Automated deep learning segmentation of high-resolution 7 Tesla postmortem MRI for quantitative analysis of structure-pathology correlations in neurodegenerative diseases

Pulkit Khandelwal[a,b], Michael Tran Duong[a], Shokufeh Sadaghiani[c], Sydney Lim[b,d], Amanda E. Denning[b,d], Eunice Chung[b,d], Sadhana Ravikumar[b,d], Sanaz Arezoumandan[c], Claire Peterson[c], Madigan Bedard[b,d], Noah Capp[c], Ranjit Ittyerah[b,d], Elyse Migdal[c], Grace Choi[c], Emily Kopp[c], Bridget Loja[c], Eusha Hasan[c], Jiacheng Li[c], Alejandra Bahena[c], Karthik Prabhakaran[d], Gabor Mizsei[d], Marianna Gabrielyan[c], Theresa Schuck[e], Winifred Trotman[c], John Robinson[e], Daniel T. Ohm[c], Edward B. Lee[e], John Q. Trojanowski[e,*], Corey McMillan[c], Murray Grossman[c,*], David J. Irwin[c], John A. Detre[c], M. Dylan Tisdall[d], Sandhitsu R. Das[b,c], Laura E. M. Wisse[f], David A. Wolk[c], Paul A. Yushkevich[b,d]

[a]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States
[b]Penn Image Computing and Science Laboratory, University of Pennsylvania, Philadelphia, PA, United States
[c]Department of Neurology, University of Pennsylvania, Philadelphia, PA, United States
[d]Department of Radiology, University of Pennsylvania, Philadelphia, PA, United States
[e]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, United States
[f]Department of Diagnostic Radiology, Lund University, Lund, Sweden
*Author deceased

Corresponding Author: Pulkit Khandelwal (pulks@seas.upenn.edu)

ABSTRACT

Postmortem MRI allows brain anatomy to be examined at high resolution and to link pathology measures with morphometric measurements. However, automated segmentation methods for brain mapping in postmortem MRI are not well developed, primarily due to limited availability of labeled datasets, and heterogeneity in scanner hardware and acquisition protocols. In this work, we present a high-resolution dataset of 135 postmortem human brain tissue specimens imaged at 0.3 mm³ isotropic using a T2w sequence on a 7T whole-body MRI scanner. We developed a deep learning pipeline to segment the cortical mantle by benchmarking the performance of nine deep neural architectures, followed by post-hoc topological correction. We evaluate the reliability of this pipeline via overlap metrics with manual segmentation in 6 specimens, and intra-class correlation between cortical thickness measures extracted from the automatic segmentation and expert-generated reference measures in 36 specimens. We also segment four subcortical structures (caudate, putamen, globus pallidus, and thalamus), white matter hyperintensities, and the normal appearing white matter, providing a limited evaluation of accuracy. We show generalizing capabilities across whole-brain hemispheres in different specimens, and also on unseen images acquired at 0.28 mm³ and 0.16 mm³ isotropic T2*w fast low angle shot (FLASH) sequence at 7T. We report associations between localized cortical thickness and volumetric measurements across key regions, and semi-quantitative neuropathological ratings in a subset of 82 individuals with Alzheimer's disease (AD) continuum diagnoses. Our code, Jupyter notebooks, and the containerized executables are publicly available at the **project webpage** (https://pulkit-khandelwal.github.io/exvivo-brain-upenn/).

Keywords: 7 T postmortem MRI, Alzheimer's disease, dementia, deep learning, image segmentation

1. INTRODUCTION

Neurodegenerative diseases are increasingly understood to be heterogeneous, with multiple distinct neuropathological processes jointly contributing to neurodegeneration in most patients, called *mixed pathology* (Schneider et al., 2007). For example, many patients diagnosed at autopsy with Alzheimer's disease (AD) also have brain lesions associated with vascular disease, TDP-43 proteinopathy, and α-synuclein pathology (Matej et al., 2019; Robinson et al., 2018). Currently, some of these pathological processes (particularly TDP-43 and α-synuclein pathologies) cannot be reliably detected with antemortem biomarkers, which makes it difficult for clinicians to determine to what extent cognitive decline in individual patients is driven by AD versus other factors. The recent modest successes of AD treatments in clinical trials (van Dyck et al., 2022) make it ever more important to derive antemortem biomarkers that can detect and quantify mixed pathology, so that treatments can be prioritized for those most likely to benefit from them.

Importantly, the understanding and utility of antemortem biomarkers are augmented by the coupling of imaging and postmortem pathology. Following autopsy, histological examination of the donor's brain tissue provides a semi-quantitative assessment of the presence and severity of various pathological drivers of neurodegeneration. Associations between these pathology measures and regional measures of neurodegeneration, such as cortical thickness, can identify patterns of neurodegeneration probabilistically linked to specific pathological drivers (de Flores et al., 2020; Frigerio et al., 2021; Wisse et al., 2020, 2021). Such association studies can use either antemortem or postmortem imaging (Wisse et al., 2017). Both approaches have their limitations. When the time between antemortem imaging and death is substantial, the postmortem pathology may not accurately match the state of pathology at the time of imaging. Postmortem MRI generally requires dedicated imaging facilities and image analysis algorithms, and whatever neurodegeneration/pathology association patterns are discovered have to be "translated" into the antemortem imaging domain for use as antemortem biomarkers. However, postmortem MRI allows imaging at much greater resolution than antemortem, allowing structure/pathology associations to be examined with greater granularity than antemortem imaging.

Furthermore, postmortem MRI of the brain can provide an advantage over antemortem MRI for visualizing detailed and intricate neuroanatomy and linking macroscopic morphometric measures such as cortical thickness to underlying cytoarchitecture and pathology (Adler et al., 2014; Alkemade et al., 2022; Augustinack et al., 2010; Beaujoin et al., 2018; García-Cabezas et al., 2020; Iglesias et al., 2015; Mancini et al., 2020; Pallebage-Gamarallage et al., 2018; Vega et al., 2021). Recent inquiries comparing postmortem imaging with histopathology have demonstrated relationships between atrophy measures and neurodegenerative pathology (Makkinejad et al., 2019; Ravikumar et al., 2021; Wisse et al., 2021; Yushkevich et al., 2021). Such associations corroborate patterns of neurodegeneration by specifically linking them with the underlying contributing pathology such as TAR DNA-binding protein 43 (TDP-43), phosphorylated tau (p-tau), and α-synuclein in Alzheimer's Disease (AD). In particular, Ravikumar et al. (2021) found significant correlations between tau pathology and thickness in the entorhinal cortex (ERC) and stratum radiatum lacunosum moleculare (SRLM) consistent with early Braak stages. Separately, Wisse et al. (2021) found significant associations of TDP-43 with thickness in the hippocampal subregions. Postmortem imaging also helps in characterizing underlying anatomy at the scale of subcortical layers (Augustinack et al., 2013; Kenkhuis et al., 2019), such as hippocampal subfields in the medial temporal lobe (MTL) (Ravikumar et al., 2020; Yushkevich et al., 2021). Several studies have also explored pathology/MRI associations in other neurodegenerative diseases, such as frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS) (Gordon et al., 2016; Irwin et al., 2015; Mackenzie et al., 2011). Previous work has identified correlations between high-resolution postmortem MRI and histopathology, to map myelin and iron deposits in cortical laminae (Bulk et al., 2020), due to oligodendrocytes and pathologic iron inclusions in astrocytes and microglia (Tisdall et al., 2021), which are a major source of iron because of myelination demands. Therefore, postmortem MRI would be helpful for validating and refining pathophysiological correlates derived from antemortem studies. Additionally, the volume of the WMH burden is an indirect marker of cerebrovascular pathology and the associations between cortical thickness and WMH (Dadar et al., 2022; Du et al., 2005; Rizvi et al., 2018) are complementary to the associations between cortical thickness and tau, TDP-43, amyloid-β and α-synuclein pathology. Also, van der Velpen et al. (2023) suggest that the subcortical brain structures are highly involved in dementia risk. Smaller volumes and thickness measurements of thalamus, amygdala, and hippocampus were associated with incident dementia.

Compared to antemortem MRI, postmortem MRI is not affected by head or respiratory motion artifacts and has much less stringent time. Compared to histology, postmortem MRI is less affected by distortion or tearing, and provides a continuous 3D representation of brain anatomy. However, both histology and postmortem MRI

are affected by changes occurring in the agonal state, during brain removal, and during tissue fixation and handling. Indeed, postmortem MRI is often used to provide a 3D reference space onto which 2D histological images are mapped. Combined analysis of postmortem MRI and histology makes it possible to link morphological changes in the brain to underlying pathology as well as to generate anatomically correct parcellations of the brain based on cytoarchitecture (Amunts et al., 2020; Schiffer, Spitzer, et al., 2021), and pathoarchitecture (Augustinack et al., 2013). Postmortem MRI could also act as a reference space to generate quantitative 3D maps of neurodegenerative proteinopathies from serial histology imaging (Ushizima et al., 2022; Yushkevich et al., 2021).

Given the rising use of high-resolution postmortem MRI in neurodegenerative disease research, automated techniques are imperative to effectively analyze such growing datasets. Particularly, in the case of structure-pathology association studies, scaling them beyond a few dozen datasets requires reliable morphometry measurements from postmortem MRI via accurate 3D segmentation and reconstruction of the structures of interest. There has been substantial work in brain MRI parcellation such as *FreeSurfer* (Fischl, 2012) and recent efforts based on deep learning (Chen et al., 2018; Henschel et al., 2020). However, these approaches focus on antemortem MRI, and there is limited work on developing automated segmentation methods for postmortem MRI segmentation. Postmortem segmentation methods have been region specific. Recent developments include automated deep learning methods for high-resolution cytoarchitectonic mapping of the occipital lobe in 2D histological sections (Eckermann et al., 2021; Kiwitz et al., 2020; Schiffer, Amunts, et al., 2021; Schiffer, Harmeling, et al., 2021; Schiffer, Spitzer, et al., 2021; Spitzer et al., 2018). The work by Iglesias et al. (2015, 2018) has developed an atlas to segment the MTL and the thalamus using manual segmentations in postmortem images. Yet, a postmortem segmentation method applicable to a variety of brain regions has yet to be described. This is attributable to several factors. Some groups have developed robust whole-brain postmortem image analysis tools (Casamitjana et al., 2024; Edlow et al., 2019; Jonkman et al., 2019; Mancini et al., 2020; Zeng et al., 2023), though overall there is limited availability of postmortem specimens, scans, segmentation algorithms, and labeled reference standard segmentations. Compared to antemortem structural MRI, postmortem MRI currently exhibits substantial heterogeneity in scanning protocols, larger image dimensions, greater textual complexity, and more profound artifacts. These issues can be addressed with new datasets and automated segmentation tools open to the public.

In this study, we expand upon our pilot study (Khandelwal, Duong, et al., 2022; Khandelwal, Sadaghiani, et al., 2022; Khandelwal et al., 2021) and develop a methodological framework to segment cortical gray matter; subcortical structures (caudate, putamen, globus pallidus, thalamus), white matter (WM), and WMH in high-resolution (0.3 x 0.3 x 0.3 mm^3) 7 T T2w postmortem MRI scans of whole-brain hemispheres. We train and evaluate our approach using 135 brain hemisphere scans from the Center for Neurodegenerative Disease Research of the University of Pennsylvania. We measure cortical thickness at several key locations in the cortex based on our automatic segmentation of the cortex, and correlate these measures with thickness measurements obtained using a user-guided semi-automated protocol. High consistency between these two sets of measures supports the use of deep learning-based automated thickness measures for postmortem brain segmentation and morphometry. We then report regional patterns of association between cortical thickness at a set of anatomical locations and neuropathology ratings (regional measures of p-tau, neuronal loss; as well as global amyloid-β, Braak staging, and CERAD ratings) obtained from histology data and WMH burden. Additionally, we show that networks trained on T2-weighted spin echo images acquired at 7 T generalize to postmortem images obtained with T2*w gradient echo fast low angle shot (FLASH) 7T MRI acquired at a resolution of 0.28 x 0.28 x 0.28 mm^3 and 0.16 x 0.16 x 0.16 mm^3.

2. MATERIALS

2.1. Donor cohort

We analyze a dataset of 135 postmortem whole-hemisphere MRI scans selected from Penn Integrated Neurodegenerative Disease Database (INDD) (Toledo et al., 2014). Patients were evaluated at the Penn Frontotemporal Degeneration Center (FTDC) or Alzheimer's Disease Research Center (ADRC) and followed to autopsy at the Penn Center for Neurodegenerative Disease Research (CNDR) as part of ongoing and previous clinical research programs (Arezoumandan et al., 2022; Irwin, 2016; Tisdall et al., 2021). The cohort included 62 female (sex assigned at birth) donors (Age: 75.37 ± 10.02 years, Age range: 53-97) and 73 male donors (Age: 73.95 ± 11.59 years, Age range: 44–101) with Alzheimer's Disease or related dementias (ADRD), such as Lewy body disease (LBD), FTLD-TDP43, progressive supranuclear palsy (PSP), or cognitively normal adults. Human brain specimens were obtained in accordance with local laws and regulations, and included informed consent from next of kin at time of death. The

Post-mortem right hemisphere brain tissue

(A) Lateral view

(B) Medial view
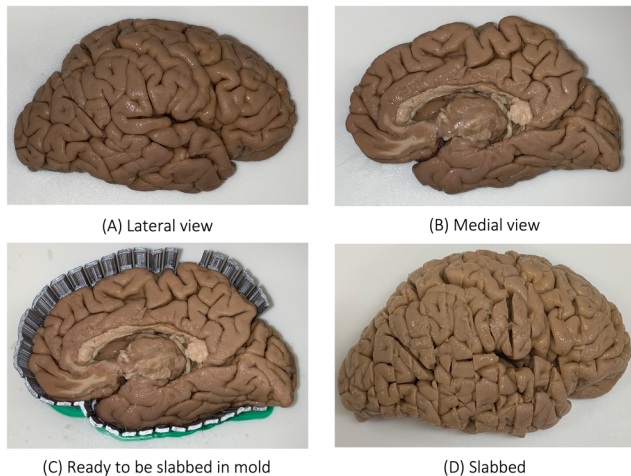
(C) Ready to be slabbed in mold

(D) Slabbed

Fig. 1. Postmortem tissue blockface photograph of a donor with diagnosis of Parkinson's disease (not demented) and Lewy body disease (deceased at the age of 79). Shown are the lateral (A) and medial (B) views of the right hemisphere. The tissue is then placed in a mold (C) and is subsequently slabbed (cut into different slices) as shown (D). See Lasserve et al., 2020 for more details. See Supplementary Figure 1 for the slices of the given brain tissue.

Table 1. Demographics of the Alzheimer's disease and related dementias (ADRD) patient cohort in the current study.

Brain donor cohort	
N	135 (Female: 62 and Male: 73)
Age (years)	74.60 ± 10.88 (range 44-101)
Race	White: 125 Black: 8 More than one race: 1 Unknown: 1
Hemisphere imaged	Right: 66 Left: 69
Postmortem interval (hours)	20.22 ± 15.05 (range: 3-105)
Fixation time (days)	298.67 ± 338.60 (range: 29-1559)

Neuropathological diagnosis	Primary	Secondary/tertiary
Alzheimer's disease	53	40
Lewy body disease	25	26
FTLD-TDP	17	-
Pick's disease	9	-
Progressive supranuclear palsy	6	1
Cerebrovascular disease	5	5
Corticobasal degeneration	4	1
Globular glial tauopathy	4	-
PART	2	10
LATE	2	15
Amyotrophic lateral sclerosis	1	1
FTLD-FUS	1	-
FTD with Parkinsonism (chromosome 17)	1	-
Argyrophilic grain disease	-	4
Hippocampal Sclerosis	-	5

Global neuropathological staging				
Ratings	0	1	2	3
Amyloid-β Thal staging (A score)	26	21	32	56
Braak three stage scheme (B score)	14	34	27	48
CERAD (C score)	49	20	16	50

AD, Alzheimer's disease; ALS, Amyotrophic lateral sclerosis; CVD, Cerebrovascular disease; LATE, Limbic-predominant age-related TDP-43 encephalopathy; LBD, Lewy body disease; CTE, chronic traumatic encephalopathy; FTLD-TDP, Frontotemporal lobar degeneration with TDP inclusions; GGT, Globular glial tauopathy; CVD, Cerebrovascular disease; CBD, Corticobasal degeneration; PART, Primary age-related tauopathy; PSP, Progressive supranuclear palsy; FUS, Fused-in-Sarcoma; tau-Misc, tauopathy unclassifiable.

patients were evaluated at FTDC and ADRC as per standard diagnostic criteria (Toledo et al., 2014), and imaged by the teams at the ADRC and the Penn Image Computing and Science Laboratory (PICSL) and the FTDC. Autopsy was performed at the CNDR. Figure 1 shows an example of a brain specimen with Parkinson's and LBD ready for autopsy. The brain is subsequently cut into 1 cm-thick slabs using a 3D-printed cutting mold, and slabs are further cut and embedded in paraffin for other ongoing studies in the laboratory. Supplementary Figure 1 illustrates the internal structures of the brain. Separately, the postmortem tissue photograph of a specimen with progressive non-fluent aphasia (PFNA) and Globular glial tauopathy (GGT) disease is shown in Supplementary Figure 2. Table 1 details the primary neuropathological diagnostic groups in the cohort with complete details tabulated in the Supplementary Spreadsheet.

2.2. MRI acquisition

During the autopsy of a specimen, one hemisphere was immersed in 10% neutral buffered formalin for at least 4 weeks prior to imaging. After the fixation time, samples were placed in Fomblin (California Vacuum Technology; Freemont, CA), enclosed in custom-built plastic bag holders. Samples were left to rest for at least 2 days to allow the air bubbles to escape from the tissue. The samples were scanned using either a custom-built small solenoid coil or a custom-modified quadrature birdcage (Varian,

Palo Alto, CA, USA) coil (Edlow et al., 2019; Tisdall et al., 2021). The samples were then placed into a whole-body 7 T scanner (MAGNETOM Terra, Siemens Healthineers, Erlangen, Germany). T2-weighted images were acquired using a 3D-encoded T2 SPACE sequence with 0.3 x 0.3 x 0.3 mm^3 isotropic resolution, 3 s repetition time (TR), echo time (TE) 383 ms, turbo factor 188, echo train duration 951 ms, and bandwidth 348 Hz/px in approximately 2–3 hours per measurement. Image reconstruction was done using the vendor's on-scanner reconstruction
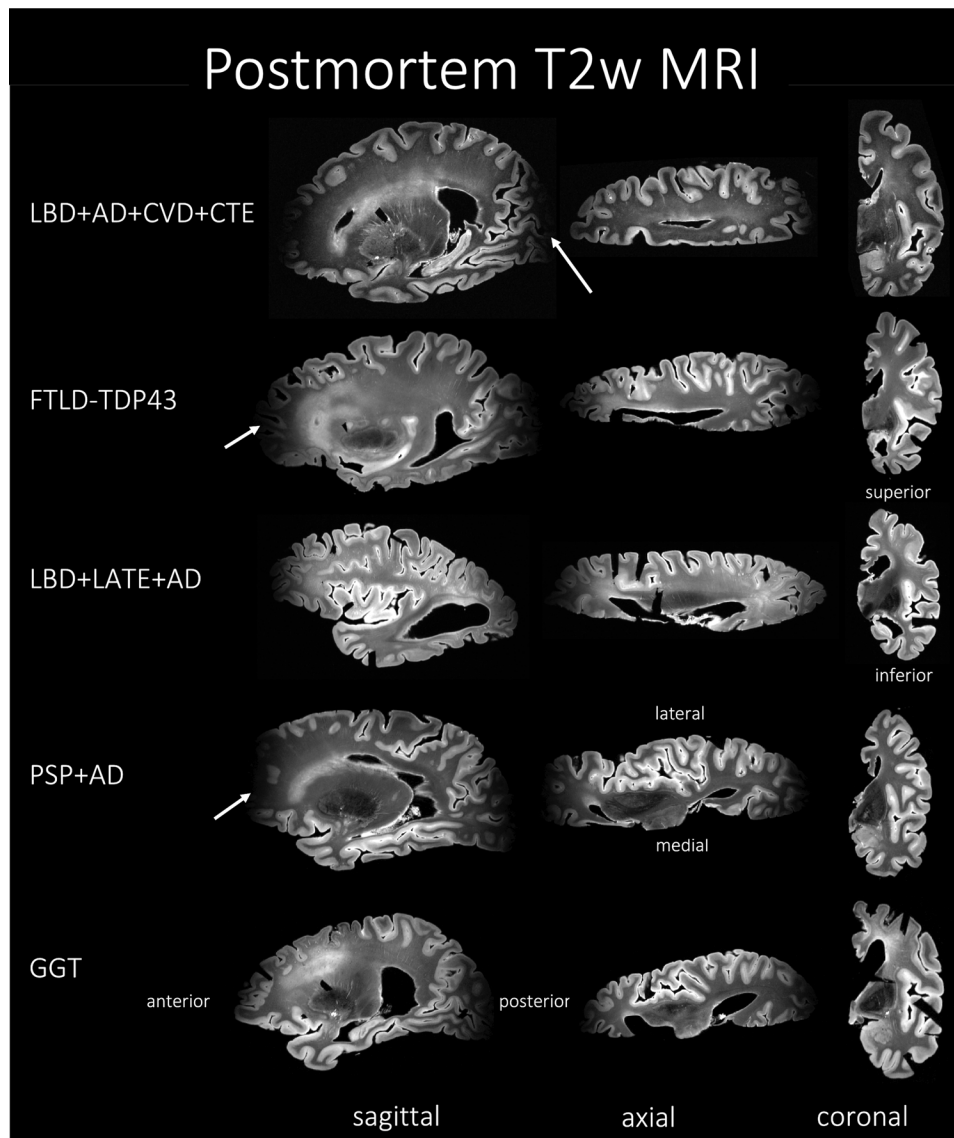
Fig. 2. MRI of the T2w sequence representative of Alzheimer's disease and related dementias (ADRD) spectrum with mixed pathology and diagnoses of five subjects. The heterogeneity among the subjects can be appreciated through the three different viewing planes. Notice that the MRI signal drops off at the anterior and posterior ends of the hemisphere, a drawback of the current acquisition protocol. AD, Alzheimer's disease; CVD, cerebrovascular disease; LATE, limbic-predominant age-related TDP-43 encephalopathy; LBD, Lewy body disease; CTE, chronic traumatic encephalopathy; FTLD-TDP, frontotemporal lobar degeneration with TDP inclusions; GGT, globular glial tauopathy; PART, primary age-related tauopathy; PSP, progressive supranuclear palsy.

software which also corrected the global frequency drift, combined the signal averages in k-space, and produced magnitude images for each echo. A total of four repeated measurements were acquired for each sample and subsequently averaged to generate the final image. Sample MRI slices are shown in Figure 2 for a range of specimens with different diseases. The image acquisition suffers from geometric distortions due to the non-linearity of the magnetic gradient field that increases towards the ends (both anterior and posterior) of the sample and B1-transmit inhomogeneity, which results in decreased image quality as shown in Figure 2.

We also acquire images at a much higher resolution using two separate T2*w FLASH MRI sequences, one at 280 microns and the other at 160 microns. We use these sequences for the generalization experiments as described in Section 4.3. Here, we briefly describe their acquisition parameters. For the 280 microns isotropic resolution: MRI data were reacquired with a 3D-encoded, 8-echo gradient-recalled echo (GRE) sequence with non-selective RF pulses. To maintain readout polarity and minimize distortions due to field inhomogeneity, each readout was followed by a flyback rephrasing gradient. The final echo was followed by an additional

completely rephrased readout to measure frequency drifts. Each line of k-space was acquired with multiple averages sequentially before advancing to the next phase-encode step. Common parameters for the sequence were: 280 microns isotropic resolution, 25° flip angle, 60 ms repetition time (TR), minimum echo time (TE) 3.48 ms, echo spacing 6.62 ms, and bandwidth 400 Hz/px. The field of view was adapted to each sample, and subsequently TRs and TEs were slightly modified based on the necessary readout duration. Total scan times were 8–10 hours for each sample.

For the 160 microns isotropic resolution: MRI data were acquired with a 3D-encoded, 3-echo GRE sequence with non-selective RF pulses with the sequence: 25° flip angle, 60 ms repetition time (TR), minimum echo time (TE), 9.37 ms, echo spacing 11.33 ms, and bandwidth 90 Hz/px similar to methods described in Tisdall et al. (2022).

2.3. Neuropathological ratings

The non-imaged hemisphere, that is, the contralateral tissue of each specimen systematically underwent histological processing for neuropathological examination at the CNDR (Toledo et al., 2014). Roughly $1.5 \times 1.5 \times 0.5$ cm^3 tissue blocks were extracted from the contralateral hemisphere. Paraffin-embedded blocks were then sectioned into 6 µm for immunohistochemistry using phosphorylated tau (AT8; Fisher, Waltham, MA; Catalogue No. ENMN1020) to detect phosphorylated p-tau deposits and p409/410 (mAb, 1:500, a gift from Dr. Manuela Neumann and Dr. E. Kremmer) to detect phosphorylated TDP-43 deposits. Immunohistochemistry evaluation was performed on the hemisphere contralateral to the hemisphere that was scanned using previously validated antibodies and established methods: NAB228 (monoclonal antibody [mAb], 1:8000, generated in the CNDR), phosphorylated tau PHF-1 (mAb, 1:1000, a gift from Dr. Peter Davies), TAR5P-1D3 (mAb, 1:500, a gift from Dr. Manuela Neumann and Dr. E. Kremmer), and Syn303 (mAb, 1:16,000, generated in the CNDR) to detect amyloid β (A β) deposits, phosphorylated tau (p-tau) deposits, phosphorylated TDP-43 deposits, and the presence of pathological conformation of α-synuclein, respectively. In 16 cortical regions, semi-quantitative ratings of p-tau, TDP-43, β-amyloid, and α-synuclein pathology, as well as neuronal loss, were visually assigned by an expert neuropathologist (E.B.L. and J.Q.T.) on a scale of 0–3, that is, "0: None," "0.5: Rare," "1: Mild," "2: Moderate," or "3: Severe" (Toledo et al., 2014). These ratings are illustrated in Figure 3. Global ratings of neurodegenerative disease progression were also derived, including A, B, and C scores (Hyman et al., 2012). Supplementary Table 2
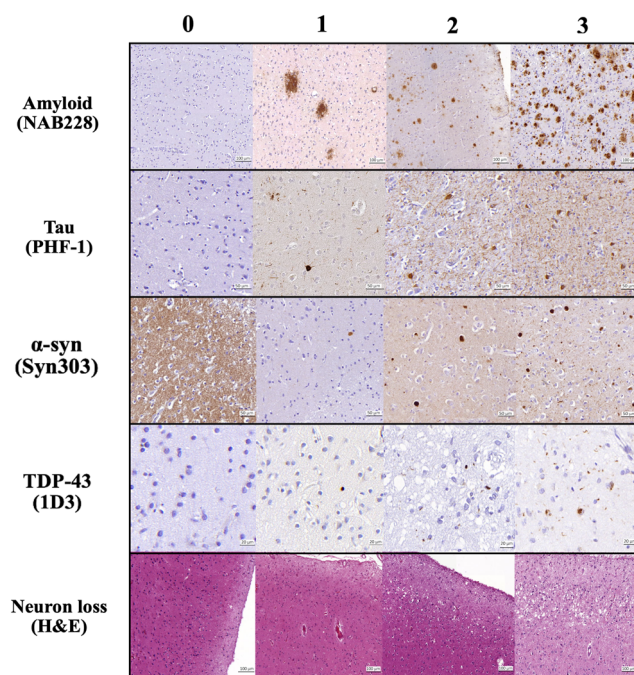

Fig. 3. Pathological assessment for each regional pathology derived from histology as per the protocol discussed in Section 2.3. The columns are severity ratings (left to right): 0-3. The shown pathologies are: amyloid- β plaques (NAB228 antibody), p-tau pathology (PHF-1 antibody), α-synuclein (Syn303 antibody), TDP-43 inclusions (pS409/410 antibody), and neuron loss (Hematoxylin and Eosin staining) in different cortical and medial temporal lobe regions included in the current work.

details the locations from where the neuropathology ratings were obtained from, either the exact (main regions) or the closest (exploratory regions), to the cortical brain regions. The justification of this matching of neuropathology regions to regions on MRI is provided in Sadaghiani et al. (2022).

2.4. Localized thickness measurement pipeline at key cortical locations

Our center has adopted an expert-supervised semi-automatic protocol to obtain localized quantitative measures of cortical thickness in all 7 T postmortem MRI scans, as described in the work by Sadaghiani et al. (2022) Section 2.3 and supplementary material S1 and Wisse et al. (2021) Section: MTL thickness measurements and supplementary material: Thickness measurement. In the current study, we use these measures as the reference standard for evaluating automated cortical segmentation. In each hemisphere, 16 cortical landmarks are identified and labeled on the MRI scan as shown in Figure 4A. To measure cortical thickness at these locations, a semi-automatic level set segmentation of the
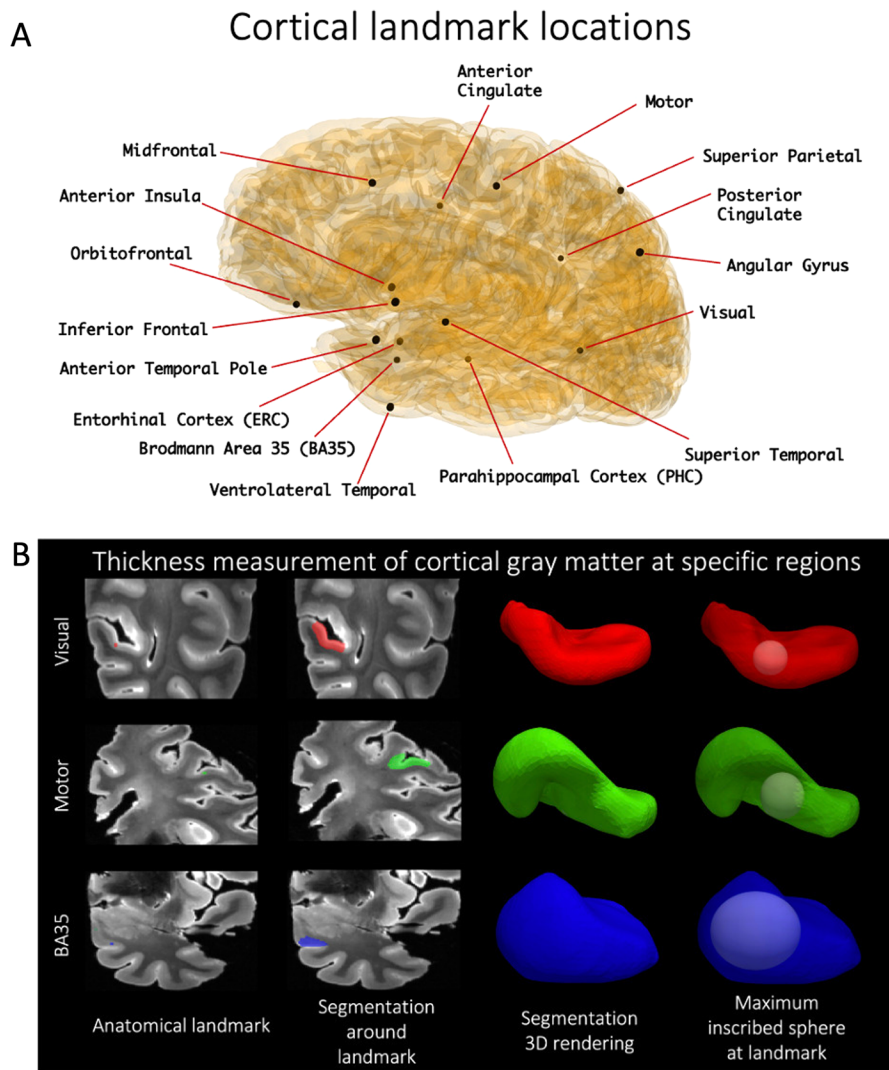
Fig. 4. Cortical thickness is measured at the 16 landmarks (A). A dot (shown here: motor, visual, BA35) is first placed to define an anatomical landmark, around which a semi-automatic level set segmentation of the surrounding cortical ribbon is provided. A maximally inscribed sphere is then computed using Voronoi skeletonization (Ogniewicz & Ilg, 1992), and the diameter of the sphere gives thickness at that landmark (B).

surrounding cortical ribbon is performed and the maximal sphere inscribed into the cortical segmentation and containing the landmark is found; the diameter of this sphere gives thickness at that landmark (Fig. 4B).

3. METHODS

In Section 3.1, we describe the manual segmentation protocols developed for cortical gray matter, the four subcortical structures, WMH, and normal-appearing white matter. Then, in Sections 3.2 and 3.3, we describe the deep learning-based pipeline and performance evaluation criteria developed for the automated segmentation of these structures. In Section 3.4, we describe the post-hoc topology correction step employed to provide geometrically accurate segmentations. Finally, in Section 3.5,

we describe the statistical models used to link neuropathological ratings of p-tau, amyloid-β neuronal loss, CERAD scores, Braak Staging with morphometry based on volume, and regional thickness measurements obtained from the automated segmentations.

3.1. Manual segmentation protocols

We developed protocols to manually segment structures in postmortem MRI: cortical gray matter, subcortical structures (caudate, putamen, globus pallidus, thalamus), WM, and WMH. We used these manual segmentations along with the corresponding MR images to train the neural networks. All manual segmentations were done by raters in ITK-SNAP (Yushkevich et al., 2019). Supplementary Table 1 shows which subjects were
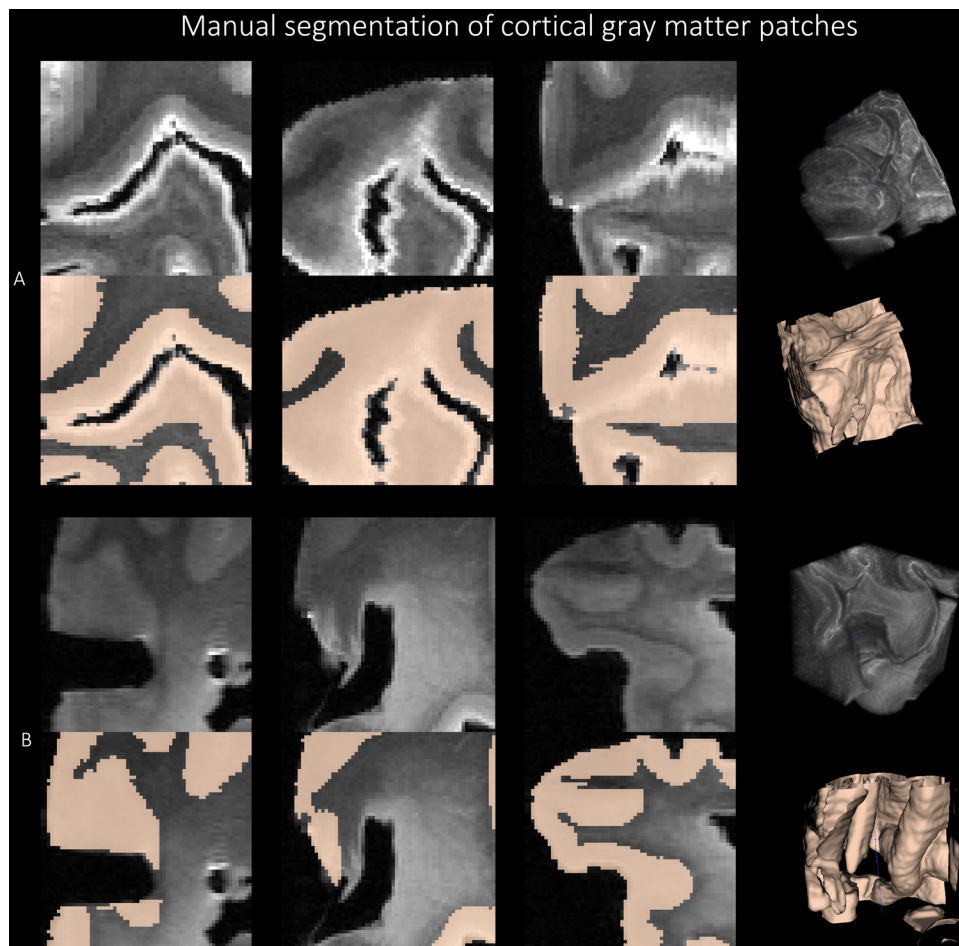
Fig. 5. Manual segmentation of the cortical gray matter patches of size 64 x 64 x 64 using the protocol explained in Figure 5. Section 3.1.1 for two subjects with FTLD-TDP (A) and GGT (B) in three viewing planes with the corresponding 3D renderings.

manually segmented to obtain the training data and the reference standard for the different labels, and the subsequent cross-validation studies as described in Sections 3.2 and 3.3.

3.1.1. Cortical gray matter

We sampled five 3D image patches of size 64 x 64 x 64 voxel^3, as shown in Figure 5 around the orbitofrontal cortex (OFC), anterior temporal cortex (ATC), inferior prefrontal cortex (IPFC), primary motor cortex (PMC), and primary somatosensory cortex (PSC) from 6 brain hemispheres, resulting in a total of 30 patches. These regions were selected as representative regions with variable levels of pathology in ADRD cases and control regions (PMC, PSC) (Tisdall et al., 2021). For example, OFC, ATC, and IPFC have high pathology in PSP, whereas, the PSC, which is generally less affected in most neurodegenerative diseases, was sampled as a negative control. In each patch, gray matter was segmented manually in ITK-SNAP. Five manual raters, divided into groups of two

(E.K. and G.C.) and three (E.H., B.L. and E.M.), labeled gray matter as the foreground, and rest of the image as the background using a combination of manual tracing and the semi-automated segmentation tool. The manual segmentation of cortical gray matter was supervised by author P.K. Figure 5 shows sample patch images and the corresponding reference standard labels with 3D renderings. We followed some guiding principles for manual segmentation: (1) we note that the white layer enveloping the cortex is not an imaging artifact, but is the outermost layer of the cortex, and thus is labeled as the gray matter; (2) adjoining gyri in deep sulci region are correctly labeled as gray matter and demarcate the deep sulci as the background; (3) in several regions, gray matter have similar intensities with the nearby WMH which were resolved by visual inspection of texture in the surrounding region; and (4) the 64 x 64 x 64 patches provided little context when segmenting the GM, therefore, the corresponding whole hemisphere image was displayed on a separate ITK-SNAP window for the user to examine the structures surrounding the given image patch. Inter-rater reliability

scores were computed for these manual segmentations in terms of Dice Coefficient (DSC): Raters 1&2: 95.26 ± 1.37 %, Raters 1&3: 94.64 ± 1.64 %, Raters 2&3: 94.54 ± 1.20 %, and Raters 4&5: 92.04 ± 4.26 %. After the review of the segmentations by an expert, the patches segmented by raters 2 (G.C.) and 5 (E.M.) were selected for training the deep learning models.

3.1.2.   Subcortical structures

Four subcortical gray matter structures (caudate, putamen, globus pallidus, thalamus) were manually segmented on seven specimens, selected to span varying pathological diagnoses (including Alzheimer's spectrum, p-tauopathies, TDP-43 encephalopathies, LBD, cerebrovascular disease and mixed disease), the range of postmortem age and levels of cerebral atrophy, tissue quality and tearing, mineral deposition and calcification, blood vessel size, and vascular pathology on imaging and autopsy. Figure 6 shows an example manual segmentation of the four subcortical structures. First, structures

were segmented across the axial plane based on several anatomical considerations. (1) The borders of the caudate (head, body, and tail) were defined by the frontal horns of the lateral ventricles (cerebrospinal fluid, anterior border) and the internal capsule (WM, boundaries of the body). Periventricular WM disease adjacent to the head of the caudate was excluded from the caudate segmentation. (2) The borders of the putamen were determined by the globus pallidus (medial) and the claustrum (lateral). (3) The globus pallidus was annotated to include the pallidum interna and externa, bounded by WM surrounding the thalamus and subthalamic nucleus. Mineral deposition (such as calcium) was noted by areas of heterogeneous T2 hypointensities and predominantly localized to the lentiform nucleus; where needed, the contours of some segmentations were adjusted to include these regions of heterogeneous signal within the lentiform structures. (4) The thalamus was segmented as the medial gray matter bounded by the subthalamic nucleus, midbrain, mamillary bodies, corpus callosum, lateral ventricles, and caudate. After segmentation in the axial
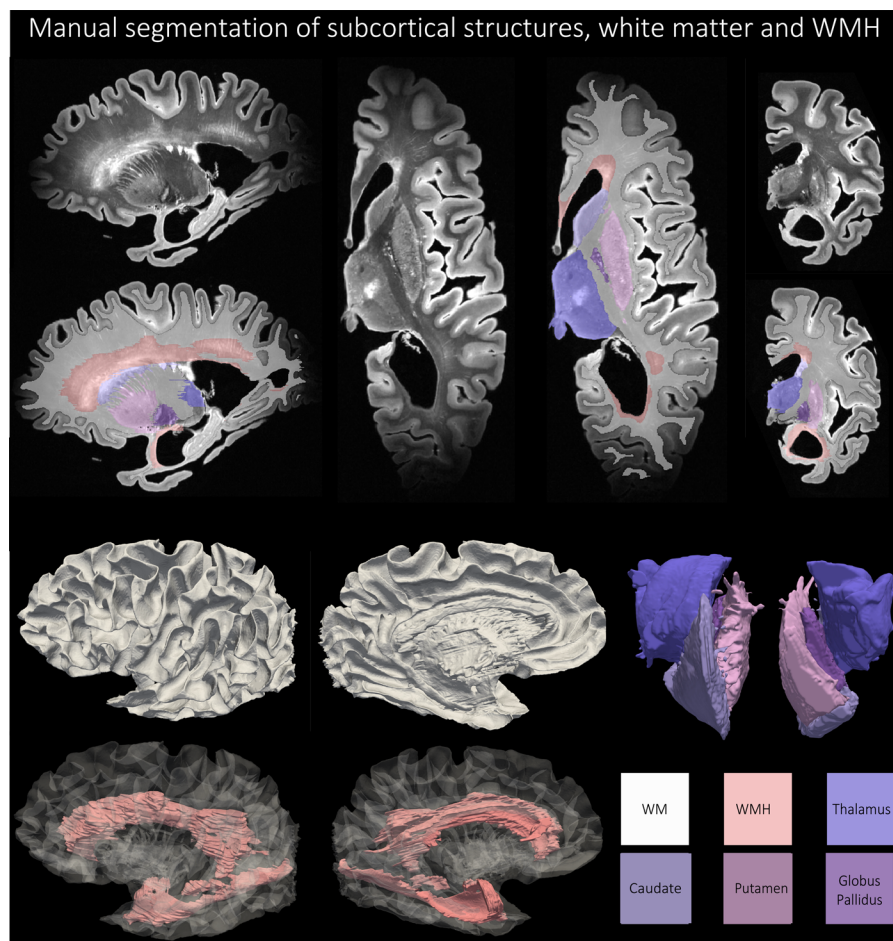

Fig. 6.   An example of manual segmentation of the thalamus, putamen, caudate, globus pallidus, and white matter hyperintensities as per the protocol described in Sections 3.1.2 and 3.1.3 for a subject with Alzheimer's Disease and Lewy body disease (87 years old).

plane, volumes were edited in the coronal and sagittal planes to ensure smoothness of the segmentation across all three planes. Boundaries between structures (such as striations between caudate and putamen) were agreed upon among authors. Manual segmentation of the four subcortical structures was performed by author, E.C. and supervised by and edited by M.T.D. Subcortical segmentations were discussed in consensus meetings with P.K., P.A.Y., S.R.D., and D.A.W.

3.1.3.　White matter hyperintensities

Nine specimens were chosen to segment WMHs across a gamut of vascular pathology with differing levels of WMH appearance, including focal small-vessel ischemia to intermediate periventricular and juxtacortical patterns to large, diffuse cerebrovascular disease. General principles for segmentation were applied as follows. (1) segmented lesions should be generally larger than 1 cm^3, (2) segmentations should appear for at least 4-6 slices to be a 1 cm^3 region, and (3) segmented WMH should include both periventricular (anterior frontal and posterior temporal/occipital horns of the lateral ventricles) and juxtacortical lesions. This distinguished WMH from insular cortex, claustrum, basal ganglia, and other gray matter structures embedded in WM. (4) WMH segmentations generally exhibited signal intensity above a threshold of ≥450-550 intensity (but this was influenced by field inhomogeneity artifacts, either between images or within the same image, often at the anterior and posterior cortical poles), given the image intensity range was normalized between 0-1000. (5) WMH segmentations included T2 hyperintense perivascular spaces and cortical venules but must also include surrounding T2 hyperintense white matter regions that occupy a region larger than 1 cm^3. (6) WMH was segmented primarily in axial plane and then assessed for contiguity and smoothness in sagittal and axial planes as well as 3D renderings. Manual segmentation was performed by the author A.E.D. and supervised and edited by M.T.D. WMH segmentations were discussed in consensus meetings with P.K., P.A.Y, S.R.D, and D.A.W. Figure 6 shows an example manual segmentation of the WMH.

3.1.4.　White matter

Two specimens with manually labeled subcortical structures and WMH, but nnU-Net-based automated segmentations for cortical mantle were thresholded to obtain the WM label, which was then manually corrected by the author A.E.D., supervised by P.K., to remove incorrectly labeled spurious voxels to fill-in holes and thereby define a clear GM/WM boundary.

3.2.　Automated deep learning-based segmentation for cortical gray matter

We first developed an approach for labeling cortical gray matter in postmortem 7 T MRI scans using the 30 patch-level cortical gray matter segmentations from six subjects (described in Section 3.1.1) for training and cross-validation. Given the exceptional performance of convolutional neural network (CNN) models for antemortem medical image segmentation (Schiffer, Spitzer, et al., 2021), our effort focused on benchmarking leading existing CNN models, rather than developing another model for the postmortem cortical segmentation task. We benchmarked the following variants of popular biomedical image segmentation deep learning models: (1) nnU-Net (Isensee et al., 2021); four variants of AnatomyNet (Zhu et al., 2019) based on squeeze-and-excitation blocks (Roy et al., 2018): (2) Channel Squeeze and Spatial Excitation AnatomyNet (SE); (3) Spatial Squeeze and Channel Excitation AnatomyNet (CE); (4) Anatomy-Net (Vanilla); (5) Spatial and Channel Squeeze and Excitation AnatomyNet (SE + CE); (6) 3D Unet-like network (Khandelwal & Yushkevich, 2020); (7) VoxResNet (Chen et al., 2018); (8) VNet (Milletari et al., 2016); and (9) Attention Unet (Oktay et al., 2018). Supplementary Material provides architectural details of the nine deep neural networks. We use PyTorch 1.5.1 and a single Nvidia Quadro RTX 5000 GPU to train the models using user-annotated images described in Sections 3.1.1-3.1.3. To train the deep neural networks, the input images were standardized, and then normalized between 0 and 1.

To ensure a fair and systematic evaluation of the nine networks, we trained and evaluated the nine network architectures within the nnU-Net framework under matched conditions (i.e., same split of data into training/validation/testing subsets; same data augmentation strategy, same hyper-parameter tuning strategy). We first evaluated the accuracy of cortical segmentation by six-fold cross-validation on the patch-level manual segmentations. We report average DSC and average Hausdorff Distance 95th percentile (HD95) across the 30 segmented patches in the six-fold cross-validation experiments.

To choose the best performing model for subsequent analyses, the nine trained models were then used to segment the whole hemisphere cortical mantle across a subset of the cohort (N = 36 for which the user-supervised semi-automated segmentations are available). We used these segmentations to compute the thickness around the 16 landmarks in these 36 subjects at the cortical landmarks. We then compared the thickness measurements to corresponding measures obtained via user-supervised semi-automated segmentations (reference standard) using our protocol (Section 2.4). In Supplementary Figure 4, for each of the 16 landmarks, we report Spearman's correlation

coefficient between the automated and the reference manual segmentation-based cortical thickness measurements, and the average fixed raters intra-class correlation coefficient (ICC) (Shrout & Fleiss, 1979). Based on the combination of cross-validation DSC accuracy, agreement between the automatic and manual segmentation-based cortical thickness measures, and overall visual inspection of the segmentations, a single model was selected for subsequent experiments. As reported in Section 4.1, this best-performing method was the **nnU-Net** model.

3.3. Automated deep learning-based segmentation for other structures

We trained another vanilla nnU-Net model to obtain the automated segmentation of subcortical structures, WMH, and normal-appearing WM. In particular, for the purpose of five-fold cross-validation studies to report DSC and HD95 scores, we trained an nnU-Net model based on the manually labeled training data for the segmentation of subcortical structures in seven subjects and WMH in nine subjects; and the manually labeled WM and the automated cortical mantle segmentations obtained from the nnU-Net model trained in the previous Section 3.2 as additional labels. Supplementary Table 1 lists the subjects used for training and evaluation of all the seven labels.

3.4. Post-hoc automated topology correction of segmentations

As shown in Figure 2, the image acquisition suffers from geometric distortions due to the non-linearity of the magnetic gradient field and might also suffer from partial volume averaging which generally introduces holes and handles in the segmentation and makes the WM surface to lack a spherical topology. Furthermore, the opposing banks of sulcus often appear to be fused, which makes it harder for the segmentation algorithm to correctly segment the cortex and thus might produce erroneous cortical thickness values. Hence, it is important to correct such mis-segmentations by enforcing a topology correction method. Therefore, we employ the methods developed in CRUISE: Cortical reconstruction using implicit surface evolution (Han et al., 2004) made available as part of the "*nighres*" package, with default settings (Huntenburg et al., 2018) for post-hoc topology correction of WM surface and constraining the GM segmentations. In particular, the method uses a fast marching-based method (Bazin & Pham, 2007), a graph-based topology correction algorithm (GTCA) (Han et al., 2002), and a topology preserving geometric deformable model (TGDM) (Han et al., 2003) to rectify holes and handles, thus producing a WM surface with a spherical topology. Separately, the Anatomically Consistent Enhancement (ACE) method is used to provide a GM representation

that has evidence of sulci where it might not otherwise exist due to the partial volume effect. Thus, we obtain geometrically accurate and topologically correct segmentations for the cortical GM and WM. In the current study, we apply this topology correction step on the automated GM/WM segmentation labels obtained from the nnU-Net model trained in the previous Section 3.3. We called the post-hoc topology corrected model as "nnU-Net-CRUISE."

3.5. Linking neuropathology ratings with morphometry

We computed Spearman's correlation between thickness measurements obtained from automated segmentations at the 16 anatomical landmarks described above with semi-quantitative pathology ratings from approximately corresponding anatomical locations in the contralateral hemisphere (regional p-tau score, regional neuronal loss) and global pathology ratings (CERAD stage, Braak stage, β-amyloid). We repeat this analysis with thickness measurements obtained from user-annotated manual reference segmentations, and thereby test the hypothesis that similar trends would be seen between pathology correlations with automated and manual thickness measures, which, in turn, would imply that automated segmentations are viable for morphometry-based studies (Section 4.4). In particular, we follow the experimental design from our recent work (Sadaghiani et al., 2022) for a subset of the cohort within the AD spectrum having AD as their primary diagnosis and also diagnosis of either: LBD, PART, LATE, CBD (N = 82) out of 135. The criteria for AD continuum are based on excluding cases with diagnoses of FLTD or non-AD tauopathy (whether primary or secondary). Finally, we normalized the WMH volumes by the corresponding WM volumes, and then computed one-sided Spearman correlation between the normalized WMH volume with regional cortical thickness and subcortical volumes for the subjects within the AD spectrum (N = 82). We include nuisance covariates of age, sex, and postmortem interval (PMI) in all of our analyses.

Table 2. Six-fold cross-validation dice coefficient (DSC) and Hausdorff distance 95th percentile (HD95) scores between reference standard and automated patch-level cortical segmentations.

Deep learning method	DSC (%)	HD95 (mm)
nnU-Net	93.98 ± 5.25	0.49 ± 0.45
AnatomyNet (SE)	94.84 ± 3.84	0.45 ± 0.42
AnatomyNet (CE)	94.91 ± 3.27	0.45 ± 0.42
AnatomyNet (Vanilla)	94.86 ± 3.83	0.46 ± 0.44
AnatomyNet (CE + SE)	94.66 ± 3.79	0.47 ± 0.44
3D Unet	93.57 ± 5.22	0.58 ± 0.51
VoxResNet	94.84 ± 4.00	0.45 ± 0.42
VNet	90.84 ± 5.93	0.99 ± 0.56
Attention Unet	93.65 ± 4.91	0.62 ± 0.66

4.   RESULTS

4.1.   Cortical gray matter segmentation

4.1.1.   *Dice coefficient volume overlap and qualitative analysis*

Table 2 tabulates the *patch*-level cortical gray matter segmentation performance of the nine different networks across six-fold cross-validation. AnatomyNet and its vari-

ants attain the highest patch-level DSC, closely followed by VoxResNet. The nnU-Net model has slightly lower DSC than the best AnatomyNet model, but the difference is less that 1%. However, since the patches used to train the segmentation networks were only sampled from select regions of the hemispheres, cross-validation accuracy on these patches is not necessarily indicative of the networks' ability to generalize to other brain regions. Figure 7 illustrates, that consistently across our specimens,
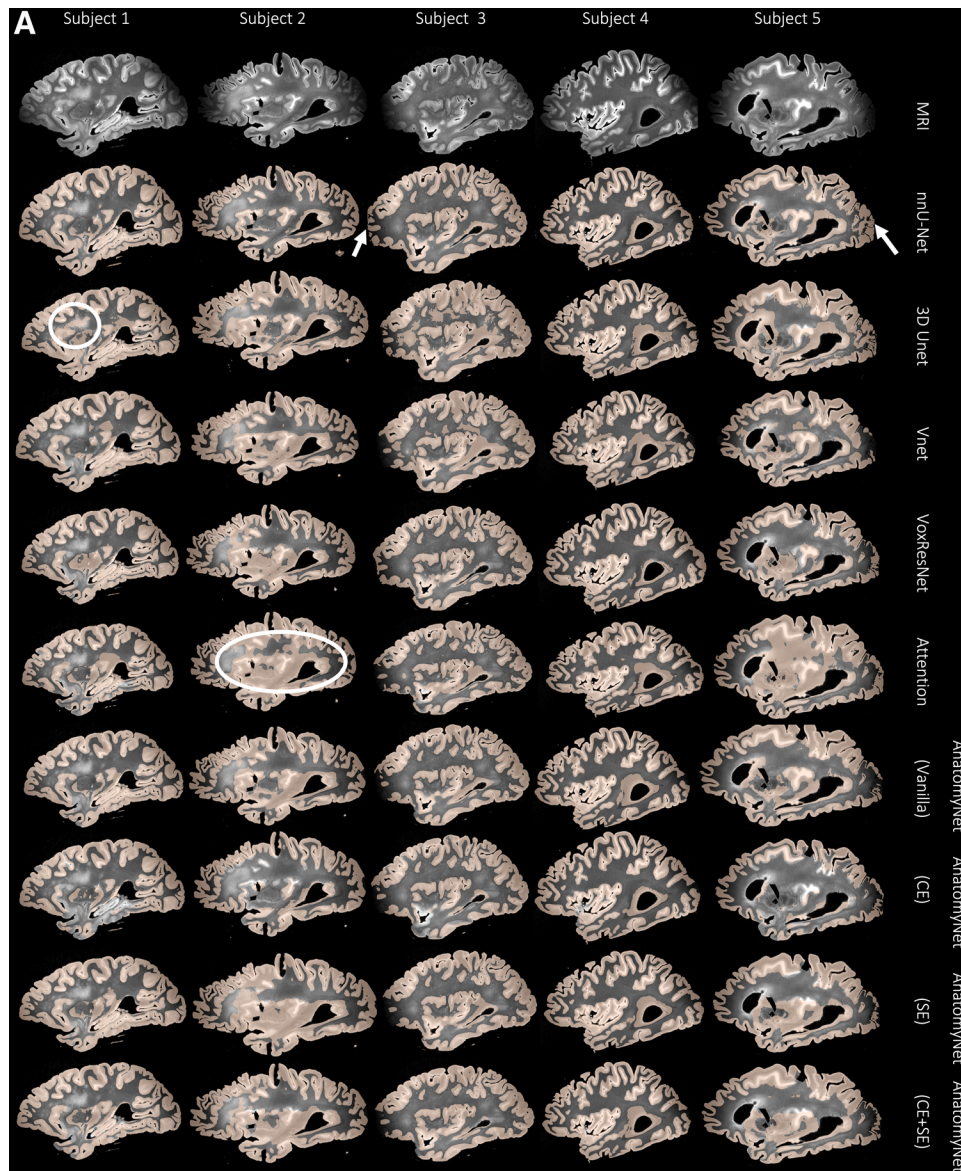

Fig. 7.   (A) For a subset of 10 subjects, each column shows an example sagittal view of the automated cortical mantle segmentation predicted across the whole-brain hemisphere using the different neural network architectures. We notice that all architectures except nnU-Net show either under- or over-segmentation of the cortical gray matter, which, together with the results reported in Table 3 (ICC results), prompted us to select nnU-Net as the preferred model for cortical gray matter segmentation. For example, notice how networks such as 3D Unet and Attention Unet incorrectly segment large chunks of WM as cortical GM (white circles). Whereas nnU-Net performs the best in difficult to segment areas such as the anterior and posterior regions of the brain with poor MRI signal. (B) Again, notice that AnatomyNet (CE) is not able to segment certain areas in the cortical mantle (white arrows). The 3D Unet and Vnet networks oversegment the cortical mantle into the WM. However, nnU-Net overcomes all of these limitations.

AnatomyNet and its variants are able to distinguish gray matter from white matter in high-contrast regions, but fail to segment the anterior and posterior regions where contrast is lower due to limitations of the MRI coil. There is also some systematic under-segmentation (see white arrows) of the cortex even in higher-contrast regions (see the white circled regions in Fig. 7). By contrast, nnU-Net clearly demarcates GM/WM boundary even in low-contrast regions, which is remarkable considering that these regions were not captured by the training patches. Supplementary Figure 3 shows the Dice scores and the HD95 scores as box plots with pairwise paired t-tests with Bonferroni correction for all the nine network archi-

tectures in 2. Mainly the comparisons between VNet, the lowest performing network, and the rest are significant. This supports our choice of nnU-Net given that it is not significantly worse performing than any of the other architectures.

4.1.2. Intra-class correlation coefficient

We compare cortical thickness (mm) at 16 cortical landmarks between automated gray matter segmentations obtained by the nine networks and the user-supervised semi-automated segmentation method, which serves as the reference standard.
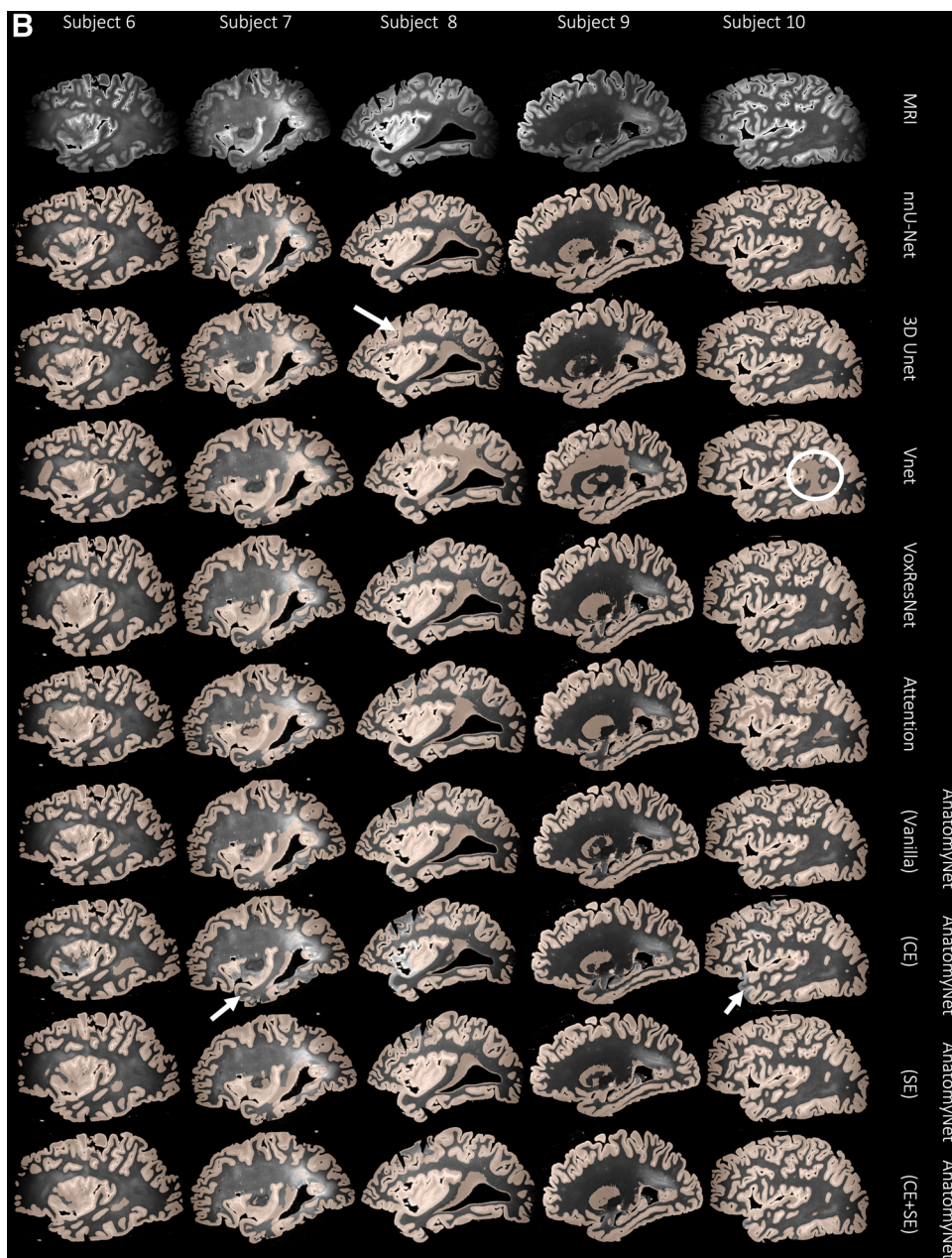

Fig. 7 (Continued)

Table 3. Average fixed raters intra-class correlation coefficient (ICC) scores for the regional cortical thickness measurements between automated nnU-Net and reference standard manual segmentations for all the nine neural network architectures, and the topologically corrected nnU-Net-CRUISE segmentations.

ROI\Network	VoxResNet	VNet	3D Unet	Attention UNet	AnatomyNet (SE)	AnatomyNet (CE+SE)	AnatomyNet (Vanilla)	AnatomyNet (CE)	nnU-Net	nnU-Net-CRUISE
Visual	0.40	0.27	0.39	0.40	0.32	0.41	0.34	0.53	0.54	0.55
Motor	0.59	0.03	0.25	0.03	0.11	0.04	0.08	0.46	0.72	0.75
Posterior cingulate	0.57	0.51	0.46	0.39	0.46	0.68	0.62	0.88	0.92	0.96
Midfrontal	0.63	0.55	0.32	0.54	0.64	0.73	0.72	0.61	0.84	0.79
Anterior cingulate	0.86	0.67	0.72	-0.12	0.34	0.56	0.51	0.70	0.94	0.88
Orbitofrontal	0.66	0.44	0.48	0.62	0.69	0.61	0.64	0.51	0.69	0.75
Superior temporal	0.52	0.27	0.44	0.34	0.34	0.35	0.38	0.30	0.54	0.79
Inferior frontal	0.13	-0.01	0.28	0.16	-0.17	-0.09	0.34	0.37	0.46	0.79
Anterior insula	-0.14	0.44	0.50	0.08	0.36	0.22	0.30	-0.16	0.67	0.77
Anterior temporal	0.53	0.28	0.52	0.58	0.24	0.12	0.20	-0.21	0.82	0.94
Ventrolaterl	0.27	0.50	0.40	0.39	0.08	0.64	0.31	0.60	0.87	0.71
Superior pareital	0.78	0.46	0.48	0.46	0.81	0.71	0.81	0.82	0.70	0.79
Angular gyrus	0.39	0.22	0.25	0.25	0.33	0.26	0.45	0.44	0.86	0.51
ERC	0.27	-0.13	0.79	0.57	0.12	0.47	0.11	0.64	0.62	0.75
BA35	0.53	0.56	0.79	0.56	0.61	0.52	0.54	0.56	0.73	0.59
PHC	0.15	0.41	0.36	0.42	0.21	0.23	0.11	0.49	0.59	0.73
Mean	**0.45**	**0.28**	**0.47**	**0.35**	**0.34**	**0.40**	**0.40**	**0.47**	**0.72**	**0.73**

Across each row, each cell is color coded, with darker shades indicating higher ICC value.

Table 3 tabulates the Average fixed raters ICC scores for all the nine networks trained as described in 3.2. We observe that nnU-Net (mean ICC = 0.72) is clearly the best among the nine patch-based models. The final nnU-Net-CRUISE model which has been topologically corrected as described in Sections 3.3 and 3.4 slightly exceeds the patch-based vanilla nnU-Net (mean ICC of 0.73) on average. The variants of AnatomyNet have mean ICC values of, Vanilla: 0.40, CE: 0.47, CE+SE: 0.40, and SE: 0.34. We observe that the variants of AnatomyNet were the top performing models when evaluated using DSC scores at patch level, but did not generalize to robustly segment the entire cortical mantle; and thereby failed to show good correlation of regional cortical thickness when compared with the reference standard reference cortical thickness. The other four models also had very low ICC values: VoxResNet: 0.45, VNet: 0.28, 3D Unet: 0.47, and Attention Unet: 0.35.

The Bland-Altman plots in Figure 8 show strong agreement between reference standard and post-hoc topologically corrected automated nnU-Net segmentation-based thickness measurements for the 16 cortical landmarks. Furthermore, Supplementary Figure 4 shows that 13 out of the 16 regions have correlation coefficient (r) greater than 0.6. We also observe high ICC scores, with 12 regions having ICC greater than 0.7. These results confirm that automated segmentations are accurate to give desirable cortical thickness measurements.

Therefore, based on quantitative evaluation in terms of DSC and HD95 scores, ICC values, and the qualitative visual inspection of the segmentations for the different neural network architectures, we conclude that nnU-Net-CRUISE is the best performing model. Figure 9 shows the post-hoc topology corrected automated nnU-Net segmentations of cortical gray matter and white matter shown in sagittal view for five randomly chosen subjects. The first columns shows the MRI slice, the automated nnU-Net segmentation before and after topology correction step are shown in columns 2 and 3. Columns 4 and 5 show the zoomed-in area, with the red arrows indicating the regions where topology correction improves the segmentation. We notice that the opposite banks of sulci are no longer fused and in fact well demarcated after correcting for topology. These segmentations will therefore provide more reliable and accurate estimates of cortical thickness.

4.2. Other structures: subcortical, WMH, and WM segmentation

Based on the superior performance of nnU-Net, as explained in the previous section, we employ nnU-Net to perform the multi-label segmentation for WMH, caudate, putamen, globus pallidus, thalamus, and WM.
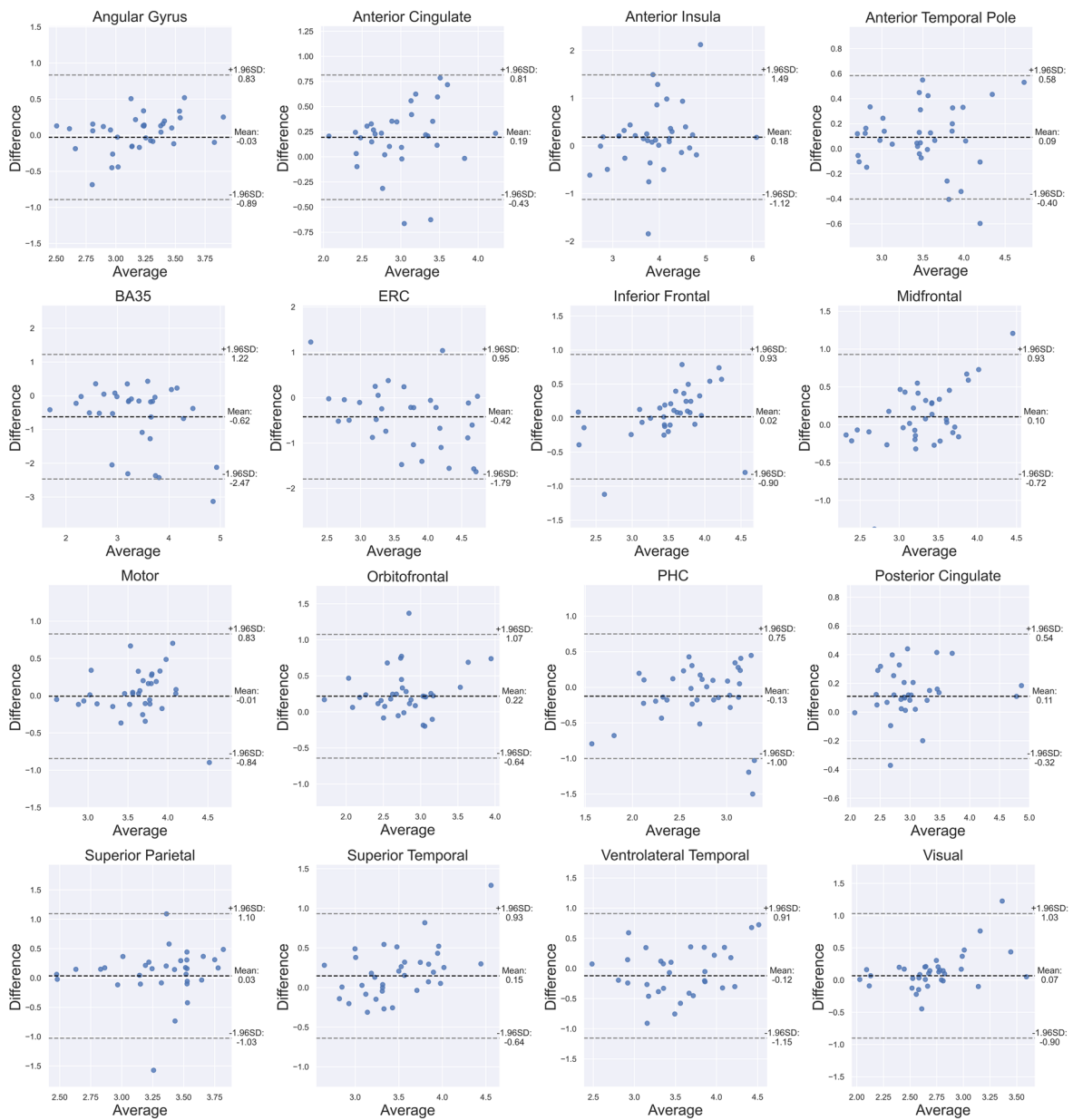
Fig. 8. Bland-Altman plots for the cortical regions. We observe good agreement between the manual and nnU-Net-CRUISE automated segmentations-based thickness (mm), which supports the hypothesis that topologically corrected deep learning-based automated segmentations are reliable for morphometric measurements. Note that the highlighted differences are in the direction: manual-automated segmentations. Supplementary Figure 3 shows the corresponding correlation plots.

The mean DSC scores for the four subcortical structures and WMH across all leave-one-out cross-validation experiments were: WMH: 79.70 %, caudate: 88.18 %, putamen: 85.20 %, globus pallidus: 80.12 %, and the thalamus: 87.29 %. Note that currently, we do not have a large sample size to perform cross-validation evaluation on the WM, due to the time-consuming process of manually correcting the segmentations. Obtaining manual WM label is beyond the scope of the current study. Figures 10 and 11 show qualitative results for all the

segmented structures in 2D and 3D respectively for the 36 subjects used for quantitative evaluation in the previous section, in sagittal view.

4.3. Generalization to other imaging sequences and protocols

Further highlighting the strong generalization properties of nnU-Net, Figure 12 illustrates that the nnU-Net model trained on 7 T 0.3 x 0.3 x 0.3 mm³ T2w images is able to
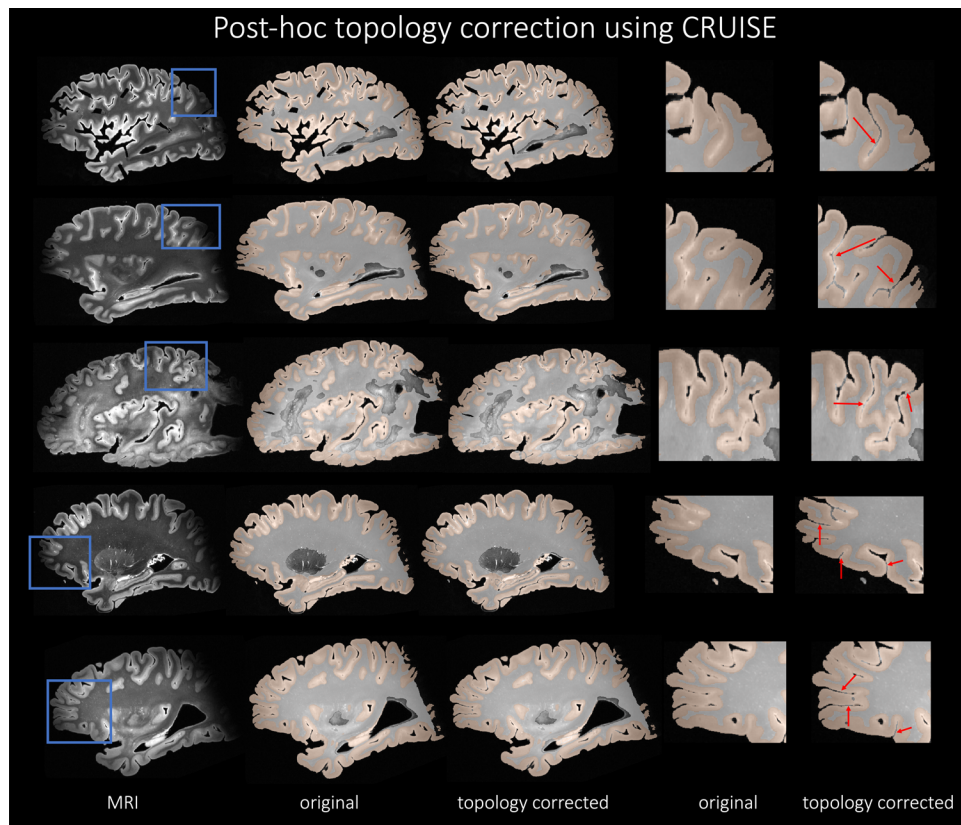
Fig. 9. nnU-Net-CRUISE. Topologically corrected automated nnU-Net segmentations of cortical gray matter and white matter shown in sagittal view for five subjects. The first columns shows the MRI slice, the automated nnU-Net segmentation before and after topology correction step (nnU-Net-CRUISE) are shown in columns 2 and 3, respectively. Columns 4 and 5 show the zoomed-in area, with the red arrows indicating the regions where topology correction improves the segmentation.

generalize well to MRI sequences and resolutions unseen during training. In particular, we evaluate the trained model on 7 T T2*w GRE FLASH sequence postmortem images acquired (Tisdall et al., 2021) at 0.28 x 0.28 x 0.28 mm³ (N = 13) and 0.16 x 0.16 x 0.16 mm³ (N = 73) resolution as shown qualitatively in Figure 12. We observe good generalization performance for GM, WM and WMH but observe some under-segmentations in the subcortical structures for both the 160 and 280 micron sequences. Currently, we only provide qualitative assessment on a representative sample of the FLASH images. The limitation lies in quantitative assessment (beyond the scope of the current study) in terms of Dice score and ICV which is currently not possible due to lack of reference manual segmentation.

4.4. Morphometry associations with underlying neuropathology

4.4.1. Regional cortical atrophy correlation patterns with underlying ratings

We correlate regional cortical thickness measures derived from nnU-Net-CRUISE gray matter segmentation with corresponding regional ratings of p-tau pathology and

neuronal loss density, amyloid-β ratings, CERAD, and Braak staging as explained in Section 3.5. Table 4 tabulates the correlation between the regional cortical thickness and the pathology ratings for measurements based on automated nnU-Net-CRUISE segmentations for the subjects within the AD continuum (N = 82). For the automated nnU-Net segmentation-based regional cortical thickness measurements, we observe significant negative one-sided Spearman's correlation which survived Bonferroni multiple corrections (corrected p-value <0.05), controlling for age, sex, and PMI:

- in posterior cingulate (r = -0.46, p = 0.0001), midfrontal (r = -0.441, p = 10^{-5}), superior temporal (r = -0.325, p = 0.002), and entorhinal cortex (r = -0.429, p = 0.0001) with amyloid-β;
- in posterior cingulate (r = -0.445, p = 0.0002), midfrontal (r = -0.327, p = 0.0018) and entorhinal cortex (r = -0.522, p = 10^{-5}), and parahippocampal cortex (r = -0.328, p = 0.0018) with Braak staging;
- in posterior cingulate (r = -0.419, p = 0.0004) and entorhinal cortex (r = -0.335, p = 0.0016) with CERAD rating;
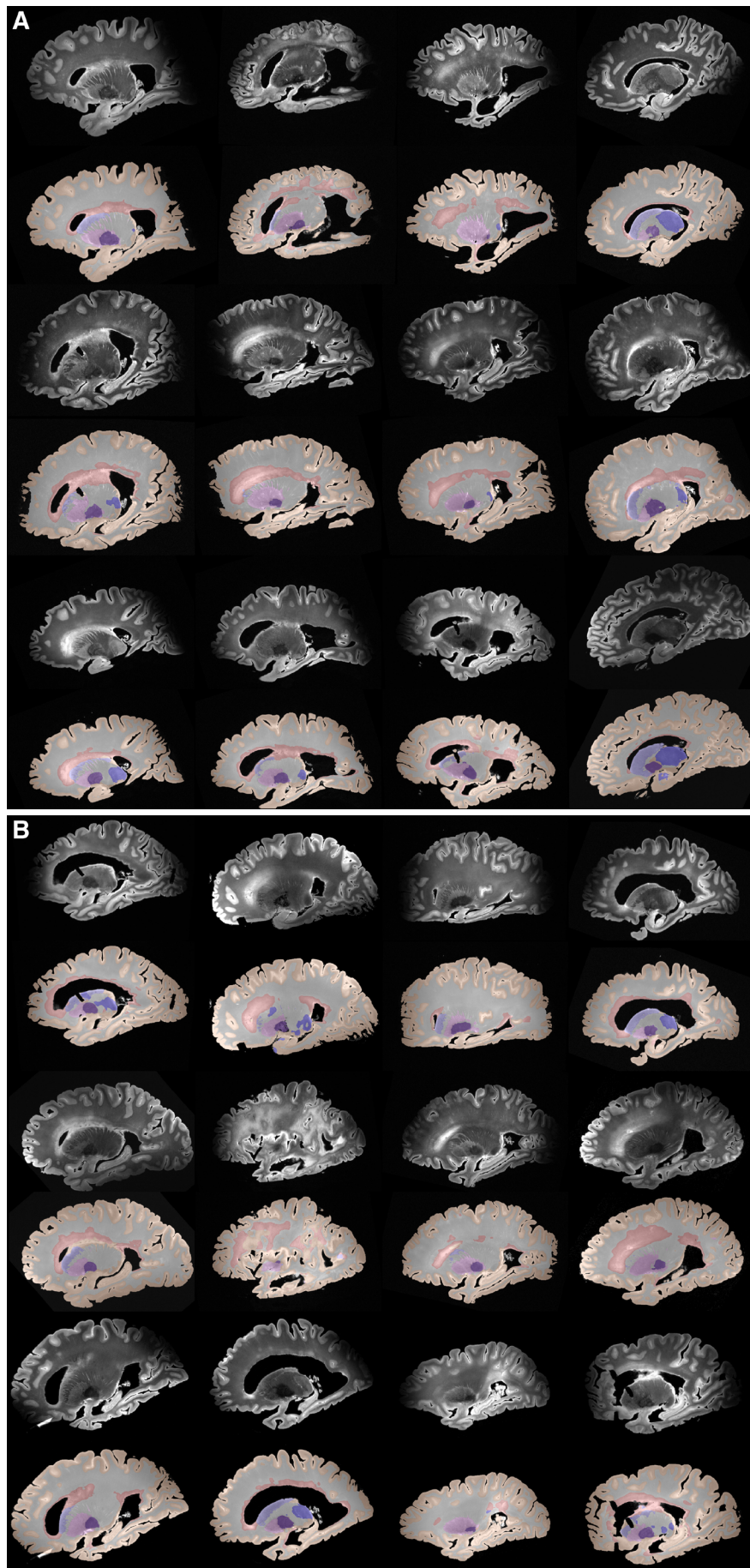
Fig. 10 (Continued)

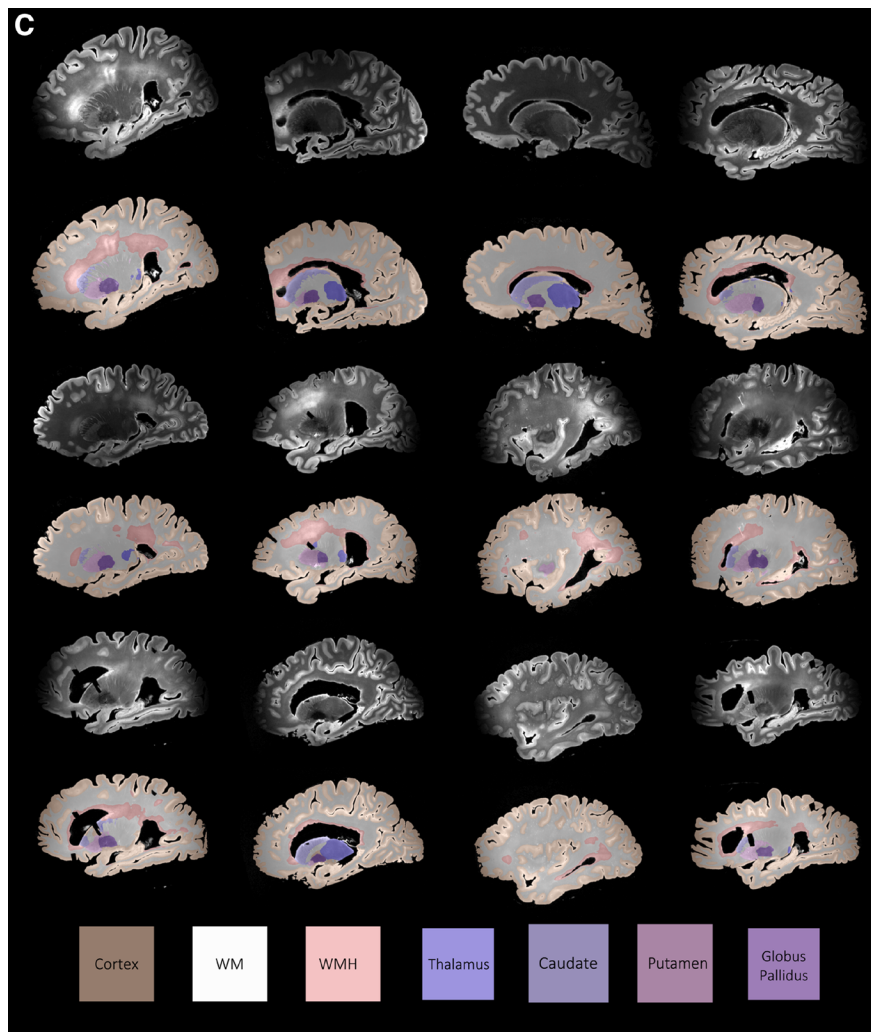**Fig. 10.** (A, B, C) Automated segmentations by nnU-Net of cortical gray matter, WMH, WM and the four subcortical structures for a sample of 36 subjects in the cohort.
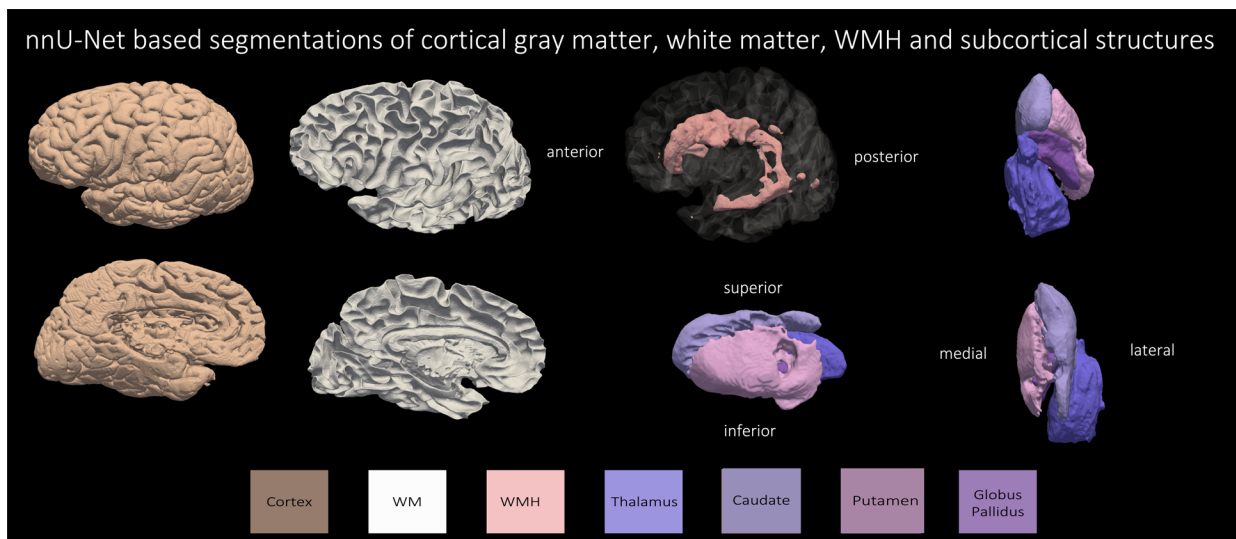
Fig. 11. Three-dimensional renderings of automated segmentations by nnU-Net of cortical gray matter, WM, WMH and the four subcortical structures for the subject with primary age-related tauopathy and cerebrovascular disease.
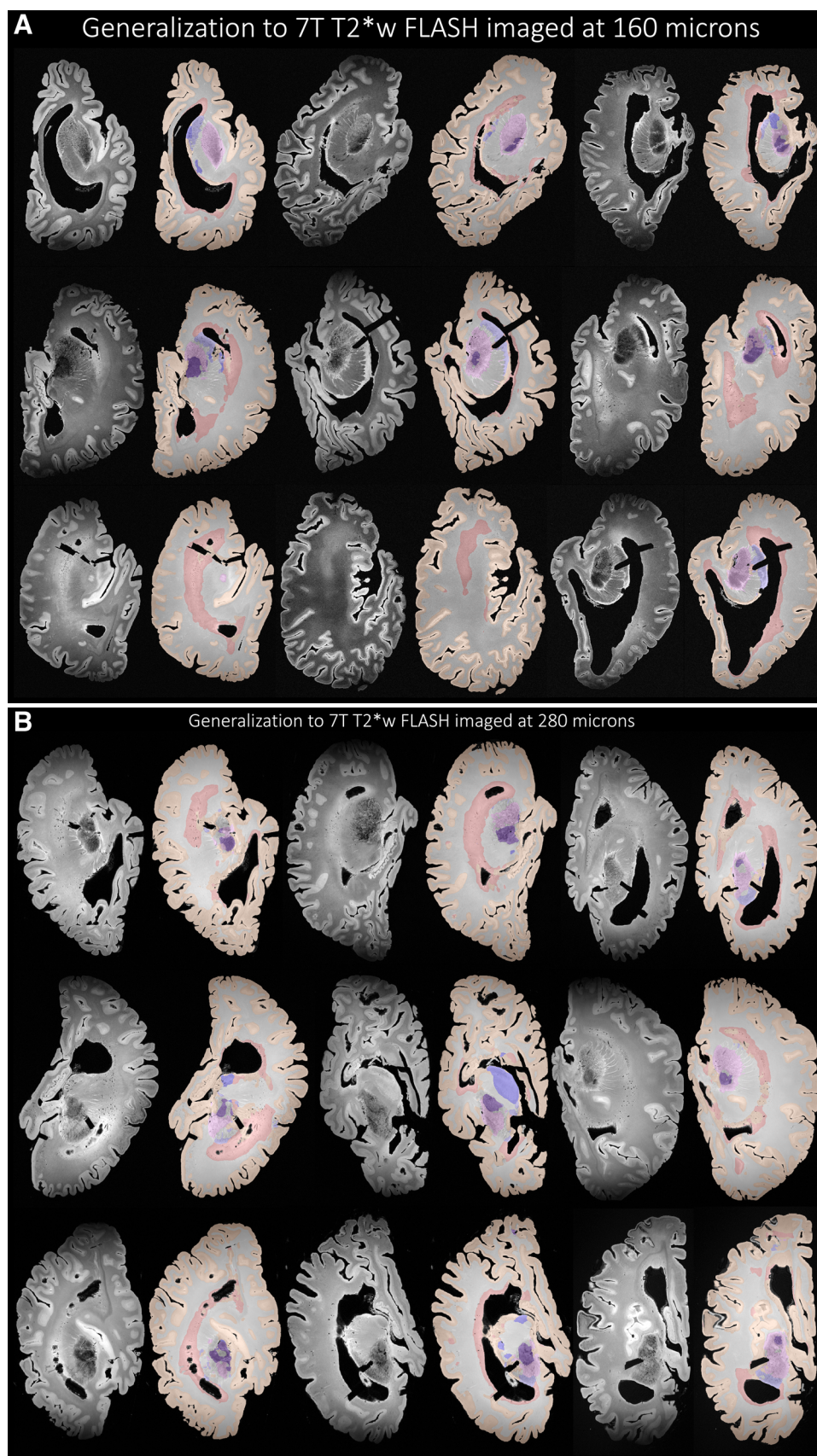
Fig. 12. Generalization to other unseen imaging sequences. (A) The nnU-Net architecture trained on 7 T T2w images at 0.3 x 0.3 x 0.3 mm^3 generalized well to images acquired at 7 T T2*w at a high resolution of 0.16 x 0.16 x 0.16 mm^3. (B) The nnU-Net architecture trained on 7 T T2w images at 0.3 x 0.3 x 0.3 mm^3 generalized well to images acquired at 7 T T2*w at a high resolution of 0.28 x 0.28 x 0.28 mm^3.

Table 4. Morphometry associations with underlying neuropathology measurements.

ROI	Abeta		Braak stage		CERAD		p-tau		Neuronal loss	
---	rho	CI	rho	CI	rho	CI	rho	CI	rho	CI
Visual	-0.176	[-1.0, 0.01]	-0.151	[-1.0, 0.04]	-0.17	[-1.0, 0.02]	-0.003	[-1.0, 0.19]	0.1	[-1.0, 0.28]
Motor	-0.139	[-1.0, 0.05]	-0.135	[-1.0, 0.06]	-0.063	[-1.0, 0.13]	NA	NA	NA	NA
Posterior cingulate	-0.46***	[-1.0, -0.27]	-0.445***	[-1.0, -0.25]	-0.419***	[-1.0, -0.23]	-0.357**	[-1.0, -0.16]	-0.303	[-1.0, -0.1]
Midfrontal	-0.441****	[-1.0, -0.28]	-0.327**	[-1.0, -0.15]	-0.271	[-1.0, -0.09]	-0.383***	[-1.0, -0.21]	-0.256	[-1.0, -0.07]
Anterior cingulate	-0.113	[-1.0, 0.09]	-0.16	[-1.0, 0.05]	-0.168	[-1.0, 0.04]	-0.185	[-1.0, 0.02]	-0.308	[-1.0, -0.11]
Orbitofrontal	0.105	[-1.0, 0.29]	0.219	[-1.0, 0.39]	0.122	[-1.0, 0.3]	NA	NA	NA	NA
Superior temporal	-0.325**	[-1.0, -0.14]	-0.289	[-1.0, -0.1]	-0.271	[-1.0, -0.09]	-0.293	[-1.0, -0.11]	-0.249	[-1.0, -0.06]
Inferior frontal	-0.178	[-1.0, 0.01]	-0.094	[-1.0, 0.1]	-0.145	[-1.0, 0.04]	-0.171	[-1.0, 0.02]	-0.145	[-1.0, 0.04]
Anterior insula	-0.202	[-1.0, -0.02]	-0.17	[-1.0, 0.02]	-0.222	[-1.0, -0.04]	-0.201	[-1.0, -0.01]	-0.215	[-1.0, -0.03]
Anterior temporal	-0.229	[-1.0, -0.04]	-0.266	[-1.0, -0.08]	-0.136	[-1.0, 0.05]	-0.238	[-1.0, -0.05]	-0.194	[-1.0, -0.01]
Ventrolaterl temporal cortex	-0.1	[-1.0, 0.13]	-0.175	[-1.0, 0.06]	-0.053	[-1.0, 0.18]	-0.114	[-1.0, 0.12]	-0.05	[-1.0, 0.18]
Superior pareital	-0.298	[-1.0, -0.11]	-0.225	[-1.0, -0.03]	-0.25	[-1.0, -0.06]	-0.295	[-1.0, -0.1]	-0.1	[-1.0, 0.1]
Angular gyrus	-0.166	[-1.0, 0.02]	-0.225	[-1.0, -0.04]	-0.109	[-1.0, 0.08]	-0.244	[-1.0, -0.06]	-0.139	[-1.0, 0.05]
Entorhinal cortex	-0.429***	[-1.0, -0.26]	-0.522****	[-1.0, -0.37]	-0.335**	[-1.0, -0.15]	-0.348**	[-1.0, -0.17]	-0.414***	[-1.0, -0.24]
Brodmann Area 35	-0.158	[-1.0, 0.04]	-0.189	[-1.0, 0.0]	-0.265	[-1.0, -0.08]	-0.179	[-1.0, 0.01]	-0.285	[-1.0, -0.1]
Parahippocampal cortex	-0.307	[-1.0, -0.13]	-0.328**	[-1.0, -0.15]	-0.25	[-1.0, -0.06]	-0.133	[-1.0, 0.06]	-0.188	[-1.0, 0.0]

Shown is the one-sided Spearman's correlation, controlling for age, sex, and PMI, between regional cortical thickness measures derived from topologically corrected nnUNet-CRUISE gray matter segmentation with corresponding regional ratings of p-tau pathology, neuronal loss density, and global ratings of amyloid-β, CERAD, and Braak staging. Each cell is color coded, with darker shades indicating more negative correlations. The asterisk indicates that the test survived Bonferroni multiple testing correction (Bonferroni, 1935). CI indicates 95% confidence interval.
*:0.01 < p ≤ 0.05; **:0.001 < p ≤ 0.01; ***:0.0001 < p ≤ 0.001; ****:0.00001 < p ≤ 0.0001.

- in posterior cingulate ($r = -0.357$, $p = 0.0024$), mid-frontal ($r = -0.383$, $p = 0.0003$), and entorhinal cortex ($r = -0.348$, $p = 0.0011$) with regional p-tau rating;
- in entorhinal cortex ($r = -0.414$, $p = 0.0001$) with regional neuronal loss rating.

Supplementary Table 3 shows the one-sided Spearman's correlation (controlling for age, sex and PMI) between the cortical thickness measures derived from the topologically corrected nnU-Net-CRUISE gray matter segmentation with the corresponding medial temporal lobe (MTL) ratings of p-tau pathology and neuronal loss density as the MTL is a region linked to early neurodegeneration in Alzheimer's disease. The MTL pathology ratings are computed by taking the average pathology of the entorhinal cortex, CA1 and subiculum, and the dentate gyrus. We notice significant results in the entorhinal cortex ($r = -0.53$, $p = 10-5$) with the MTL p-tau rating; and in posterior cingulate ($r = -0.396$, $p = 0.0008$), entorhinal cortex ($r = -0.47$, $p = 10-5$), and BA35 ($r = -0.321$, $p = 0.0026$) with the MTL neuronal loss density.

Separately, Supplementary Table 4 tabulates the morphometry associations with the underlying neuropathology for the AD continuum (N = 21) subjects from the subset of the cohort which have the manual segmentations (N = 36) as described in Section 3.2. We observe that the analysis based on manual segmentations shows a similar trend as the nnU-Net-CRUISE automated segmentations, suggesting that automated segmentations obtained from the developed pipeline provide meaningful associations with the underlying neuropathological measurements, are thus reliable for validation of clinical ratings, and can act as surrogates for the time-consuming user-supervised segmentations. Finally, Supplementary Figures 7-13 depict all the correlations as a plot along with the p-values to appreciate the data distribution.

4.4.2. Normalized WMH volume correlation patterns with regional cortical thickness and subcortical volumes

Table 5 compares the correlation between the regional cortical thickness and subcortical volumes with the normalized WMH (WMH volume divided by the total WM volume) volume for measurements based on automated nnU-Net-CRUISE for the subjects in the AD continuum (N = 82). We observed significant negative one-sided Spearman's correlation which survived Bonferroni multiple corrections in posterior cingulate ($r = -0.448$, $p = 0.0001$) and midfrontal cortex ($r = -0.325$, $p = 0.0019$) with normalized WMH volume for the automated nnU-Net

Table 5. White matter hyperintensities volume correlations.

ROI	rho	CI
Visual	0.243	[-1.0, 0.41]
Motor cortex	-0.256	[-1.0, -0.07]
Posterior cingulate	-0.448 ***	[-1.0, -0.26]
Midfrontal	-0.325 **	[-1.0, -0.15]
Anterior cingulate	-0.262	[-1.0, -0.06]
Orbitofrontal	0.355	[-1.0, 0.51]
Superior temporal pole	-0.205	[-1.0, -0.02]
Inferior frontal	-0.182	[-1.0, 0.01]
Anterior insula	0.011	[-1.0, 0.2]
Anterior temporal pole	-0.050	[-1.0, 0.14]
Ventrolateral temporal cortex	-0.142	[-1.0, 0.09]
Superior parietal	-0.193	[-1.0, 0.0]
Angular gyrus	-0.243	[-1.0, -0.06]
Entorhinal cortex	-0.293	[-1.0, -0.11]
Brodmann Area 35	-0.156	[-1.0, 0.04]
Parahippocampal cortex	-0.148	[-1.0, 0.04]
Caudate	-0.399***	[-1.0, -0.23]
Globus Pallidus	-0.193	[-1.0, -0.01]
Putamen	-0.312	[-1.0, -0.13]
Thalamus	-0.352***	[-1.0, -0.18]

Shown is the one-sided Spearman's correlation between normalized WMH volume (obtained by dividing the WMH volume with the corresponding WM) with regional cortical thickness based on nnU-Net-CRUISE. For the four subcortical structures, partial one-sided Spearman's correlation, with uncorrected p-value in brackets, is shown. All the tests were controlled for age, sex, and PMI. The asterisk indicates that the test survived Bonferroni multiple tests correction. Each cell is color coded, with darker shades indicating more negative correlations. CI indicates 95% confidence interval. *:$0.01 < p \leq 0.05$; **:$0.001 < p \leq 0.01$; ***:$0.0001 < p \leq 0.001$; ****:$0.00001 < p \leq 0.0001$.

segmentation-based regional thickness measurements; and in caudate ($r = -0.399$, $p = 0.0001$) and thalamus ($r = -0.352$, $p = 0.0007$) for the subcortical structures when correlated with the normalized WMH volume. All the tests were controlled for age, sex, and PMI. Supplementary Figures 14–16 depict these correlations as a plot along with the p-values to observe the data distribution.

5. DISCUSSION

5.1. Segmentation pipeline

To our knowledge, the current study is the most comprehensive assessment of automated segmentation of 7 T postmortem human brain MRI. Our evaluation is performed in a large cohort of 135 brain specimens with a range of neurodegenerative pathologies, and focuses on multiple tasks: cortical gray matter segmentation, subcortical gray matter structure segmentation, as well as white matter and WMH segmentation. For cortical segmentation, we evaluated nine deep learning architectures using direct metrics of segmentation accuracy (cross-validation DSC), derived morphological metrics (ICC of regional

cortical thickness with the reference standard; comparison of associations with pathology), and visual assessment.

Our paper stands apart from recent work on automated segmentation of postmortem brain MRI, which has either been performed in lower-resolution 3T MRI scans (Mancini et al., 2020), or in smaller high-resolution datasets (Jonkman et al., 2019). Mancini et al. (2020) used Free-Surfer (Fischl, 2012) and a Bayesian modeling technique, SAMSEG (Puonti et al., 2016), to map a single postmortem specimen imaged at 3 T but did not evaluate on higher resolution 7 T. In addition to segmenting the cortical gray matter, Mancini et al. (2020) parcellate the cortex into anatomical regions. A separate study (Kotrotsou et al., 2014) also segmented the entire hemisphere but applied to a smaller dataset of 7 subjects with a slice thickness of 1.5 mm and imaged at 3 T. Both these methods relied on multi-atlas based image segmentation which is dependent on registration between high-resolution postmortem and low-resolution antemortem atlases. Registration between inter-modality postmortem and antemortem currently remains challenging, especially for higher-resolution 7T postmortem MRI (Casamitjana et al., 2021; Casamitjana et al., 2022; Daly et al., 2021). Other recent studies on postmortem human brain morphometry (Adler et al., 2014, 2018; Augustinack et al., 2014; DeKraker et al., 2018, 2021; Iglesias et al., 2015; Ravikumar et al., 2021; Wisse et al., 2021; Yushkevich et al., 2021) focused on specific areas such as the hippocampus or the MTL, and relied on manual segmentation to guide inter-subject registration and atlas generation.

Our evaluation demonstrates that deep learning-based segmentation pipelines, particularly nnU-Net, can generate high-quality segmentations of cortical gray matter, subcortical structures, normal-appearing white matter, and white matter lesions even with very limited training data. With inference time of around 15 minutes on a CPU, our pipeline represents the first step towards fast, automated, and reliable brain mapping of high-resolution postmortem whole-hemisphere MRI. Our nnU-Net based pipeline generalized well to areas of low contrast unseen during training, as well as to other MRI protocols, and resolutions. Furthermore, our pipeline imposed post-hoc constraints on the automated segmentations to produce geometrically accurate and topologically correct segmentations using deformable surface-based methods, which have shown tremendous success in the antemortem literature. Therefore, we were able to correctly label the challenging boundary deep sulci and the cortical gray matter, leading to more reliable cortical thickness estimates. Moreover, thickness measures derived from the deep learning-based automated segmentations concur with the reference standard, and

similar associations between thickness and pathology are detected using automatically derived and reference standard thickness measurements, albeit with the latter usually having higher effect sizes. This suggests that fully automated cortical thickness analysis is feasible for postmortem MRI. Indeed, with further improvements to accuracy (e.g., a larger training set covering more of the hemisphere), automated postmortem segmentation may make the labor-intensive and subjective semi-automated approach to cortical thickness measurement unnecessary. Thus, the study suggests the feasibility of a fully automated group-wise cortical thickness analysis in postmortem MRI analogous to the way FreeSurfer is used for antemortem MRI morphometry, which forms the basis of our future research direction.

The limitations of the current pipeline include a relatively small whole hemisphere cortical gray matter segmentation training set which limited our ability to use direct metrics such as DSC to evaluate overall segmentation accuracy. Indeed, the method that performed best in terms of cross-validation DSC (AnatomyNet) performed worse in areas unseen during training than nnU-Net. To address this limitation, in future work we plan to train the method on manual segmentations of whole hemispheres; however, generating such a dataset will require a significant additional investment in time and effort. A limitation in the cross-validation experiment comparing the nine network architectures (Table 2) is that the 30 patches were not stratified by specimen during cross-validation, that is, patches from the same specimen appeared in both training and test subsets, which may have led to slight overestimation of Dice coefficients in Table 2. However, the patches within each subject were very different in terms of anatomical location, and crucially, the main validation experiment in this paper (ICC between expert-guided and automatic cortical thickness measurements in Table 4) is free of such data leakage. We still rely on manually placed landmarks to measure thickness in specific anatomical regions, and we do not show 3D maps of thickness as is common in antemortem morphometry studies. For example, the variability in the placement of the BA35 landmarks may explain our observation of the stronger thickness/pathology associations in the ERC than in BA35, even though BA35 is thought to have early tau pathology. Moreover, we noticed that four regions (BA35, superior parietal, angular gyrus, and the visual cortices) have the lowest ICC scores between the manual reference and automated nnU-Net-CRUISE segmentations. In Supplementary Figure 6, we show a sample segmentation where we observe some discrepancy between the manual reference and automated nnU-Net-CRUISE segmentations. Some of the possible reasons for these discrepancies are in fact: over- or under-segmentation,

MRI artefacts, or, as mentioned above, the lack of precision in the placement of landmark on which the thickness computation is directly dependent.

Due to limited availability of reference standard segmentations, parcellating the brain into subregions, as in done in most antemortem MRI, currently remains a challenge. We intend to address this limitation by semi-manually annotating the brain into different cortical and subcortical structures, guided by anatomical priors derived from antemortem MRI studies. We could then develop techniques for groupwise normalization studies by building a template for postmortem MRI. Towards this goal, we are currently developing deep learning-based methods for registration between postmortem and antemortem MRI. Our work is limited to ADRD, but in future we plan to evaluate our methodological pipeline on a cohort of non-demented specimens obtained from a separate postmortem dataset (Boon et al., 2019; Frigerio et al., 2021; Jonkman et al., 2019).

Certain limitations of the postmortem imaging should be noted. Wisse et al. (2017) compared the cortical thickness of the MTL substructures between antemortem and postmortem; and also postmortem (3T) and postmortem (9.4T) MRI. They found differences in thickness on different MRI scans and attributed to the various factors such as: (1) difference in the actual size between the antemortem and postmortem tissue is due to an actual difference in size as studies have suggested that tissue changes may occur during or after death since the agonal state causes hypoxia and ischemia which results in brain swelling. (2) an increase in size could result from brain extraction, for example, by a relief of intracranial pressure after autopsy. (3) formalin fixation could cause the underlying differences by shrinking the brain after several weeks.

Separately, at the moment, we do not have a way to quantitatively assess the accuracy of the T2*w FLASH images. In follow-up work, we plan to collect reference data by manually correcting the mis-segmentations and then re-train the deep learning model in a few-shot learning setting to achieve a boost in segmentation performance for T2*w FLASH MRI. Then, we will quantitatively assess the segmentation performance for the said generalization procedure using reference user-supervised manually corrected reference segmentation.

5.2. Neuropathology associations

With the help of the proposed segmentation and morphometry pipeline, we were able to conduct studies of postmortem MRI that have not been possible before: for example, we replicated some of the findings from Sadaghiani et al. (2022), and drew associations between WMH volume, cortical thickness, and subcortical volumes. Here, we discuss some of the interesting findings. Strong postmortem image analysis frameworks allow us to better understand the distinct roles and degrees by which antemortem pathology affects neurodegeneration. Prior work shows how amyloid-β, p-tau, TDP-43, etc. have differential influences on atrophy (Dugger & Dickson, 2017; Matej et al., 2019; Negash et al., 2011; Robinson et al., 2018), and an automated pipeline and dataset as shown here can bolster the quantitative interrogation of these open questions. It will allow future work to study links between macro structure and other local processes beyond pathology, including inflammatory markers, gene expression, etc. We also note that measuring WMH in postmortem imaging adds value to histological studies, as we have less clear measures of vascular burden with traditional autopsy, which will allow probing of WMH for better understanding their pathologic correlates given their non-specific nature.

The current study demonstrates the associations between regional cortical thickness measurements with the underlying semi-quantitative neuropathological ratings for the AD cohort. Negative correlations between p-tau and cortical thickness were found to be significant in angular gyrus and midfrontal regions, which is in line with previous research in antemortem (Das et al., 2019; Harrison et al., 2021; LaPoint et al., 2017; Whitwell et al., 2018; Xia et al., 2017) and postmortem *in situ* MRI (Frigerio et al., 2021) studies. Tau pathology is concurrent with neuronal loss in ADRD (Dawe et al., 2011; Jack Jr et al., 2018; Ohm et al., 2021) and the loss of neurons is likely a key source of cortical atrophy. We observed that cortical thickness showed significant negative correlation with neuronal loss in BA35 and entorhinal cortex, regions where p-tau pathology has predicted the atrophy rate (La Joie et al., 2020; LaPoint et al., 2017; Xie et al., 2018) in antemortem studies. Significant negative correlations were observed between Braak staging and cortical thickness in midfrontal, ERC and BA35, regions consistent with high p-tau uptake in positron emission tomography (PET) imaging with cortical thickness on MRI. Tau pathology in Braak regions plays an important role in cortical atrophy and cognitive decline during the course of AD. Similar findings are reported for global cortical thickness with *in situ* postmortem MRI in Frigerio et al. (2021). The relationship between amyloid-β and neurodegeneration is thought to be rather indirect (Gómez-Isla & Frosch, 2022; Jack Jr et al., 2018). Nevertheless, we did find strong negative correlations between thickness and amyloid-β in midfrontal, inferior frontal, and ventrolateral temporal cortex, brain regions implicated in working memory capacity (Barbey et al., 2013; Chiou & Ralph, 2018). Our observation of a significant negative correlation of CERAD

scores with cortical thickness in the superior parietal region is consistent with previous studies relating CERAD with cortical thickness (Paajanen et al., 2013; Santos et al., 2011).

Lastly, WMH has been implicated in age-related cognitive decline and AD, which is characterized by atrophy in the cortical mantle and the MTL (Dadar et al., 2022; Du et al., 2005; Rizvi et al., 2018). In our study, we observed significant negative correlations between normalized WMH and thickness in posterior cingulate and superior temporal regions. Previous work (Reijmer et al., 2015) showed that the disruption of structural and functional connectivity has an impact on executive functioning and memory among individuals with high WMH volume. To this point, our study found that subcortical atrophy was significantly negatively correlated with WMH volume in caudate and thalamus, suggesting more global effects on brain volume.

In our dataset, among the 135 specimens, 82 had AD pathology with existence of co-pathologies. Future studies may apply this dataset and pipeline to help disentangle the differential contributions of unique pathologies to individual atrophy patterns. Separately, we are aware that the pathology measures and MRI segmentation-based measures were obtained from contralateral hemispheres, which could potentially weaken the observed associations. But pathology in AD is usually largely symmetrical between the hemispheres, and therefore leaves less room for biases in the observed correlations, as claimed in a recent study (Ravikumar et al., 2021) which showed that correlations between MTL thickness maps and both contralateral and ipsilateral semi-quantitative p-tau pathology scores did not detect substantially different correlation patterns. Overall, the fact that strong thickness-pathology associations were observed for many brain regions even in this sub-optimal setting, supports our overall hypothesis that thickness measures derived using the proposed automatic pipeline are suitable for brain morphometry-pathology association studies. In future work, we plan to examine associations between these thickness measures and ipsilateral quantitative measures of pathology matched via multi-modality image registration.

Another limitation is that our study relies on semi-quantitative measures of neuropathology, which are subjective and might not reflect a linear pathology burden. We are currently in the process of obtaining neuropathology measurements from the same hemisphere histology, and developing machine learning-based quantitative pathological ratings to further validate our work. The current study provides support for future work to use larger datasets and quantitative pathology measures to describe the contribution of multiple pathologies to brain morphology in neurodegenerative diseases. But, overall

we observe a similar trend as described in our recent work (Sadaghiani et al., 2022), which looked at regional cortical thickness with p-tau burden. These limitations could be avoided by expanding our analysis to a larger dataset, which we are actively working towards.

Lastly, we should mention that our postmortem imaging project was launched shortly before the COVID-19 pandemic, which greatly interfered with our ability to maintain consistency in formalin fixation. This has been addressed in our Center's more recent autopsies, where we aim for consistent 60 days fixation. Like other limitations, we hypothesize that if formalin fixation had been more consistent, that would only lead to stronger associations between structure and pathology. The plot in Supplementary Figure 5 shows the mismatch in thickness (in mm) between automated (nnU-Net-CRUISE) and manual reference segmentations versus fixation time (in days). The plot does not reveal a systematic relationship between fixation time and thickness mismatch, neither in terms of bias nor in terms of variance. This is consistent with the literature that shows relative plateauing of T2 values in postmortem MRI after initial 1-2 months fixation (Dawe et al., 2009). We conclude that the large variation in fixation times in our study did not significantly affect the thickness computation based on the obtained segmentations.

6. CONCLUSION

While there is increased interest in using high-resolution postmortem MRI of the human brain for discovering associations between brain structure and pathology, automated tools for the analysis of such complex images have received much less attention compared to *antemortem* MRI. Our study used a relatively large dataset of 135 high-resolution T2w 7 T postmortem whole-hemisphere MRI scans to evaluate multiple deep learning image segmentation architectures and to develop an automatic segmentation pipeline that labels cortical gray matter, four subcortical structures (caudate, globus pallidus, putamen, and thalamus), WMH, and normal-appearing white matter. We report good agreement between thickness measures derived from our deep learning pipeline with the reference standard of semi-automated thickness measurement. Our analysis linking morphometry measures and pathology demonstrated that automated analysis of postmortem MRI yields similar findings to a labor-intensive semi-automated approach, and more broadly, that automated segmentation of postmortem MRI can complement and inform antemortem neuroimaging in neurodegenerative diseases. We have released our pipeline as a stand-alone containerized tool that can be readily applied to other postmortem brain datasets.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Human brain specimens were obtained in accordance with local laws and regulations, and include informed consent from next of kin at time of death or where possible, pre-consent during life.

CONSENT FOR PUBLICATION

All authors have reviewed the contents of the manuscript being submitted, approved of its contents, and validated the accuracy of the data and consented to publication.

DATA AND CODE AVAILABILITY

We have provided the code, scripts, and Jupyter notebooks to reproduce the findings of this study at the project webpage. The MRI data will be available upon request due to compliance and ethical issues.

AUTHOR CONTRIBUTIONS

Pulkit Khandelwal: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Roles/Writing—original draft, and Writing—review and editing. Michael Tran Duong: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, and Writing—review and editing. Shokufeh Sadaghiani: Conceptualization, Data curation, Resources, and Methodology. Sydney Lim: Data curation, Resources, and Methodology. Amanda E. Denning: Data curation, Resources, and Methodology. Eunice Chung: Data curation, Resources, and Methodology. Sadhana Ravikumar: Conceptualization, Validation, and Writing—review and editing. Sanaz Arezoumandan: Data curation, Resources. Claire Peterson: Data curation, Resources. Madigan Bedard: Data curation, Resources. Noah Capp: Data curation, Resources, And Software. Ranjit Ittyerah: Data curation, Resources. Elyse Migdal: Data curation, Resources, and Methodology. Grace Choi: Data curation, Resources, and Methodology. Emily Kopp: Data curation, Resources, and Methodology. Bridget Loja: Data curation, Resources, and Methodology. Eusha Hasan: Data curation, Resources, and Methodology. Jiacheng Li: Data curation, Resources, and Methodology. Alejandra Bahena: Data curation, Resources. Karthik Prabhakaran: Data curation, Resources. Gabor Mizsei: Data curation, Resources. Marianna Gabrielyan: Data curation, Resources. Theresa Schuck: Data curation, Resources. Winifred Trotman: Data curation, Project administration, Resources. John Robinson: Data curation, Resources. Daniel T. Ohm: Data curation, Resources, and Project administration. Edward B. Lee: Data curation, Funding acquisition, Investigation, Project administration, Resources, and Writing—review and editing. John Q. Trojanowski: Data curation, Project administration, Resources, and Funding acquisition. Corey McMillan: Data curation, Resources, and Funding acquisition. Murray Grossman: Conceptualization, Data curation, and Funding acquisition. David J. Irwin: Data curation, Project administration, Resources, Funding acquisition, Validation, and Writing—review and editing. John A. Detre: Data curation, Project administration, Resources, Validation, and Writing—review and editing. M. Dylan Tisdall: Data curation, Project administration, Resources, Validation, and Writing—review and editing. Sandhitsu R. Das: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, and Writing—review and editing. Laura E. M. Wisse: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, and Writing—review and editing. David A. Wolk: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Roles/Writing—original draft, and Writing—review and editing. Paul A. Yushkevich: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, and Writing—review and editing.

FUNDING

DECLARATION OF COMPETING INTEREST

David A. Wolk has received grant support from Merck, Biogen, and Eli Lilly/Avid. David A. Wolk received consultation fees from Neuronix, Eli Lilly, and Qynaps and is on the Data and Safety Monitoring Board for a clinical trial run by Functional Neuromodulation. John Q. Trojanowski received revenue from the sale of Avid to Eli Lilly as co-inventor on imaging-related patents submitted by the University of Pennsylvania. David J. Irwin is member of the science advisory board of Denali Therapeutics. Sandhitsu R. Das received consultation fees from Rancho Biosciences and Nia Therapeutics. The other authors have nothing to disclose.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00171.

REFERENCES

Adler, D. H., Pluta, J., Kadivar, S., Craige, C., Gee, J. C., Avants, B. B., & Yushkevich, P. A. (2014). Histology-derived volumetric annotation of the human hippocampal subfields in postmortem MRI. *NeuroImage*, *84*, 505–523. https://doi.org/10.1016/j.neuroimage.2013.08.067

Adler, D. H., Wisse, L. E., Ittyerah, R., Pluta, J. B., Ding, S. L., Xie, L., Wang, J., Kadivar, S., Robinson, J. L., Schuck, T., Trojanowski, J. Q., Grossman, M., Detre, J. A., Elliott, M. A., Toledo, J. B., Liu, W., Pickup, S., Miller, M. I., Das, S. R., … Yushkevich, P. A. (2018). Characterizing the human hippocampus in aging and Alzheimer's disease using a computational atlas derived from ex vivo MRI and histology. *Proceedings of the National Academy of Sciences*, *115*, 4252–4257. https://doi.org/10.1073/pnas.1801093115

Alkemade, A., Bazin, P. L., Balesar, R., Pine, K., Kirilina, E., Möller, H. E., Trampel, R., Kros, J. M., Keuken, M. C., Bleys, R. L., Swaab, D. F., Herrler, A., Weiskopf, N., & Forstmann, B. U. (2022). A unified 3D map of microscopic architecture and MRI of the human brain. *Science Advances*, *8*, eabj7892. https://doi.org/10.1126/sciadv.abj7892

Amunts, K., Mohlberg, H., Bludau, S., & Zilles, K. (2020). Julich-brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*, *369*, 988–992. https://doi.org/10.1126/science.abb4588

Arezoumandan, S., Xie, S. X., Cousins, K. A., Mechanic-Hamilton, D. J., Peterson, C. S., Huang, C. Y., Ohm, D. T., Ittyerah, R., McMillan, C. T., Wolk, D. A., Yushkevich, P., Trojanowski, J. Q., Lee, E. B., Grossman, M., Phillips, J. S., & Irwin, D. J. (2022). Regional distribution and maturation of tau pathology among phenotypic variants of Alzheimer's disease. *Acta Neuropathologica*, *114*, 1103–1116. https://doi.org/10.1007/s00401-022-02472-x

Augustinack, J. C., Helmer, K., Huber, K. E., Kakunoori, S., Zöllei, L., & Fischl, B. (2010). Direct visualization of the perforant pathway in the human brain with ex vivo diffusion tensor imaging. *Frontiers in Human Neuroscience*, *4*, 42. https://doi.org/10.3389/fnhum.2010.00042

Augustinack, J. C., Magnain, C., Reuter, M., van der Kouwe, A. J., Boas, D., & Fischl, B. (2014). MRI parcellation of ex vivo medial temporal lobe. *NeuroImage*, *93*, 252–259. https://doi.org/10.1016/j.neuroimage.2013.05.053

Augustinack, J. C., van der Kouwe, A. J., & Fischl, B. (2013). Medial temporal cortices in ex vivo magnetic resonance imaging. *Journal of Comparative Neurology*, *521*, 4177–4188. https://doi.org/10.1002/cne.23432

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, *49*, 1195–1205. https://doi.org/10.1016/j.cortex.2012.05.022

Bazin, P. L., & Pham, D. L. (2007). Topology correction of segmented medical images using a fast marching algorithm. *Computer Methods and Programs in Biomedicine*, *88*, 182–190. https://doi.org/10.1016/j.cmpb.2007.08.006

Beaujoin, J., Palomero-Gallagher, N., Boumezbeur, F., Axer, M., Bernard, J., Poupon, F., Schmitz, D., Mangin, J. F., & Poupon, C. (2018). Post-mortem inference of the human hippocampal connectivity and microstructure using ultra-high field diffusion MRI at 11.7 T. *Brain Structure and Function*, *223*, 2157–2179. https://doi.org/10.1007/s00429-018-1617-1

Bonferroni, C. E. (1935). Il calcolo delle assicurazioni su gruppi di teste. In Studi in onore del professore salvatore ortu carboni (pp. 13–60).

Boon, B. D., Pouwels, P. J., Jonkman, L. E., Keijzer, M. J., Preziosa, P., van de Berg, W. D., Geurts, J. J., Scheltens, P., Barkhof, F., Rozemuller, A. J. M., Bouwman, F. H., & Steenwijk, M. D. (2019). Can post-mortem MRI be used as a proxy for in-vivo? A case study. *Brain Communications*, *1*(1), fcz030. https://doi.org/10.1093/braincomms/fcz030

Bulk, M., Abdelmoula, W. M., Geut, H., Wiarda, W., Ronen, I., Dijkstra, J., & van der Weerd, L. (2020). Quantitative MRI and laser ablation-inductively coupled plasma-mass spectrometry imaging of iron in the frontal cortex of healthy controls and Alzheimer's disease patients. *NeuroImage*, *215*, 116808. https://doi.org/10.1016/j.neuroimage.2020.116808

Casamitjana, A., Lorenzi, M., Ferraris, S., Peter, L., Modat, M., Stevens, A., Fischl, B., Vercauteren, T., & Iglesias, J. E. (2022). Robust joint registration of multiple stains and MRI for multimodal 3D histology reconstruction: Application to the Allen human brain atlas. *Medical Image Analysis*, *75*, 102265. https://doi.org/10.1016/j.media.2021.102265

Casamitjana, A., Mancini, M., & Iglesias, J. E. (2021). Synth-by-Reg (SbR): Contrastive learning for synthesis-based registration of paired images. In *International Workshop on Simulation and Synthesis in Medical Imaging* (pp. 44–54). Springer. https://doi.org/10.1007/978-3-030-87592-3_5

Casamitjana, A., Mancini, M., Robinson, E., Peter, L., Annunziata, R., Althonayan, J., Crampsie, S., Blackburn, E., Billot, B., Atzeni, A., Puonti, O., Balbastre, Y., Schmidt, P., Hughes, J., Augustinack, J. C., Edlow, B. L., Zöllei, L., Thomas, D. L. Kliemann, D., … Iglesias, J. E. (2024). A next-generation, histological atlas of the human brain and its application to automated brain MRI segmentation. *bioRxiv*. https://doi.org/10.1101/2024.02.05.579016

Chen, H., Dou, Q., Yu, L., Qin, J., & Heng, P. A. (2018). Voxresnet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, *170*, 446–455. https://doi.org/10.1016/j.neuroimage.2017.04.041

Chiou, R., & Ralph, M. A. L. (2018). The anterior-ventrolateral temporal lobe contributes to boosting visual working memory capacity for items carrying semantic information. *NeuroImage*, *169*, 453–461. https://doi.org/10.1016/j.neuroimage.2017.12.085

Dadar, M., Manera, A. L., Ducharme, S., & Collins, D. L. (2022). White matter hyperintensities are associated with grey matter atrophy and cognitive decline in Alzheimer's

disease and frontotemporal dementia. *Neurobiology of Aging*, *111*, 54–63. https://doi.org/10.1016/j.neurobiolaging.2021.11.007

Daly, A. C., Geras, K. J., & Bonneau, R. (2021). A convolutional neural network for common coordinate registration of high-resolution histology images. *Bioinformatics*, *37*, 4216–4226. https://doi.org/10.1093/bioinformatics/btab447

Das, S. R., Xie, L., Wisse, L. E., Vergnet, N., Ittyerah, R., Cui, S., Yushkevich, P. A., Wolk, D. A., & Alzheimer's Disease Neuroimaging Initiative. (2019). In vivo measures of tau burden are associated with atrophy in early Braak stage medial temporal lobe regions in amyloid-negative individuals. *Alzheimer's & Dementia*, *15*, 1286–1295. https://doi.org/10.1016/j.jalz.2019.05.009

Dawe, R. J., Bennett, D. A., Schneider, J. A., & Arfanakis, K. (2011). Neuropathologic correlates of hippocampal atrophy in the elderly: A clinical, pathologic, postmortem MRI study. *PLoS One*, *6*, e26286. https://doi.org/10.1371/journal.pone.0026286

Dawe, R. J., Bennett, D. A., Schneider, J. A., Vasireddi, S. K., & Arfanakis, K. (2009). Postmortem MRI of human brain hemispheres: T2 relaxation times during formaldehyde fixation. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *61*, 810–818. https://doi.org/10.1002/mrm.21909

de Flores, R., Wisse, L. E., Das, S. R., Xie, L., McMillan, C. T., Trojanowski, J. Q., Robinson, J. L., Grossman, M., Lee, E., Irwin, D. J., Yushkevich, P. A., & Wolk, D. A. (2020). Contribution of mixed pathology to medial temporal lobe atrophy in Alzheimer's disease. *Alzheimer's & Dementia*, *16*, 843–852. https://doi.org/10.1002/alz.12079

DeKraker, J., Ferko, K. M., Lau, J. C., Köhler, S., & Khan, A. R. (2018). Unfolding the hippocampus: An intrinsic coordinate system for subfield segmentations and quantitative mapping. *NeuroImage*, *167*, 408–418. https://doi.org/10.1016/j.neuroimage.2017.11.054

DeKraker, J., Köhler, S., & Khan, A. R. (2021). Surface-based hippocampal subfield segmentation. *Trends in Neurosciences*, *44*, 856–863. https://doi.org/10.1016/j.tins.2021.06.005

Du, A. T., Schuff, N., Chao, L. L., Kornak, J., Ezekiel, F., Jagust, W. J., Kramer, J. H., Reed, B. R., Miller, B. L., Norman, D., Chui, H. C., & Weiner, M. W. (2005). White matter lesions are associated with cortical atrophy more than entorhinal and hippocampal atrophy. *Neurobiology of Aging*, *26*, 553–559. https://doi.org/10.1016/j.neurobiolaging.2004.05.002

Dugger, B. N., & Dickson, D. W. (2017). Pathology of neurodegenerative diseases. *Cold Spring Harbor Perspectives in Biology*, *9*, a028035. https://doi.org/10.1101/cshperspect.a028035

Eckermann, M., Schmitzer, B., van der Meer, F., Franz, J., Hansen, O., Stadelmann, C., & Salditt, T. (2021). Three-dimensional virtual histology of the human hippocampus based on phase-contrast computed tomography. *Proceedings of the National Academy of Sciences*, *118*, e2113835118. https://doi.org/10.1073/pnas.2113835118

Edlow, B. L., Mareyam, A., Horn, A., Polimeni, J. R., Witzel, T., Tisdall, M. D., Augustinack, J. C., Stockmann, J. P., Diamond, B. R., Stevens, A., Tirrell, L. S., Folkerth, R. D., Wald, L. L., Fischl, B., & van der Kouwe, A. (2019). 7 Tesla MRI of the ex vivo human brain at 100 micron resolution. *Scientific Data*, *6*, 1–10. https://doi.org/10.1038/s41597-019-0254-8

Fischl, B. (2012). Freesurfer. *NeuroImage*, *62*, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Frigerio, I., Boon, B. D., Lin, C. P., Galis-de Graaf, Y., Bol, J., Preziosa, P., Twisk, J., Barkhof, F., Hoozemans, J. J., Bouwman, F. H., Rozemuller, A. J. M., van de Berg, W. D. J., & Jonkman, L. E. (2021). Amyloid-β, p-tau and reactive microglia are pathological correlates of MRI cortical atrophy in Alzheimer's disease. *Brain Communications*, *3*, fcab281. https://doi.org/10.1093/braincomms/fcab281

García-Cabezas, M. Á., Hacker, J. L., & Zikopoulos, B. (2020). A protocol for cortical type analysis of the human neocortex applied on histological samples, the atlas of von Economo and Koskinas, and magnetic resonance imaging. *Frontiers in Neuroanatomy*, *14*, 576015. https://doi.org/10.3389/fnana.2020.576015

Gómez-Isla, T., & Frosch, M. P. (2022). Lesions without symptoms: Understanding resilience to Alzheimer disease neuropathological changes. *Nature Reviews Neurology*, *18*, 323–332. https://doi.org/10.1038/s41582-022-00642-9

Gordon, E., Rohrer, J. D., & Fox, N. C. (2016). Advances in neuroimaging in frontotemporal dementia. *Journal of Neurochemistry*, *138*, 193–210. https://doi.org/10.1111/jnc.13656

Han, X., Pham, D. L., Tosun, D., Rettmann, M. E., Xu, C., & Prince, J. L. (2004). Cruise: Cortical reconstruction using implicit surface evolution. *NeuroImage*, *23*, 997–1012. https://doi.org/10.1016/j.neuroimage.2004.06.043

Han, X., Xu, C., Braga-Neto, U., & Prince, J. L. (2002). Topology correction in brain cortex segmentation using a multiscale, graph-based algorithm. *IEEE Transactions on Medical Imaging*, *21*, 109–121. https://doi.org/10.1109/42.993130

Han, X., Xu, C., & Prince, J. L. (2003). A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 755–768. https://doi.org/10.1109/tpami.2003.1201824

Harrison, T. M., Du, R., Klencklen, G., Baker, S. L., & Jagust, W. J. (2021). Distinct effects of beta-amyloid and tau on cortical thickness in cognitively healthy older adults. *Alzheimer's & Dementia*, *17*, 1085–1096. https://doi.org/10.1002/alz.12249

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

Huntenburg, J. M., Steele, C. J., & Bazin, P. L. (2018). Nighres: Processing tools for high-resolution neuroimaging. *GigaScience*, *7*, giy082. https://doi.org/10.1093/gigascience/giy082

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., Dickson, D. W., Duyckaerts, C., Frosch, M. P., Masliah, E., Mirra, S. S., Nelson, P. T., Schneider, J. A., Thal, D. R., Thies, B., Trojanowski, J. Q., Vinters, H. V., & Montine, T. J. (2012). National institute on aging–Alzheimer's association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & Dementia*, *8*, 1–13. https://doi.org/10.1007/s00401-011-0910-3

Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., Roy, N., Frosch, M. P., McKee, A. C., Wald, L. L., Fischl, B., Van Leemput, K., & Alzheimer's Disease Neuroimaging Initiative. (2015). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *NeuroImage*, *115*, 117–137. https://doi.org/10.1016/j.neuroimage.2015.04.042

Iglesias, J. E., Insausti, R., Lerma-Usabiaga, G., Bocchetta, M., Van Leemput, K., Greve, D. N., Van der Kouwe, A., Fischl, B., Caballero-Gaudes, C., & Paz-Alonso, P. M. (2018). A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *NeuroImage*, *183*, 314–326. https://doi.org/10.1016/j.neuroimage.2018.08.012

Irwin, D. J. (2016). Progressive supranuclear palsy, corticobasal syndrome, and other tauopathies. In *International Neurology* (pp. 157–160). Wiley. https://doi.org/10.1002/9781118777329.ch44

Irwin, D. J., Cairns, N. J., Grossman, M., McMillan, C. T., Lee, E. B., Van Deerlin, V. M., Lee, V. M. Y., & Trojanowski, J. Q. (2015). Frontotemporal lobar degeneration: Defining phenotypic diversity through personalized medicine. *Acta Neuropathologica*, *129*, 469–491. https://doi.org/10.1007/s00401-014-1380-1

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

Jack, Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Contributors. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*, 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jonkman, L. E., Galis-de Graaf, Y., Bulk, M., Kaaij, E., Pouwels, P. J., Barkhof, F., Rozemuller, A. J., van der Weerd, L., Geurts, J. J., & van de Berg, W. D. (2019). Normal aging brain collection amsterdam (nabca): A comprehensive collection of postmortem high-field imaging, neuropathological and morphometric datasets of non-neurological controls. *NeuroImage: Clinical*, *22*, 101698. https://doi.org/10.1016/j.nicl.2019.101698

Kenkhuis, B., Jonkman, L. E., Bulk, M., Buijs, M., Boon, B. D., Bouwman, F. H., Geurts, J. J., van de Berg, W. D., & van der Weerd, L. (2019). 7t MRI allows detection of disturbed cortical lamination of the medial temporal lobe in patients with Alzheimer's disease. *NeuroImage: Clinical*, *21*, 101665. https://doi.org/10.1016/j.nicl.2019.101665

Khandelwal, P., Duong, M. T., Chung, E., Sadaghiani, S., Lim, S. A., Ravikumar, S., Arezoumandan, S., Peterson, C., Bedard, M. L., Capp, N., Ittyerah, R., Migdal, E., Choi, G., Kopp, E., Patino, B. L., Hasan, E., Li, J., Prabhakaran, K., Mizsei, G., … Yushkevich, P. A. (2022). Deep learning for ultra high resolution t2-weighted 7 Tesla ex vivo magnetic resonance imaging reveals differential subcortical atrophy across neurodegenerative diseases. *Alzheimer's & Dementia*, *18*, e062628. https://doi.org/10.1002/alz.062628

Khandelwal, P., Sadaghiani, S., Chung, E., Lim, S. A., Duong, M. T., Ravikumar, S., Arezoumandan, S., Peterson, C., Bedard, M. L., Capp, N., Ittyerah, R., Migdal, E., Choi, G., Kopp, E., Patino, B. L., Hasan, E., Li, J., Prabhakaran, K., Mizsei, G., … Yushkevich, P. A. (2022). Deep learning pipeline for cortical gray matter segmentation and thickness analysis in ultra high resolution t2w 7 Tesla ex vivo MRI across neurodegenerative diseases reveals associations with underlying neuropathology. *Alzheimer's & Dementia*, *18*, e065737. https://doi.org/10.1002/alz.065737

Khandelwal, P., Sadaghiani, S., Ravikumar, S., Lim, S., Arezoumandan, S., Peterson, C., Chung, E., Bedard, M., Capp, N., Ittyerah, R., Migdal, E., Choi, G., Kopp,

E., Loja, B., Hasan, E., Li, J., Prabhakaran, K., Mizsei, G., Gabrielyan, M., … Yushkevich, P. A. (2021). Gray matter segmentation in ultra high resolution 7 Tesla ex vivo t2w MRI of human brain hemispheres. *arXiv*, arXiv:2110.07711. https://doi.org/10.1002/alz.065737

Khandelwal, P., & Yushkevich, P. (2020). Domain generalizer: A few-shot meta learning framework for domain generalization in medical imaging. In *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning* (pp. 73–84). Springer. https://doi.org/10.1007/978-3-030-60548-3_8

Kiwitz, K., Schiffer, C., Spitzer, H., Dickscheid, T., & Amunts, K. (2020). Deep learning networks reflect cytoarchitectonic features used in brain mapping. *Scientific Reports*, *10*, 1–15. https://doi.org/10.1038/s41598-020-78638-y

Kotrotsou, A., Bennett, D. A., Schneider, J. A., Dawe, R. J., Golak, T., Leurgans, S. E., Yu, L., & Arfanakis, K. (2014). Ex vivo MR volumetry of human brain hemispheres. *Magnetic Resonance in Medicine*, *71*, 364–374. https://doi.org/10.1002/mrm.24661

La Joie, R., Visani, A. V., Baker, S. L., Brown, J. A., Bourakova, V., Cha, J., Chaudhary, K., Edwards, L., Iaccarino, L., Janabi, M., Lesman-Segev, O. H., Miller, Z. A., Perry, D. C., P O'Neil, J., Pham, J., Rojas, J. C., Rosen, H. J., Seeley, W. W., Tsai, R. M., … Rabinovici, G. D. (2020). Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Science Translational Medicine*, *12*, eaau5732. https://doi.org/10.1126/scitranslmed.aau5732

LaPoint, M. R., Chhatwal, J. P., Sepulcre, J., Johnson, K. A., Sperling, R. A., & Schultz, A. P. (2017). The association between tau PET and retrospective cortical thinning in clinically normal elderly. *NeuroImage*, *157*, 612–622. https://doi.org/10.1016/j.neuroimage.2017.05.049

Lasserve, J., Lim, S. A., Wisse, L., Ittyerah, R., Ravikumar, S., Lavery, M., Robinson, J. L., Schuck, T., Grossman, M., Lee, E. B., Yushkevich, P. A., Tisdall, D. M., Prabhakaran, K., Mizsei, G., Artacho-Perula, E., Martin, M. M. I. d. O., Jimenez, M. d. M. A., Munoz, M., Romero, F. J. M., … Insausti, R. (2020). Optimized extraction of the medial temporal lobe for postmortem MRI based on custom 3D printed molds. *Alzheimer's & Dementia*, *16*, e043254. https://doi.org/10.1002/alz.043254

Mackenzie, I. R., Neumann, M., Baborie, A., Sampathu, D. M., Du Plessis, D., Jaros, E., Perry, R. H., Trojanowski, J. Q., Mann, D., & Lee, V. M. (2011). A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathologica*, *122*, 111–113. https://doi.org/10.1007/s00401-011-0845-8

Makkinejad, N., Schneider, J. A., Yu, J., Leurgans, S. E., Kotrotsou, A., Evia, A. M., Bennett, D. A., & Arfanakis, K. (2019). Associations of amygdala volume and shape with transactive response DNA-binding protein 43 (TDP-43) pathology in a community cohort of older adults. *Neurobiology of Aging*, *77*, 104–111. https://doi.org/10.1016/j.neurobiolaging.2019.01.022

Mancini, M., Casamitjana, A., Peter, L., Robinson, E., Crampsie, S., Thomas, D. L., Holton, J. L., Jaunmuktane, Z., & Iglesias, J. E. (2020). A multimodal computational pipeline for 3D histology of the human brain. *Scientific Reports*, *10*, 1–21. https://doi.org/10.1038/s41598-020-69163-z

Matej, R., Tesar, A., & Rusina, R. (2019). Alzheimer's disease and other neurodegenerative dementias in comorbidity: A clinical and neuropathological overview. *Clinical Biochemistry*, *73*, 26–31. https://doi.org/10.1016/j.clinbiochem.2019.08.005

Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565–571). IEEE. https://doi.org/10.1109/3dv.2016.79

Negash, S., Bennett, D. A., Wilson, R. S., Schneider, J. A., & Arnold, S. E. (2011). Cognition and neuropathology in aging: Multidimensional perspectives from the rush religious orders study and rush memory and aging project. *Current Alzheimer Research*, *8*, 336–340. https://doi.org/10.2174/156720511795745302

Ogniewicz, R. L., & Ilg, M. (1992). Voronoi skeletons: Theory and applications. In *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 63–69). IEEE. https://doi.org/10.1109/cvpr.1992.223226

Ohm, D. T., Fought, A. J., Martersteck, A., Coventry, C., Sridhar, J., Gefen, T., Weintraub, S., Bigio, E., Mesulam, M. M., Rogalski, E., & Geula, C. (2021). Accumulation of neurofibrillary tangles and activated microglia is associated with lower neuron densities in the aphasic variant of Alzheimer's disease. *Brain Pathology*, *31*, 189–204. https://doi.org/10.1111/bpa.12902

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv*, arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999

Paajanen, T., Hänninen, T., Aitken, A., Hallikainen, M., Westman, E., Wahlund, L. O., Sobow, T., Mecocci, P., Tsolaki, M., Vellas, B., Muehlboeck, S., Spenger, C., Lovestone, S., Simmons, A., Soininen, H., & AddNeuroMed Consortium (2013). CERAD neuropsychological total scores reflect cortical thinning in prodromal Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders Extra*, *3*, 446–458. https://doi.org/10.1159/000356725

Pallebage-Gamarallage, M., Foxley, S., Menke, R. A., Huszar, I. N., Jenkinson, M., Tendler, B. C., Wang, C., Jbabdi, S., Turner, M. R., Miller, K. L., & Ansorge, O. (2018). Dissecting the pathobiology of altered MRI signal in amyotrophic lateral sclerosis: A post mortem whole brain sampling strategy for the integration of ultra-high-field MRI and quantitative neuropathology. *BMC Neuroscience*, *19*, 1–24. https://doi.org/10.1186/s12868-018-0416-1

Puonti, O., Iglesias, J. E., & Van Leemput, K. (2016). Fast and sequence-adaptive whole-brain segmentation using parametric bayesian modeling. *NeuroImage*, *143*, 235–249. https://doi.org/10.1016/j.neuroimage.2016.09.011

Ravikumar, S., Wisse, L. E., Ittyerah, R., Lim, S., Lavery, M., Xie, L., Robinson, J. L., Schuck, T., Grossman, M., Lee, E. B., Tisdall, M. D., Prabhakaran, K., Detre, J. A., Das, S. R., Mizsei, G., Artacho-Pérula, E., de Onzono Martin, M. M. I., Jiménez, M. M. A., Muñoz, M., … Yushkevich, P. A. (2020). Building an ex vivo atlas of the earliest brain regions affected by Alzheimer's disease pathology. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 113–117). IEEE. https://doi.org/10.1109/isbi45749.2020.9098427

Ravikumar, S., Wisse, L. E., Lim, S., Ittyerah, R., Xie, L., Bedard, M. L., Das, S. R., Lee, E. B., Tisdall, M. D., Prabhakaran, K., Lane, J., Detre, J. A., Mizsei, G., Trojanowski, J. Q., Robinson, J. L., Schuck, T., Grossman, M., Artacho-Pérula, E., Martin, M. M. I. O., … Yushkevich, P. A. (2021). Ex vivo MRI atlas of the human medial temporal lobe: Characterizing neurodegeneration due to tau pathology. *Acta Neuropathologica Communications*, *9*, 1–14. https://doi.org/10.1186/s40478-021-01275-7

Reijmer, Y. D., Schultz, A. P., Leemans, A., O'sullivan, M., Gurol, M. E., Sperling, R., Greenberg, S. M., Viswanathan, A., & Hedden, T. (2015). Decoupling of structural and functional brain connectivity in older adults with white matter hyperintensities. *NeuroImage*, *117*, 222–229. https://doi.org/10.1016/j.neuroimage.2015.05.054

Rizvi, B., Narkhede, A., Last, B. S., Budge, M., Tosto, G., Manly, J. J., Schupf, N., Mayeux, R., & Brickman, A. M. (2018). The effect of white matter hyperintensities on cognition is mediated by cortical atrophy. *Neurobiology of Aging*, *64*, 25–32. https://doi.org/10.1016/j.neurobiolaging.2017.12.006

Robinson, J. L., Lee, E. B., Xie, S. X., Rennert, L., Suh, E., Bredenberg, C., Caswell, C., Van Deerlin, V. M., Yan, N., Yousef, A., Hurtig, H. I., Siderowf, A., Grossman, M., McMillan, C. T., Miller, B., Duda, J. E., Irwin, D. J., Wolk, D., Elman, L., … Trojanowski, J. Q. (2018). Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain*, *141*, 2181–2193. https://doi.org/10.1093/brain/awy146

Roy, A. G., Navab, N., & Wachinger, C. (2018). Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE Transactions on Medical Imaging*, *38*, 540–549. https://doi.org/10.1109/tmi.2018.2867261

Sadaghiani, S., Trotman, W., Lim, S. A., Chung, E., Ittyerah, R., Ravikumar, S., Khandelwal, P., Prabhakaran, K., Lavery, M. L., Ohm, D. T., Gabrielyan, M., Das, S. R., Schuck, T., Capp, N., Peterson, C. S., Migdal, E., Artacho-Pérula, E., Jiménez, M. D. M. A., Rabal, M. D. P. M., … Yushkevich, P. A. (2022). Associations of phosphorylated tau pathology with whole-hemisphere ex vivo morphometry in 7 Tesla MRI. *Alzheimer's & Dementia*. https://doi.org/10.1002/alz.12884

Santos, V. D., Thomann, P. A., Wüstenberg, T., Seidl, U., Essig, M., & Schröder, J. (2011). Morphological cerebral correlates of cerad test performance in mild cognitive impairment and Alzheimer's disease. *Journal of Alzheimer's Disease*, *23*, 411–420. https://doi.org/10.3233/jad-2010-100156

Schiffer, C., Amunts, K., Harmeling, S., & Dickscheid, T. (2021). Contrastive representation learning for whole brain cytoarchitectonic mapping in histological human brain sections. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (pp. 603–606). IEEE. https://doi.org/10.1109/isbi48211.2021.9433986

Schiffer, C., Harmeling, S., Amunts, K., & Dickscheid, T. (2021). 2D histology meets 3D topology: Cytoarchitectonic brain mapping with graph neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 395–404). Springer. https://doi.org/10.1007/978-3-030-87237-3_38

Schiffer, C., Spitzer, H., Kiwitz, K., Unger, N., Wagstyl, K., Evans, A. C., Harmeling, S., Amunts, K., & Dickscheid, T. (2021). Convolutional neural networks for cytoarchitectonic brain mapping at large scale. *NeuroImage*, *240*, 118327. https://doi.org/10.1016/j.neuroimage.2021.118327

Schneider, J. A., Arvanitakis, Z., Bang, W., & Bennett, D. A. (2007). Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology*, *69*, 2197–2204. https://doi.org/10.1212/01.wnl.0000271090.28148.24

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420. https://doi.org/10.1037//0033-2909.86.2.420

Spitzer, H., Kiwitz, K., Amunts, K., Harmeling, S., & Dickscheid, T. (2018). Improving cytoarchitectonic segmentation of human brain areas with self-supervised Siamese networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 663–671). Springer. https://doi.org/10.1007/978-3-030-00931-1_76

Tisdall, M. D., Ohm, D. T., Lobrovich, R., Das, S. R., Mizsei, G., Prabhakaran, K., Ittyerah, R., Lim, S., McMillan, C. T., Wolk, D. A., Gee, J., Trojanowski, J. Q., Lee, E. B., Detre, J. A., Yushkevich, P., Grossman, M., & Irwin, D. J. (2022). Ex vivo MRI and histopathology detect novel iron-rich cortical inflammation in frontotemporal lobar degeneration with tau versus TDP-43 pathology. *NeuroImage: Clinical*, *33*, 102913. https://doi.org/10.1016/j.nicl.2021.102913

Tisdall, M. D., Ohm, D. T., Lobrovich, R., Das, S. R., Mizsei, G., Prabhakaran, K., Ittyerah, R., Lim, S., McMillan, C. T., Wolk, D. A., Gee, J., Trojanowski, J. Q., Lee, E. B., Detre, J. A., Yushkevich, P., Grossman, M., & Irwin, D. J. (2021). Joint ex vivo MRI and histology detect iron-rich cortical gliosis in tau and TDP-43 proteinopathies. *bioRxiv*. https://doi.org/10.1101/2021.04.14.439064

Toledo, J. B., Van Deerlin, V. M., Lee, E. B., Suh, E., Baek, Y., Robinson, J. L., Xie, S. X., McBride, J., Wood, E. M., Schuck, T., Irwin, D. J., Gross, R. G., Hurtig, H., McCluskey, L., Elman, L., Karlawish, J., Schellenberg, G., Chen-Plotkin, A., Wolk, D., … Trojanowski, J. Q. (2014). A platform for discovery: The university of Pennsylvania integrated neurodegenerative disease biobank. *Alzheimer's & Dementia*, *10*, 477–484. https://doi.org/10.1016/j.jalz.2013.06.003

Ushizima, D., Chen, Y., Alegro, M., Ovando, D., Eser, R., Lee, W., Poon, K., Shankar, A., Kantamneni, N., Satrawada, S., Amaro Junior, E., Heinsen, H., Tosun, D., & Grinberg, L. T. (2022). Deep learning for Alzheimer's disease: Mapping large-scale histological tau protein for neuroimaging biomarker validation. *NeuroImage*, *248*, 118790. https://doi.org/10.1016/j.neuroimage.2021.118790

van der Velpen, I. F., Vlasov, V., Evans, T. E., Ikram, M. K., Gutman, B. A., Roshchupkin, G. V., Adams, H. H., Vernooij, M. W., & Ikram, M. A. (2023). Subcortical brain structures and the risk of dementia in the Rotterdam study. *Alzheimer's & Dementia*, *19*, 646–657. https://doi.org/10.1002/alz.12690

van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L. D., & Iwatsubo, T. (2022). Lecanemab in early Alzheimer's disease. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa2212948

Vega, A. R., Chkheidze, R., Jarmale, V., Shang, P., Foong, C., Diamond, M. I., White, C. L., & Rajaram, S. (2021). Deep learning reveals disease-specific signatures of white matter pathology in tauopathies. *Acta Neuropathologica Communications*, *9*, 1–13. https://doi.org/10.1186/s40478-021-01271-x

Whitwell, J. L., Graff-Radford, J., Tosakulwong, N., Weigand, S. D., Machulda, M. M., Senjem, M. L., Spychalla, A. J., Vemuri, P., Jones, D. T., Drubach, D. A., Knopman, D. S., Boeve, B. F., Ertekin-Taner, N., Petersen, R. C., Lowe, V. J., Jack, C. R., Jr, & Josephs, K. A. (2018). Imaging correlations of tau, amyloid, metabolism, and atrophy in typical and atypical Alzheimer's disease. *Alzheimer's & Dementia*, *14*, 1005–1014. https://doi.org/10.1016/j.jalz.2018.02.020

Wisse, L., Adler, D., Ittyerah, R., Pluta, J., Robinson, J., Schuck, T., Trojanowski, J., Grossman, M., Detre, J.,

Elliott, M., Toledo, J. B., Liu, W., Pickup, S., Das, S. R., Wolk, D. A., & Yushkevich, P. A. (2017). Comparison of in vivo and ex vivo MRI of the human hippocampal formation in the same subjects. *Cerebral Cortex*, *27*, 5185–5196. https://doi.org/10.1093/cercor/bhw299

Wisse, L., Ravikumar, S., Ittyerah, R., Lim, S., Lane, J., Bedard, M., Xie, L., Das, S., Schuck, T., Grossman, M., Lee, E. B., Tisdall, M. D., Prabhakaran, K., Detre, J. A., Mizsei, G., Trojanowski, J. Q., Artacho-Pérula, E., de Iñiguez de Onzono Martin, M. M., Arroyo-Jiménez, M. M., … Yushkevich, P. A. (2021). Downstream effects of polypathology on neurodegeneration of medial temporal lobe subregions. *Acta Neuropathologica Communications*, *9*, 1–11. https://doi.org/10.1186/s40478-021-01225-3

Wisse, L., Ravikumar, S., Ittyerah, R., Lim, S. A., Lane, J., Lavery, M., Xie, L., Robinson, J. L., Schuck, T., Grossman, M., Lee, E. B., Tisdall, D. M., Prabhakaran, K., Detre, J. A., Das, S. R., Mizsei, G., Artacho-Perula, E., de Onzono Martin, M. M. I., Jimenez, M. M. A., Munoz, M., … Yushkevich, P. A. (2020). High-resolution postmortem MRI reveals tdp-43 association with medial temporal lobe subregional atrophy: Biomarkers: Leveraging postmortem collections to validate neuroimaging. *Alzheimer's & Dementia*, *16*, e045744. https://doi.org/10.1002/alz.045744

Xia, C., Makaretz, S. J., Caso, C., McGinnis, S., Gomperts, S. N., Sepulcre, J., Gomez-Isla, T., Hyman, B. T., Schultz, A., Vasdev, N., Johnson, K. A., & Dickerson, B. C. (2017). Association of in vivo [18F]AV-1451 tau PET imaging results with cortical atrophy and symptoms in typical and atypical Alzheimer disease. *JAMA Neurology*, *74*, 427–436. https://doi.org/10.1001/jamaneurol.2016.5755

Xie, L., Das, S. R., Wisse, L. E., Ittyerah, R., Yushkevich, P. A., Wolk, D. A., & Alzheimer's Disease Neuroimaging Initiative (2018). Early tau burden correlates with higher rate of atrophy in transentorhinal cortex. *Journal of Alzheimer's Disease*, *62*, 85–92. https://doi.org/10.3233/jad-170945

Yushkevich, P. A., López, M. M., Martin, M. M. I. d. O., Ittyerah, R., Lim, S., Ravikumar, S., Bedard, M. L., Pickup, S., Liu, W., Wang, J., Hung, L. Y., Lasserve, J., Vergnet, N., Xie, L., Dong, M., Cui, S., McCollum, L., Robinson, J. L., Schuck, T., … Insausti, R. (2021). Three-dimensional mapping of neurofibrillary tangle burden in the human medial temporal lobe. *Brain*, *144*, 2784–2797. https://doi.org/10.1093/brain/awab262

Yushkevich, P. A., Pashchinskiy, A., Oguz, I., Mohan, S., Schmitt, J. E., Stein, J. M., Zukić, D., Vicory, J., McCormick, M., Yushkevich, N., Schwartz, N., Gao, Y., & Gerig, G. (2019). User-guided segmentation of multi-modality medical imaging datasets with itk-snap. *Neuroinformatics*, *17*, 83–102. https://doi.org/10.1007/s12021-018-9385-x

Zeng, X., Puonti, O., Sayeed, A., Herisse, R., Mora, J., Evancic, K., Varadarajan, D., Balbastre, Y., Costantini, I., Scardigli, M., Ramazzotti, J., DiMeo, D., Mazzamuto, G., Pesce, L., Brady, N., Cheli, F., Pavone, F. S., Hof, P. R., Frost, R., … Fischl, B. (2023). Segmentation of supragranular and infragranular layers in ultra-high resolution 7t ex vivo MRI of the human cerebral cortex. *bioRxiv*, 2023–12. https://doi.org/10.1101/2023.12.06.570416

Zhu, W., Huang, Y., Zeng, L., Chen, X., Liu, Y., Qian, Z., Du, N., Fan, W., & Xie, X. (2019). Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical Physics*, *46*, 576–589. https://doi.org/10.1002/mp.13300