ou Pendências e Lista de Inadequações:

Emenda considerada 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).

- 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.

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: 4.916.667

- 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 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_1807154_E8.pdf	10/08/2021 23:50:20		Aceito
Outros	Sub_projeto_DIOGO.pdf	10/08/2021 22:56:49	Marcondes Cavalcante França Junior	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_MRI_doencas_neurodegenerativas_emenda8.doc	10/08/2021 22:56:34	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Carta_CEP.pdf	10/08/2021 22:56:15	Marcondes Cavalcante França Junior	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_MRI_doencas_neurodegenerativas_emenda7.pdf	09/07/2021 22:51:52	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Carta_emenda7signed.pdf	09/07/2021 22:42:13	Marcondes Cavalcante França Junior	Aceito

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: 4.916.667

Declaração de Pesquisadores	Carta_Resposta.pdf	08/05/2021 12:41:18	Maria Fernanda Ribeiro Bittar	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE15_17anos.docx	10/04/2021 20:03:39	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_CONTROL E_responsavel.pdf	10/04/2021 19:59:30	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_CONTROL E.pdf	10/04/2021 19:59:22	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_CASO_responsavel_emenda6.pdf	10/04/2021 19:59:11	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_controle14_17anos.docx	10/04/2021 19:58:22	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_11a14anos.docx	10/04/2021 19:58:14	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_7a10anos.docx	10/04/2021 19:58:06	Marcondes Cavalcante França Junior	Aceito
Outros	Ficha_clinica_SCA3_V1.doc	10/04/2021 19:57:33	Marcondes Cavalcante França Junior	Aceito
Outros	Ficha_clinica_SCA3_V0.doc	10/04/2021 19:57:19	Marcondes Cavalcante França Junior	Aceito
Outros	COMPROVANTE_R4_Gabriel.pdf	10/04/2021 19:56:54	Marcondes Cavalcante França Junior	Aceito
Outros	COMPROVANTE_doutorado_Joyce.jpg	10/04/2021 19:56:36	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Cartaz_divulgacao.pdf	10/04/2021 19:56:00	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Adendo_GabrielSilva_C.pdf	10/04/2021 19:55:47	Marcondes Cavalcante França	Aceito

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: 4.916.667

Declaração de Pesquisadores	Adendo_GabrielSilva_C.pdf	10/04/2021 19:55:47	Junior	Aceito
Outros	Projeto_doutorado_V1_Joyce_Macedo.pdf	10/04/2021 19:55:03	Marcondes Cavalcante França Junior	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_MRI_doencas_neurodegenerativas_emenda6.pdf	10/04/2021 19:54:37	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Solicitacao_emenda6_assinado.pdf	10/04/2021 19:54:12	Marcondes Cavalcante França Junior	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_MRI_doencas_neurodegenerativas_2020_emenda5.doc	04/11/2020 18:36:26	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Emenda_inclusao_pesquisadores.docx	04/11/2020 18:35:39	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_sob_risco_EM.docx	10/04/2020 19:47:38	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_sob_risco_EM.pdf	10/04/2020 19:47:11	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_pacientes_EM.pdf	10/04/2020 19:47:03	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Heredodegenerativas_pacientes_EM.docx	10/04/2020 19:46:53	Marcondes Cavalcante França Junior	Aceito
Declaração de Manuseio Material Biológico / Biorepositório / Biobanco	Biorrepositorio.pdf	10/04/2020 19:46:16	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Justificativa_da_emenda.pdf	10/04/2020 19:46:03	Marcondes Cavalcante França Junior	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_MRI_doencas_neurodegenerativas_2020_emenda3.pdf	24/02/2020 21:54:49	Marcondes Cavalcante França Junior	Aceito
Declaração de Pesquisadores	Retorno_ambulatorio.pdf	24/02/2020 21:51:43	Marcondes Cavalcante França Junior	Aceito

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: 4.916.667

Declaração de Pesquisadores	Carta_ao_CEP_EM3.pdf	24/02/2020 21:48:31	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	relato_de_caso_TCLE.pdf	24/02/2020 21:40:07	Marcondes Cavalcante França Junior	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	relato_de_caso_TCLE.docx	24/02/2020 21:15:03	Maria Fernanda Ribeiro Bittar	Aceito
Folha de Rosto	Folha_de_rosto.pdf	09/02/2018 16:48:24	Maria Fernanda Ribeiro Bittar	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CAMPINAS, 19 de Agosto de 2021

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

9.2 Comprovante de Submissão do Manuscrito

Comprovante de submissão do manuscrito para periódico indexado.

Detailed Status Information

Manuscript #	NPJDIGITALMED-11218
Current Revision #	0
Submission Date	17th Jun 24 21:50:37
Current Stage	Manuscript Received
Title	Automated Dentate Nucleus Segmentation from QSM Images Using Deep Learning
Running Head	Automated segmentation of the dentate nuclei
Manuscript Type	Article
Collections	N/A
Corresponding Author	Dr Thiago Rezende (thiago.jrezende@gmail.com) (Department of Neurology, School of Medical Sciences, University of Campinas (Unicamp), Campinas, Brazil)
Contributing Authors	Mr Diogo Shiraiishi, Dr Susmita Saha, Dr Isaac Adanyeguh, Dr Sirio Cocozza, Louise Corben, Dr Andreas Deistung, Professor Martin Delatycki, Dr Imis Dogan, Professor William Gaetz, Professor Nellie Georgiou-Karistianis, Mr Simon Graf, Dr Marina Grisoli, Professor Pierre-Gilles Henry, Mr Gustavo Jarola, Dr James Joers, Dr Christian Langkammer, Professor Christophe Lenglet, Mr Jakun Li, Mr Camilia Lobo, Professor Eric Lock, Professor David Lynch, Dr Thomas Mareel, Professor Alberto Martinez, Dr Serena Monti, Dr Anna Nigri, Professor Massimo Pandolfo, Professor Kathrin Reetz, Professor Timothy Roberts, Dr Sandro Romanzetti, Dr David Rudko, Dr Alessandra Scaravilli, Professor Jörg Schulz, Dr S Subramony, Dr Dagmar Timmann, Professor Marcondes França, Dr Ian Harding, Mx. TRACK-FA Neuroimaging Consortium
Authorship	Yes
Abstract	Purpose: To develop a dentate nucleus (DN) segmentation tool using deep learning (DL) applied to brain quantitative susceptibility mapping (QSM) images. Materials and Methods: Brain QSM images from 132 healthy controls and 170 individuals with cerebellar ataxia or multiple sclerosis were collected from nine different datasets worldwide for this retrospective study. Manual delineation of the DN (gray matter and white matter hilus) was first undertaken by experienced raters with a robust quality control process. Performance of automated segmentation was compared following training using several DL architectures. A two-step approach was implemented, composed of a localization model followed by DN segmentation. Results: The manual tracing protocol produced ground-truth data with high intra-rater (average ICC 0.906) and inter-rater reliability (average ICC 0.776). Initial DL architecture exploration indicated that the nnU-Net framework performed best. The two-step localization plus segmentation pipeline achieved a Dice score of 0.898±0.031 and 0.894±0.036 for left and right DN, respectively. In external validation, our algorithm outperformed the leading existing automated tool (left/right DN Dice 0.863±0.038/0.843±0.066 vs. 0.568±0.222/0.582±0.239). The model demonstrated generalizability across unseen datasets during the training step. The measures showed a superior correlation index with manual annotations and performed well in both isotropic and anisotropic QSM datasets. Conclusion: We provide a model that accurately and efficiently segments the DN from brain QSM images. The model can be readily deployed for use in observational, natural history, and treatment trials for biomarker discovery.
Subject Terms	Health sciences/Neurology Health sciences/Health care/Medical imaging/Magnetic resonance imaging Health sciences/Neurology/Neurological disorders/Movement disorders Health sciences/Neurology/Neurological disorders/Neurodegenerative diseases Biological sciences/Computational biology and bioinformatics/Machine learning
Research Square author dashboard	I understand that my manuscript and associated personal data will be shared with Research Square for the delivery of the author dashboard.
Preprint Deposition	No, my co-authors and I would not like to benefit from <i>In Review</i>
Competing interests policy	There is a conflict of interest TJRR is employed by Biogen, receives a salary, and is a grant recipient from Friedreich's Ataxia Research Alliance (FARA). DHS is an Itaú Unibanco SA employee and a grant recipient from Friedreich's Ataxia Research Alliance (FARA). SC received fees from Amicus for the advisory board. LAC is funded by Friedreich's Ataxia Research Alliance (FARA) and is a consultant for Biogen Pharmaceuticals. ID is funded by Friedreich's Ataxia Research Alliance (FARA). NGK is funded by Friedreich's Ataxia Research Alliance (FARA) and CHDI Foundation Inc (New York, USA). PGH is a grant recipient from Friedreich's Ataxia Research Alliance (FARA) and the National Institute of Health (NIH) P41EB027061, P30NS076408, S10OD017974. JMJ is funded by Friedreich's Ataxia Research Alliance (FARA) and National Institute of Health (NIH). CLK reports grant FWF P35887. CL received research grants from Minoryx Therapeutics, research support from Biogen Inc and is funded by Friedreich's Ataxia Research Alliance (FARA) and NIH P41 EB027061. JL is funded by Friedreich's Ataxia Research Alliance (FARA). TM reports funding from National Institute of Health (NIH) U01 NS104326. SM is funded by the Italian MUR for the project "SEE LIFE - StrEnghEning the Italian InFrastructure of Euro- biomedicine." AN is funded by the Italian Ministry of Health (RRC). KR is funded by Friedreich's Ataxia Research Alliance (FARA). SHS is industry support research for Reata, Biogen, Biohaven, Avidity Biosciences, Fulcrum therapeutics, Vertex, Arthex, PTC, Reneo, Larimar, has a consulting work for Reata, Biogen, Fulcrum, participated in speaking engagements for Biogen, MDA, Bionews, Medscape, and is funded by NIH, FDA, Muscular Dystrophy Association, Wyck Foundation, Friedreich's Ataxia Research Alliance, and National Ataxia Foundation. DT reports grants DFG, DE 2516/1-1 and TI 239/17-1. MCFJ reports funding from FAPESP (São Paulo Research Foundation). IHH reports Friedreich's Ataxia Research Alliance general research grant and NHMRC Ideas and Investigator Grants (2026191, 1184403). All other authors declare no financial or non-financial competing interests.
Clinical Trial	No
Applicable Funding Source	Friedreich's Ataxia Research Alliance (FARA General Research Grant) [Rezende] Friedreich's Ataxia Research Alliance (FARA General Research Grant) [Shiraiishi] Australian National Health and Medical Research Council (NHMRC Ideas Grant 1184403) [Saha] Australian National Health and Medical Research Council (NHMRC Ideas Grant 1184403) [Harding]
Previous Interactions	None of the above

Stage	Start Date
Manuscript Received	17th Jun 24 21:50:36