



Cortical Thickness Estimation: A Comparison of FreeSurfer and Three Voxel-Based Methods in a Test–Retest Analysis and a Clinical Application

Juan Velázquez¹ · Julieta Mateos² · Erick H. Pasaye³ · Fernando A. Barrios³ · Jorge A. Marquez-Flores¹

Received: 22 October 2020 / Accepted: 10 May 2021 / Published online: 18 May 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The cortical thickness has been used as a biomarker to assess different cerebral conditions and to detect alterations in the cortical mantle. In this work, we compare methods from the FreeSurfer software, the Computational Anatomy Toolbox (CAT12), a Laplacian approach and a new method here proposed, based on the Euclidean Distance Transform (EDT), and its corresponding computational phantom designed to validate the calculation algorithm. At region of interest (ROI) level, within- and inter-method comparisons were carried out with a test–retest analysis, in a subset comprising 21 healthy subjects taken from the Multi-Modal MRI Reproducibility Resource (MMRR) dataset. From the Minimal Interval Resonance Imaging in Alzheimer’s Disease (MIRIAD) data, classification methods were compared in their performance to detect cortical thickness differences between 23 healthy controls (HC) and 45 subjects with Alzheimer’s disease (AD). The validation of the proposed EDT-based method showed a more accurate and precise distance measurement as voxel resolution increased. For the within-method comparisons, mean test–retest measures (percentages differences/intraclass correlation/Pearson correlation) were similar for FreeSurfer (1.80%/0.90/0.95), CAT12 (1.91%/0.83/0.91), Laplacian (1.27%/0.89/0.95) and EDT (2.20%/0.88/0.94). Inter-method correlations showed moderate to strong values ($R > 0.77$) and, in the AD comparison study, all methods were able to detect cortical alterations between groups. Surface- and voxel-based methods have advantages and drawbacks regarding computational demands and measurement precision, while thickness definition was mainly associated to the cortical thickness absolute differences among methods. However, for each method, measurements were reliable, followed similar trends along the cortex and allowed detection of cortical atrophies between HC and patients with AD.

Keywords FreeSurfer · CAT12 · EDT · Cortical thickness · Alzheimer · S Disease · Test–retest analysis

Introduction

Cortical thickness has been extensively studied because it is one of the most sensitive biomarkers used to assess different cerebral conditions, ranging from changes under normal development, such as aging (Salat et al. 2004; Hutton et al. 2009), to neurological disorders (Rosas et al. 2002; Cardinale et al. 2014; Clarkson et al. 2011). The cerebral cortex follows a highly convoluted gyrification pattern with gyri and sulci across the entire structure. It is delimited by the white matter (WM)/gray matter (GM) surface at the interior, and by the pial surface at the outermost part of the brain. Given this geometry, cortical thickness can be measured only if the WM/GM and pial surfaces are well determined (Fischl and Dale 2000). Methodologies employed to quantify the cortical thickness from MRI data have been classified as surface-based and voxel-based (Seiger et al. 2018).

Handling Editor: Irena Rektorova.

✉ Fernando A. Barrios
fbarrios@unam.mx

✉ Jorge A. Marquez-Flores
jorge.marquez@icat.unam.mx

¹ Universidad Nacional Autónoma de México, Instituto de Ciencias Aplicadas Y Tecnología, Circuito Exterior S/N, Ciudad Universitaria, 04510, Coyoacán, Mexico City, México

² Graduate Program in Computer Science and Engineering, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510, Coyoacán, Ciudad de México, México

³ Universidad Nacional Autónoma de México, Instituto de Neurobiología, Boulevard Juriquilla 3001, 76230, Querétaro, Querétaro, México

A surface-based approach requires a mesh model to render the cortical surfaces, whereas a voxel-based approach works directly on the original grid of voxels, making this methodology less computationally expensive (Clarkson et al. 2011) and independent from a fitted surface model.

Regarding a surface-based approach, FreeSurfer is a popular toolbox to perform cortical thickness measurements. Despite its well-known computational cost to ensure an accurate cortical topology (Clarkson et al. 2011; Fischl et al. 2001), FreeSurfer has been widely applied to in-vivo datasets, post-mortem (Rosas et al. 2002) and ex-vivo (Cardinale et al. 2014), making this software reliable, robust and accurate. However, there are some reasons an in-vivo cortical thickness gold standard has not been settled. Histology measurements do not provide reliable results due to structural changes in the cortex (e.g. shrinking) related to the fixation of the post-mortem brain (Lüsebrink et al. 2013). Further, ex-vivo measurements have been made only region-specific and cannot be considered valid for the whole cerebral cortex (Seiger et al. 2018). Finally, there is no mathematical definition on how to measure the thickness of highly curved structures (Lüsebrink et al. 2013).

In view of more efficient methods and given the lack of a real gold standard, different voxel-based approaches have been proposed. Nevertheless, the most important limitation has been their decreased accuracy calculations due to partial volume effects affecting convoluted structures, which can lead to a less robust segmentation of the tissues of interest (Clarkson et al. 2011; Hutton et al. 2008). Based on this framework, Jones et al. (2000) proposed the first procedures by solving Laplace's equation and computing streamlines between the WM/GM and GM/cerebrospinal fluid (CSF) interfaces, which serve as the trajectories to estimate the cortical thickness. Improvements to provide better cortical boundaries consist in stacking layers with one voxel of thickness around the WM, identifying sulcal regions by expecting a certain thickness value (Hutton et al. 2008), and others rely on the skeletonization of the CSF to better delineate the GM/CSF boundary (Hutton et al. 2009). Another voxel-based approach is distributed in the Computational Anatomy Toolbox (CAT: <http://www.neuro.uni-jena.de/cat/>) for the Statistical Parametric Mapping (SPM: <http://www.fil.ion.ucl.ac.uk/spm/>) software. Given the WM/GM and GM/CSF segmentations, a Projection Based Thickness (PBT) method (Seiger et al. 2018) is employed, which consists in estimating the distance from the WM/GM boundary to project the local maxima to other GM voxels by taking into account information from the neighboring voxels and from blurred sulci to generate correct cortical thickness maps (Dahnke et al. 2013; Righart et al. 2017; Seiger et al. 2018).

Previous studies comparing the cortical thickness obtained by different methodologies have included FreeSurfer and Laplace's method (Clarkson et al. 2011; Li et al.

2015), and more recently, between FreeSurfer and CAT12 (Righart et al. 2017; Seiger et al. 2018). In this work, within- and inter-method comparisons were carried out using FreeSurfer, CAT12, the Laplacian thickness and a Euclidean Distance Transform (EDT). We have already proposed and applied EDT-based methods in previous shape-analysis works; for example, a morphological average uses the EDT of anatomical shapes to extract a representative model for craniofacial morphometry (Márquez et al. 2005). Also, we introduced an EDT-based method to measure the width of cortical sulci in brains of patients with Alzheimer's disease (AD) and controls (Mateos et al. 2020). It was validated with a mathematical exact analysis and a corresponding voxelized computational phantom, modeling width variations and the effect of discretization, voxel resolution and shape orientation. The phantom in the present work follows a similar approach where the effect of voxel resolution was assessed on a set of concentric spheres (Das et al. 2009), and a distribution of analytically determined distances between eccentric spheres were compared to those of the EDT algorithm. In real brain images, region of interest (ROI)-wise cortical thickness measurements, applying each method, were performed on a test–retest dataset to study the measurement reliability. Finally, as a clinical application, the detection of brain atrophy between healthy controls (HC) and subjects with AD was assessed.

Methods

Subjects and Data Acquisition

Multi-Modal MRI Reproducibility Resource (MMRR) Dataset

Data for the test–retest analysis were taken from the freely available MMRR dataset (Landman et al. 2011). We analyzed T1-weighted images of the complete database (21 subjects), comprising 10 females and 11 males (31.8 ± 9.5 years [mean age \pm standard deviation]) with no history of neurological conditions. The volunteers were scanned and rescanned with a short break between sessions (two in the same day); a total of 42 sessions were completed in a two-week interval. See detailed information in Landman et al. (2011).

Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) Dataset

As a potential clinical application, the MIRIAD dataset (Malone et al. 2013) was included comprising a total of 46 AD subjects and 23 HC. One female with AD was excluded as the data of the session (baseline) we analyzed were not available. Therefore, for the analysis, 45 subjects (26 females) diagnosed with mild-moderate probable AD

(Mini-Mental State Examination < 27), and 23 healthy controls (11 females) were considered. The AD and HC subjects were age-matched at 69.1 ± 7.1 and 69.7 ± 7.2 years, respectively. See detailed information in Malone et al. (2013).

Cortical Thickness Estimation with Current Software

FreeSurfer

Under FreeSurfer software version 6.0 (<http://surfer.nmr.mgh.harvard.edu/>), all subjects were processed using the command *recon-all* with default parameters. The pipeline consists of several stages. First, with a Talairach transformation (Talairach and Tournoux 1988), the original volume was registered to a standard space and the white matter points were labeled based on their location and intensity, followed by an intensity normalization procedure (Dale et al. 1999; Fischl et al. 2004). The skull was then stripped (Ségonne et al. 2004) and the hemispheres were separated based on the expected location of the corpus callosum and pons, while the cerebellum and brain stem were removed. Following the intensity gradients of the GM/WM boundary (white surface), topological correction to create accurate and topologically correct surfaces was achieved with a subsequent deformation to follow the intensity gradients of the CSF/GM boundary (pial surface) (Fischl et al. 2001). Finally, the cortical thickness was estimated as the average of two distances: the distance from a point on the white surface to the closest point on the pial surface and the distance from that point back to the nearest point on the white surface (Clarkson et al. 2011; Rosas et al. 2002). No manual editing was performed in any case, but each output was visually inspected. For the MMRR dataset, obvious issues regarding skull stripping, intensity normalization and tissue segmentation were not visible. For the MIRIAD dataset, no skull stripping failures were present although there were four AD subjects and one HC with soft issues regarding intensity normalization and tissue classification. However, we considered this is of no major concern as a comparison to other cortical thickness measurement methods, using the same FreeSurfer segmentation, was carried out.

Computational Anatomy Toolbox (CAT12)

As an alternative and relatively new software to perform cortical thickness estimates, all subjects were processed with CAT12 version r1430 (<http://www.neuro.uni-jena.de/cat/>) under SPM12 version 7487 (<http://www.fil.ion.ucl.ac.uk/spm/>) using Matlab (R2017b). Before running the pipeline, we set the origin in each cerebral volume at the anterior commissure. Afterwards, we segmented the original volume specifying the surface and thickness estimation for ROI analysis in the options. To calculate the cortical

thickness, tissue segmentation was used to estimate the WM distance and project the local maxima to other GM voxels by using a neighbor relationship described by the WM distance (Dahnke et al. 2013). The reconstruction process included topology correction relying on spherical harmonics (Yotter et al. 2011a), spherical mapping to reparameterize the surface mesh into a common coordinate system (Yotter et al. 2011b) and spherical registration (Ashburner 2007). It is important to mention that CAT12 is a stand-alone segmentation pipeline of structural brain MR images, as an extension to SPM12, where a range of morphometric methods offered are optional, including cortical thickness estimation. In this study, CAT12 has been used with default settings to allow comparisons against a truly alternative method taking advantage of a fast and a fully automated pipeline.

Laplacian

Supplied by ANTs (Advanced Normalization Tools, <http://stnava.github.io/ANTs/>) version 2.3.1, we implemented the Laplacian thickness method as previously described (Jones et al. 2000). The input for the command was the segmented GM and WM provided by FreeSurfer after the cortical reconstruction process. We used these volumes for a more direct comparison, reducing the software-dependent segmentation procedure. The thickness maps were capped at 5 mm and smoothed using a Gaussian filter of sigma equal to 1 mm.

Proposed EDT for Cortical Thickness Estimation

The input for this method was the segmented WM and GM taken from FreeSurfer. The EDT is a transformation D carried out on the WM to obtain a distance map d as a function of each voxel in the volume of interest. The voxel value is given by the Euclidean distance from the coordinates of that voxel to the closest point on the boundary ∂WM as expressed by:

$$[D(WM)](p) = d(p, \partial WM) \triangleq \min_{q \in \partial WM} |p - q|, p \in \mathbb{Z}^3$$

The next step was to crop and delimit the EDT to the region ranging all along the GM only. This was achieved by modulating the original distance map with the GM volume. The outcome was a cortical thickness map capped at 5 mm and smoothed using a Gaussian filter of 1 mm.

Validation of the EDT

Eccentric Spheres

Two eccentric spheres were designed to model a source of varying distances between an inner sphere ($r = 5$ mm),

centered at C_i , and an outer sphere ($R=8$ mm) centered at C_o . The relative displacement of the inner to the outer sphere was of 1 and 2 mm in the x and y direction, respectively. This phantom simulates thickness variations of the GM and does not model the shape of the brain or other features. Also, this simple geometry allowed us to obtain an analytical ground truth to compare with the discrete EDT measurement. Thus, we carried out the analytical calculation as follows (see Fig. 1). First, the general equation of the sphere was used to compute a unit normal vector (\hat{n}_{inner}) at every point of the inner sphere (p_i). Then, using the parameterized equation of a line in the space, we projected the normal vector until the intersection with the outer sphere (p_o) was found, and this value was recorded as d_1 . The coordinates at the intersection (p_o) were taken to compute a unit normal vector to the outer sphere (\hat{n}_{outer}), which was projected back to the interior part of the geometry until the intersection with the inner sphere (p_i) was reached and its distance was recorded as d_2 . Finally, the thickness at (p_o) was measured as the average between the distances d_1 and d_2 .

Concentric Spheres

Centered at the same coordinates, inner and outer spherical surfaces of radius $r=5$ and $R=8$ mm, respectively, were modeled so that a theoretical exact thickness of 3 mm was expected at every point on the outer surface. To study the effect of spatial resolution and discretization, the same

geometries were voxelized at isotropic voxel sizes of 0.1, 0.5 and 1.0 mm (see Fig. 2). The only difference with respect to the theoretical value is produced by two factors: the discretization error and machine precision, giving the latter a much lower error than the former.

Statistical Analysis

Cortical thickness data were extracted using regular commands in FreeSurfer and CAT12, while a script was devised for the Laplacian and the EDT methods. ROI-wise mean and standard deviation were obtained over the 34 regions of the Desikan-Killiany atlas (Desikan et al. 2006). For within-method comparisons, percent difference, paired sample t-test, intraclass (ICC) and Pearson (R) correlation coefficients between scans were calculated to assess the reliability measurement. For inter-method comparisons, the first scan of the MMRR dataset was taken and the Pearson correlation coefficient was calculated to measure the agreement on a between-methods basis. Finally, to assess the method capability for detecting group differences (HC, AD) in cortical atrophy, effect sizes (Cohen’s d) and Welch’s t-tests for unequal variances were computed. Significance was defined at $p < 0.05$ and t-tests were corrected for multiple comparisons with a False Discovery Rate (FDR correction) of 0.05.

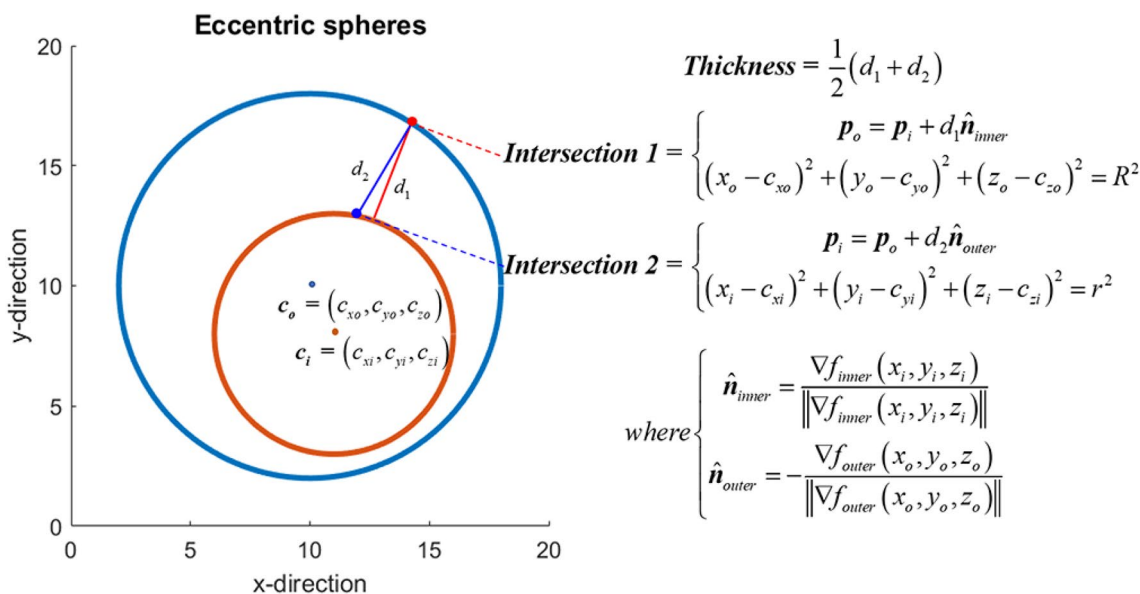


Fig. 1 Slice of a set of eccentric spherical surfaces to analytically determine the thickness point wise. First, a unit normal vector (\hat{n}_{inner}) to every point (p_i) of the inner sphere, centered at c_i , is computed. The normal vector is projected until the intersection with the outer sphere is found (Intersection 1), and the distance is recorded as d_1 .

Then, Intersection 1 (point p_o on the outer sphere, centered at c_o) is used to compute and project a unit normal vector (\hat{n}_{outer}) back to the inner sphere (p_i) to find Intersection 2. This distance is recorded as d_2 . Finally, the exact thickness at p_o is obtained as the average of d_1 and d_2

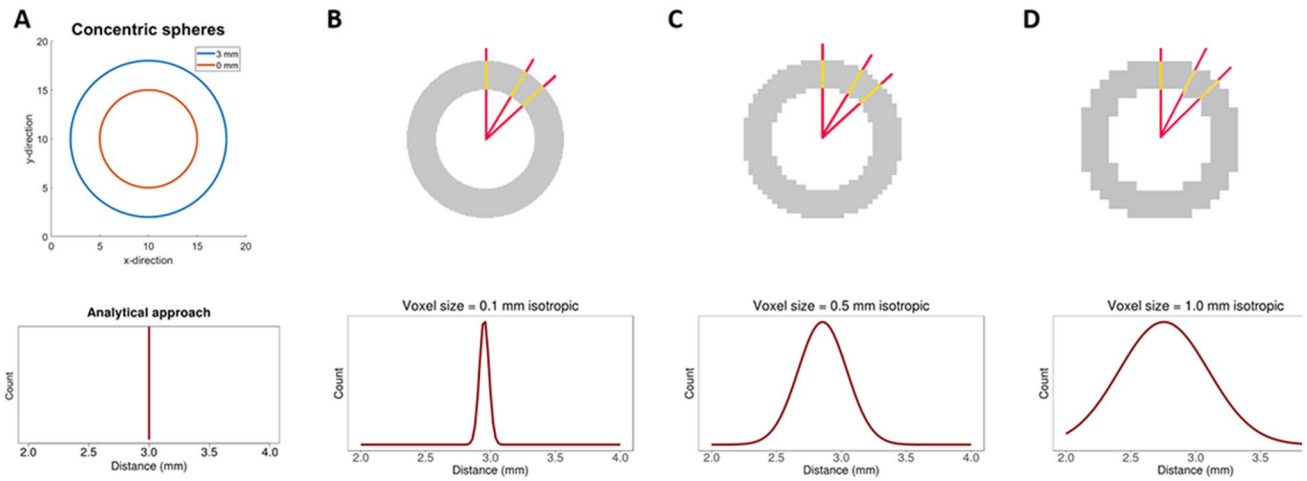


Fig. 2 Accuracy and precision of the EDT represented by a slice of the concentric spheres (top row) and its corresponding histogram (bottom row) of the (A) analytically measured thickness and the voxelized approach at voxel sizes of (B) 0.1 mm, (C) 0.5 mm and (D) 1.0 mm isotropic

Results

Evaluation of the EDT used on synthetic images at different spatial resolution showed the following: For the concentric spheres (mathematical models), the analytically measured thickness was 3 mm at every point of the outer sphere (Fig. 2A). In the voxelized geometries (computational models), a narrow distribution of thicknesses

was still near 3 mm at a voxel size of 0.1 mm isotropic (Fig. 2B) with an average of 2.95 ± 0.04 mm. However, the histogram broadened as the distances was measured on images of lower spatial resolution of 0.5 mm (average: 2.86 ± 0.19 mm) and 1.0 mm (average: 2.75 ± 0.34 mm), resulting in less precise and accurate estimations of the thickness (see Fig. 2).

Regarding the geometry of eccentric spheres, the data distribution of thicknesses was described using a *violin plot*

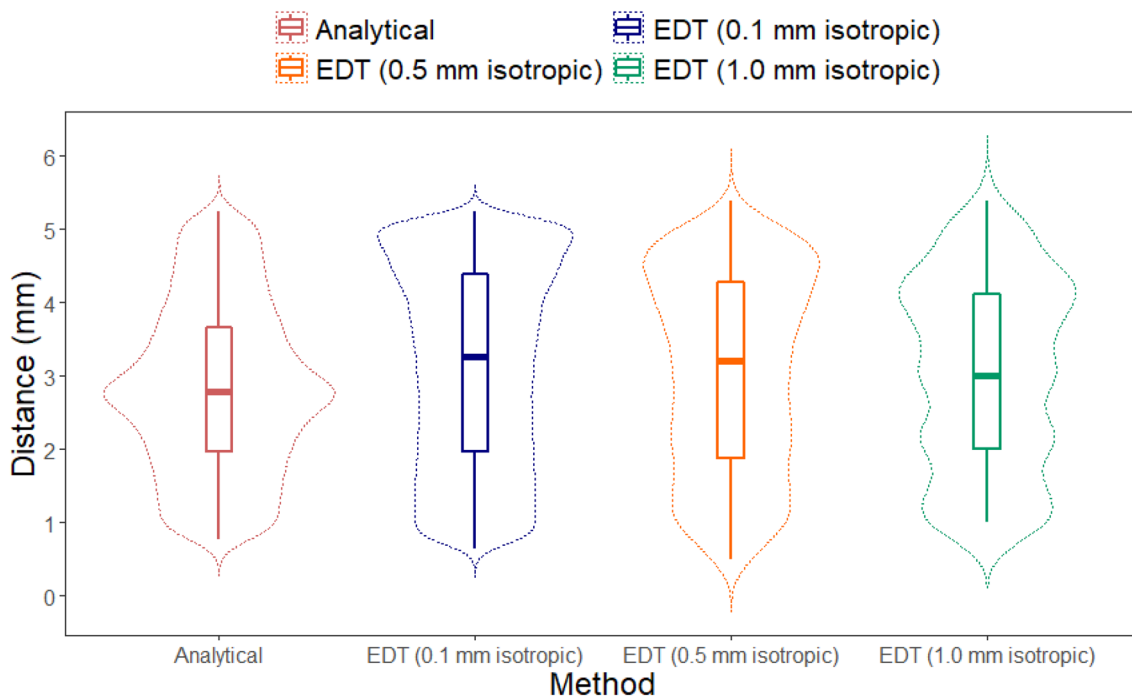


Fig. 3 Analytical and EDT thickness distribution between eccentric spheres. Analytically, the thickness was determined as detailed in Fig. 1. On the other hand, the EDT was applied at a different spatial resolution. A similar Coefficient of Variation, CV (standard deviation/mean), is shown

(see Fig. 3). In the analytical model, a distribution more densely concentrated around its mean (2.84 ± 0.14 mm) was obtained. The violin shape was similar between the EDT at 0.1- and 0.5-mm voxel with mean values of 3.15 ± 1.35 and 3.06 ± 1.34 mm, respectively. The data distribution was more concentrated at higher values which we assumed was due to a rounding-up voxel effect. Conversely, at an isotropic voxel size of 1.0 mm (average: 3.04 ± 1.28 mm), a poorer sphere segmentation was given so that the distance map resulted in zero-valued voxels with short distances, and thus, in a lower mean value. Despite differences in the data distribution, the Coefficient of Variation, CV (ratio of the standard deviation to the mean), showed robust results among measurements with values between 0.40 and 0.44.

For the test–retest analysis, in general, most of the ROIs obtained a percent difference between 1 and 3%, while the ICC and Pearson correlation coefficients were greater than 0.85 and 0.91, respectively. The extreme case was in the entorhinal and parahippocampal region measured with CAT12, delivering the highest percent differences and poorer correlations between scans. Another complicated region with low correlation values, regarding the voxel-based approaches only, was the temporal pole, transverse temporal and the insula. Detailed information of these measurements is displayed in Table 1. A paired sample t-test of equal variances was performed for each ROI. Previously, the assumption of normality and equal variances was tested (and fulfilled the criteria) running the Shapiro–Wilk’s test and Bartlett’s test, respectively. Despite some low ICCs and Pearson correlation coefficients shown in Table 1, the t-test suggested a non-significant mean difference between scans in every ROI (FDR correction) for surface- and voxel-based approaches. To complement this description, the cortical thickness (measured with four methods) on the scan-rescan images is depicted in Fig. 4. The first trend was the significant higher values obtained with the Laplacian method with respect to FreeSurfer and the EDT for all 34 ROIs. Likewise, CAT12 provided higher values compared to FreeSurfer and the EDT in 29 and 26 out of 34 ROIs, respectively. Pronounced differences were detected at the entorhinal, parahippocampal and the posterior cingulate.

For inter-method comparisons, Pearson correlation coefficients between methods are shown in Table 2. Comparison between FreeSurfer and CAT12 resulted in the highest correlation coefficient, followed by the Laplacian approach against the EDT and FreeSurfer against the EDT. When CAT12 was compared to the voxel-based methods, moderate correlation coefficients were obtained against the EDT ($R=0.79$) and Laplacian ($R=0.77$) approaches.

Results concerning the MIRIAD cohort are displayed in Fig. 5, where mean cortical thickness was plotted against the ROIs for each method. Again, for all 34 ROIs, significant higher values were obtained using the Laplace method with

a greater difference when compared against FreeSurfer and the EDT. Nevertheless, a very similar trend among methods along the range of regions under study was observed.

Regarding the within-method comparison, an evident decrease in cortical thickness in diseased subjects was shown, although the opposite was revealed in a few ROIs. A thorough comparison, based on Cohen’s *d* effect size and Welch’s t-test (unequal variances), indicated the similar significant differences (FDR correction) between HC and AD as follows. Out of 34 ROIs, a significant group difference was found in 21 ROIs when measuring with FreeSurfer. Likewise, significant differences were found in 24 ROIs and 18 ROIs for the Laplacian and EDT methods, respectively. In terms of detecting group differences, CAT12 differed the most with respect to other methods, but most Cohen’s *d* values were the highest followed by the Laplacian and EDT methods, and FreeSurfer. For all methods, pronounced differences ($d > 1.00$) were found mainly in temporal brain regions. Detailed information is shown in Table 3.

Discussion

Due to the lack of an in-vivo gold standard of cortical thickness, an underexploited mathematical model and its corresponding computational phantom were designed to validate the proposed EDT. Results reported here have shown that the accuracy and precision of the EDT method, compared to a mathematical standard, vary as spatial resolution decreases. Apparent inconsistencies between high- and low-resolution data, for arbitrary geometries, were strongly attributed to partial volume effects. To alleviate this issue mainly affecting the voxel-based methods, it might be worth spending more time on image acquisition to enhance voxel resolution and obtain more precise measurements.

For cortical thickness estimations, comparisons of three currently used methods (FreeSurfer, CAT12 and the Laplacian approach) were conducted, adding the EDT-based measurement to the volume-based methodology list. For all within-method comparisons in the test–retest analysis, non-significant cortical thickness differences were found, and three test–retest measures were devised suggesting a strong correlation between scans in almost all ROIs under investigation. Inter-methods comparisons, taking the measurements of the first scans of the MMRR dataset, showed from moderate to strong significant correlations for all observations, performing slightly better when the surface-based method was involved. In more detail, the highest correlation coefficient was observed for FreeSurfer against CAT12. This might be explained as these are pipelines with specialized stages to better estimate the cortical thickness including topology correction, in the case of FreeSurfer and adaptation of blurred sulci and gyri in

Table 1 Test–retest cortical thickness measurement using the Desikan–Killiany atlas

ROI	FS6			CAT12			Laplacian			EDT		
	Δ (%)	ICC	R	Δ (%)	ICC	R	Δ (%)	ICC	R	Δ (%)	ICC	R
1-Banks superior temporal sulcus-(B)	1.06	0.95	0.98	1.17	0.88	0.94	1.43	0.93	0.98	1.67	0.94	0.97
2-Caudal anterior cingulate-(CACg)	1.71	0.97	0.98	1.58	0.97	0.98	1.29	0.90	0.95	1.45	0.96	0.98
3-Caudal middle frontal-(CMF)	2.38	0.78	0.89	2.20	0.78	0.91	1.31	0.85	0.92	2.14	0.89	0.94
4-Cuneus-(Cu)	1.52	0.96	0.98	1.41	0.88	0.94	1.97	0.91	0.96	2.81	0.89	0.95
5-Entorhinal-(En)	4.18	0.83	0.92	6.96	0.35	0.58	1.08	0.91	0.95	4.63	0.77	0.88
6-Fusiform-(Fu)	1.19	0.93	0.97	1.71	0.90	0.95	0.77	0.86	0.95	1.62	0.84	0.93
7-Inferior parietal-(IP)	2.06	0.81	0.92	1.29	0.90	0.95	1.57	0.80	0.89	2.15	0.85	0.93
8-Inferior temporal-(IT)	1.14	0.94	0.97	1.47	0.86	0.93	0.86	0.87	0.96	1.73	0.89	0.96
9-Isthmus cingulate-(IstCg)	1.80	0.90	0.95	1.62	0.96	0.98	1.18	0.85	0.92	2.28	0.86	0.93
10-Lateral occipital-(LO)	1.67	0.92	0.97	0.93	0.96	0.98	1.80	0.86	0.94	2.99	0.86	0.94
11-Lateral orbital frontal-(LORF)	1.57	0.88	0.94	1.17	0.93	0.97	1.01	0.86	0.93	1.94	0.80	0.91
12-Lingual-(Lg)	1.56	0.94	0.97	1.67	0.75	0.88	1.54	0.90	0.96	2.63	0.91	0.96
13-Medial orbital frontal-(MORF)	1.95	0.91	0.96	1.98	0.83	0.94	1.23	0.86	0.94	2.66	0.80	0.91
14-Middle temporal-(MT)	1.13	0.92	0.96	1.55	0.59	0.77	0.73	0.92	0.96	1.49	0.87	0.94
15-Parahippocampal-(PaH)	1.73	0.95	0.97	4.00	0.54	0.73	1.17	0.92	0.96	2.34	0.92	0.96
16-Paracentral-(PaC)	1.86	0.91	0.96	1.57	0.81	0.91	1.77	0.92	0.96	2.02	0.94	0.97
17-Pars opercularis-(Op)	1.60	0.88	0.94	1.35	0.90	0.95	0.91	0.94	0.97	1.75	0.92	0.96
18-Pars orbitalis-(Or)	1.76	0.92	0.96	1.32	0.95	0.97	1.06	0.87	0.94	2.40	0.90	0.95
19-Pars triangularis-(Tr)	1.47	0.94	0.97	1.42	0.94	0.97	1.18	0.94	0.97	2.37	0.91	0.96
20-Pericalcarine-(PerCa)	2.84	0.85	0.92	2.43	0.66	0.81	3.00	0.91	0.96	4.10	0.84	0.93
21-Postcentral-(PoC)	0.88	0.97	0.98	1.14	0.96	0.99	1.19	0.95	0.97	1.42	0.95	0.97
22-Posterior cingulate-(PoCg)	1.45	0.93	0.96	1.21	0.92	0.96	1.03	0.90	0.95	1.25	0.92	0.97
23-Precentral-(PreC)	1.47	0.90	0.95	1.35	0.91	0.96	1.30	0.96	0.98	1.80	0.95	0.97
24-Precuneus-(PreCu)	1.68	0.90	0.95	1.01	0.93	0.96	1.42	0.91	0.95	1.70	0.90	0.95
25-Rostral anterior cingulate-RoACg)	2.08	0.91	0.95	1.70	0.94	0.97	1.18	0.89	0.94	1.84	0.88	0.93
26-Rostral middle frontal-(RoMF)	1.51	0.85	0.93	1.57	0.88	0.95	1.43	0.79	0.91	1.97	0.87	0.93
27-Superior frontal-(SF)	2.05	0.88	0.94	1.96	0.82	0.92	1.07	0.84	0.92	1.62	0.94	0.97
28-Superior parietal-(SP)	2.03	0.88	0.94	1.67	0.92	0.96	1.78	0.87	0.93	2.05	0.91	0.95
29-Superior temporal-(ST)	1.07	0.98	0.99	1.93	0.80	0.89	0.59	0.97	0.98	1.10	0.97	0.99
30-Supramarginal-(SM)	1.45	0.88	0.95	1.11	0.93	0.97	1.02	0.89	0.95	1.27	0.94	0.97
31-Frontal pole-(FPol)	2.45	0.89	0.95	3.30	0.81	0.92	1.15	0.90	0.95	2.96	0.88	0.93
32-Temporal pole-(TPol)	2.55	0.73	0.87	3.59	0.45	0.66	0.39	0.77	0.88	2.94	0.62	0.81
33-Transverse temporal-(TrT)	2.39	0.92	0.96	1.83	0.94	0.97	1.58	0.95	0.98	2.59	0.96	0.98
34-Insula-(Ins)	1.85	0.86	0.93	2.81	0.73	0.85	1.34	0.75	0.86	2.99	0.79	0.89
Mean	1.80	0.90	0.95	1.91	0.83	0.91	1.27	0.89	0.95	2.20	0.88	0.94

The significance is $P < 0.05$, corrected for multiple comparisons

For each region of interest (ROI) of the Desikan–Killiany atlas, Δ (%) mean percentage difference (test–retest reliability), intraclass (ICC) and Pearson (R) correlation coefficients were determined between subjects' scans. Measurement methods included FreeSurfer (FS6), the Computational Anatomy Toolbox (CAT12), the Laplacian approach and the Euclidean Distance Transform (EDT) approach

the PBT mapping for CAT12. High correlations between FreeSurfer and the remaining voxel-based methods might be attributed to the use of the same segmented WM and GM outputted from FreeSurfer. However, there is a striking lower correlation between the voxel-based methods which we ascribed to the segmentation process. While the Laplacian and EDT techniques ($R = 0.90$) took the same

WM and GM volumes, CAT12 produced their own tissue segmentation.

Inter-method comparisons have shown that absolute values obtained for each method were not directly comparable, and caution should be taken when studying group differences and comparing among methods. On the other hand, inter-method disagreement in absolute values might also be attributed to systematic differences due to distinct

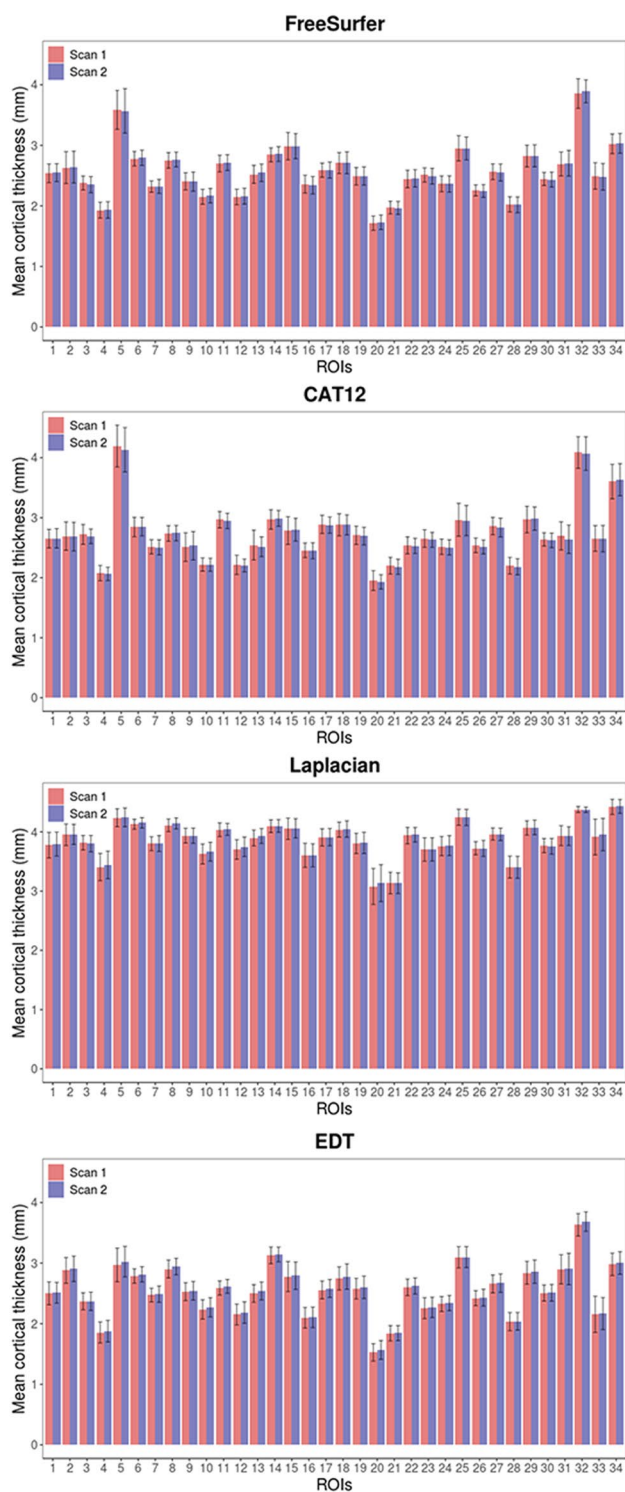


Fig. 4 Mean cortical thickness and standard deviation using FreeSurfer, CAT12, the Laplacian method and EDT in the MMRR dataset. For all methods, non-significant differences ($P < 0.05$, corrected for multiple comparisons) were found between the scan-rescan images for 34 ROIs of the Desikan-Killiany atlas. Numbers on ROIs-axis indicate a region described in Table 1

Table 2 Linear correlations to assess cortical thickness agreement between methods

	FS6	CAT12	Laplacian	EDT
FS6		0.92	0.85	0.89
CAT12			0.77	0.79
Laplacian				0.90
EDT				

Each value represents the Pearson correlation coefficient calculated between methods including FreeSurfer (FS6), the Computational Anatomy Toolbox (CAT12), the Laplacian approach and the Euclidean Distance Transform (EDT) approach. All correlations were significant ($P < 0.05$)

distance definitions, as has been pointed out in other works (Das et al. 2009; Seiger et al. 2018). The Laplacian approach differed the most in terms of the absolute values delivered. This may be explained as follows: in the Laplacian approach, the cortical thickness was estimated as a curved distance defined along computed streamlines in the GM (Hutton et al. 2008; Jones et al. 2000). Therefore, the measurement resemblance among other inter-method comparisons (FreeSurfer, CAT12 and EDT) was attributed to a straight-line definition since they share the essence of a Euclidean metric. However, the distance definition was still different; FreeSurfer computed the cortical thickness as an average nearest neighbor points (Rosas et al. 2002) of two fully reconstructed cortical surfaces, which in turn can be a source of discrepancy due to its fitting model nature that is not exempt from disregarding local, irregular variations of the cortex. Slight overestimations of CAT12 were also associated with the calculation algorithm based on a local maxima projection method adapting for blurred sulci and gyri (Righart et al. 2017; Seiger et al. 2018). Finally, the EDT distance definition was also different from the rest, as cortical thickness was estimated as the distance between closest corresponding points given by the EDT. Overall, despite these characteristics, most ROI cortical thickness patterns were similar among methods.

As a clinical application, the MIRIAD dataset was used to detect atrophies due to a neurodegenerative condition. In terms of significant ROI-wise group differences, effect sizes were slightly higher for CAT12. This may suggest an outperformance of CAT12 to discriminate AD from HC as was indicated in a previous study comparing against FreeSurfer (Seiger et al. 2018). As CAT12 is a fully automated pipeline with its own tissue classification, topological correction and cortical thickness mapping, it is difficult to identify the source of the claimed better performance. Significant group differences were more in agreement among FreeSurfer, Laplacian and EDT methods. Once again, it seemed that results were more alike when the same segmentation volumes were used within the calculation algorithm.

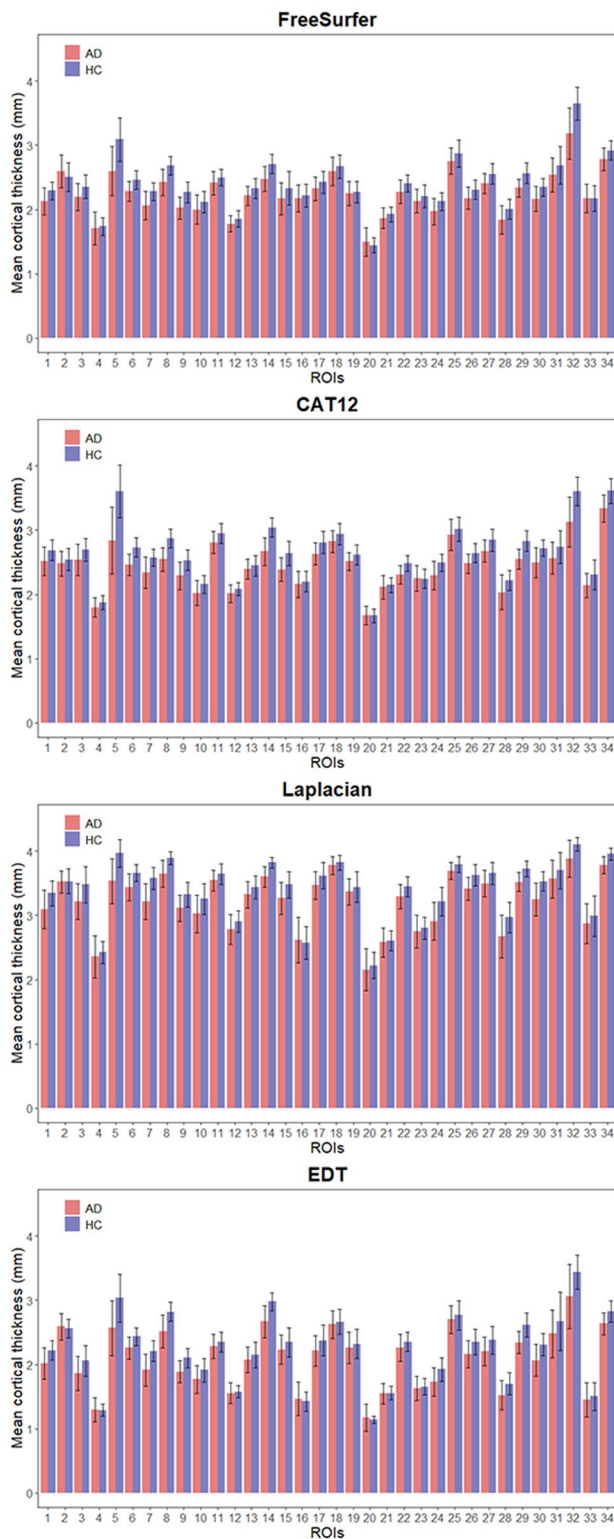


Fig. 5 Mean cortical thickness and standard deviation for FreeSurfer, CAT12, the Laplacian method and EDT in the MIRIAD dataset. Measurements included 34 ROIs of the Desikan-Killiany atlas comparing subjects with Alzheimer’s disease (AD: red) and healthy controls (HC: blue). Numbers on the ROI axis indicate a region described in Table 1

This, pattern repeated from the correlation analysis, might highlight the importance of segmentation when comparing results among techniques.

In sum, all methods performed similar for almost all ROIs with undesirable measures found in the entorhinal cortex, insula and temporal regions (parahippocampal, transverse temporal and temporal pole). These results were in line with published works where variable results were attributed to a poor volume segmentation and surface reconstruction of the temporal regions (Han et al. 2006), due to a very likely low contrast and susceptibility artifacts in the image acquisition (Lüsebrink et al. 2013). Another complicated structure is the insula (Blustajn et al. 2019; Li et al. 2015) and entorhinal (Li et al. 2015, Price et al. 2011) mainly due to their size and to challenges in detecting the boundaries to distinguish them from their surroundings. Because tissue segmentation appeared to be the main source of misleading results in highly convoluted regions, it could be suggested to optimize segmentation algorithms and acquisition protocols for specific brain areas. Despite these observations, efficient alternative methods attaining reliable estimates of cortical thickness measures can be valuable when computational resources are not enough or when shorter processing times of large datasets is a priority. FreeSurfer version 7 has been released and according to the FreeSurfer Release Notes (<https://surfer.nmr.mgh.harvard.edu/fswiki/ReleaseNotes>), the complete reconstruction pipeline is 20–25% faster than previous versions along with improved algorithms to process volumes at higher spatial resolution. Although these new features, voxel-based methods are still more efficient and they are also capable of managing data with increased resolution, overcoming one of the most important drawbacks of these kind of techniques.

Conclusions

Comparison results were limited to the use of FreeSurfer 6. As the latest stable release included a triangular mesh of the white surface of better quality, bias field correction using ANTS N4, among others, trends observed in this work may change mainly when comparing against CAT12, assuming that the Laplacian and EDT techniques take FreeSurfer segmentation as input. Another limitation was associated to CAT12 in terms of the comparison of cortical thickness metrics against the other methods, because CAT12 applied their own processing steps different to those of FreeSurfer.

We have introduced a voxel-based method using the EDT, validated by an analytical model and its corresponding computer phantom, and implemented to estimate the cortical thickness. A “beta” version of the EDT-based calculation cortical thickness in discrete brain datasets, written in Matlab will be available free, by written demand to the first

Table 3 Cortical thickness comparison between healthy controls and the Alzheimer diseased population, reporting the effect size (Cohen's d) using the Desikan-Killiany atlas

ROI	FS6 Cohen's d	CAT12 Cohen's d	Laplacian Cohen's d	EDT Cohen's d
1-B	0.84*	0.83*	0.92*	0.93*
2-CACg	− 0.38	0.33	0.04	− 0.16
3-CMF	0.78*	0.70*	0.92*	0.73*
4-Cu	0.14	0.50*	0.23	− 0.04
5-En	1.33*	1.57*	1.37*	1.13*
6-Fu	1.20*	1.69*	1.16*	1.08*
7-IP	1.09*	1.08*	1.46*	1.33*
8-IT	1.35*	1.89*	1.38*	1.32*
9-IstCg	1.44*	1.21*	1.05*	1.29*
10-LO	0.56*	0.77*	0.83*	0.68*
11-LOrF	0.51	0.81*	0.61*	0.38
12-Lg	0.57*	0.53*	0.57*	0.15
13-MOrF	0.74*	0.31	0.56*	0.41
14-MT	1.27*	1.85*	1.61*	1.40*
15-PaH	0.63*	1.35*	0.88*	0.52
16-PaC	0.23	0.20	− 0.14	− 0.17
17-Op	0.54	1.01*	0.72*	0.67*
18-Or	0.37	0.67*	0.36	0.20
19-Tr	0.13	0.71*	0.33	0.25
20-PerCa	− 0.26	− 0.03	0.22	− 0.16
21-PoC	0.37	0.21	0.14	0.05
22-PoCg	0.79*	1.24*	0.88*	0.45
23-PreC	0.39	− 0.04	0.22	0.14
24-PreCu	0.83*	0.99*	1.13*	0.95*
25-RoACg	0.55	0.41	0.76*	0.31
26-RoMF	0.80*	1.06*	1.22*	0.91*
27-SF	0.93*	0.98*	0.80*	0.77*
28-SP	0.79*	0.76*	0.97*	0.81*
29-ST	1.57*	1.76*	1.45*	1.51*
30-SM	1.02*	1.07*	1.24*	1.08*
31-FPol	0.54	0.67*	0.45	0.46
32-TPol	1.28*	1.38*	0.92*	0.86*
33-TrT	− 0.01	0.78*	0.37	0.20
34-Ins	0.79*	1.30*	1.40*	1.11*
Mean (significant group differences)	0.98	1.08	1.03	1.03

For each region of interest (ROI) of the Desikan-Killiany atlas, the effect size (Cohen's d) is given for each measurement method, including FreeSurfer (FS6), the Computational Anatomy Toolbox (CAT12), the Laplacian approach and the Euclidean Distance Transform (EDT) approach.

Significant mean cortical thickness difference ($P < 0.05$, FDR-correction) between groups is indicated with *.

The values at the bold represent the mean of the significant effect size for the given method

and last authors emails. In conclusion, the current work has shown reliable test–retest cortical thickness measures with surface- and voxel-based methods, and good agreement regarding their ability to detect atrophies in the cortex. Inter-method absolute differences reported might be attributed to systematic errors as evidenced. Each methodology has its

own foundations which make every technique unique; surface-based methods attain higher precision calculations if a reliable and accurate model is reconstructed, at the expense of more computational demands. Besides, voxel-based approaches have been limited by spatial resolution, as was also suggested by the computational phantom for the EDT.

Acknowledgements This work was supported by grant number 479002, and registration number 629404, from CONACYT (Consejo Nacional de Ciencia y Tecnología, Mexico), for a Master's in Science in Medical Physics at the UNAM. Data used in the preparation of this article were obtained from the MIRIAD database. We are also grateful to M.Sc. L. González-Santos for technical support and J. González-Norris for editing of the manuscript. The MIRIAD investigators did not participate in analysis or writing of this report. The MIRIAD dataset is made available through the support of the UK Alzheimer's Society (Grant RF116).

Funding The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC) and funding from the UK Alzheimer's Society and the Medical Research Council. Also receives funding from the EPSRC (EP/H046410/1) and the Comprehensive Biomedical Research Centre (CBRC) Strategic Investment Award (Ref. 168). And support by the Medical Research Council (grant number MR/J014257/1). This work was supported by the National Institute for Health Research (NIHR) Biomedical Research Unit in Dementia based at University College London Hospitals (UCLH), University College London (UCL). The Multi-Modal MRI Reproducibility Resource (MMRR) dataset received funding from NIH/NCCR P41RR15241 NIH/NINDS 1R01NS056307. The MIRIAD or MMRIRR investigators did not participate in analysis or writing of this report.

Data Availability Data used in the preparation of this article were obtained from the public MIRIAD database (<https://www.ucl.ac.uk/drc/research/methods/minimal-interval-resonance-imaging-alzheimers-disease-miriad>) And the Multi-Modal MRI Reproducibility Resource at (<https://www.nitrc.org/projects/multimodal>).

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethics Approval All studies counted with authorization of their IRB.

Informed consent All authors have read and consented to the submission of the final version of the manuscript to the journal.

References

- Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38(1):95–113
- Blustajn J, Krystal S, Taussig D, Ferrand-Sorbets S, Dorfmueller G, Fohlen M (2019) Optimizing the detection of subtle insular lesions on MRI when insular epilepsy is suspected. *Am J Neuroradiol* 40(9):1581–1585
- Cardinale F, Chinnici G, Bramerio M, Mai B, Sartori I, Cossu M, Lo Russo G, Castana L, Colombo N, Caborni C, De Momi E, Ferrigno G (2014) Validation of FreeSurfer-estimated brain cortical thickness: comparison with histologic measurements. *Neuroinformatics* 12(4):535–542
- Clarkson MJ, Cardoso MJ, Ridgeway GR, Modat M, Leung KK, Rohrer JD, Fox NC, Ourselin S (2011) A comparison of voxel and surface based cortical thickness estimation methods. *Neuroimage* 57(3):856–865
- Dahnke R, Yotter RA, Gaser C (2013) Cortical thickness and central surface estimation. *Neuroimage* 65:336–348
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis I. Segmentation and surface reconstruction. *Neuroimage* 9(2):179–194
- Das SR, Avants BB, Grossman M, Gee JC (2009) Registration based cortical thickness measurement. *Neuroimage* 45(3):867–879
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31(3):968–980
- Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci USA* 97(20):11050–11055
- Fischl B, Liu A, Dale AM (2001) Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging* 20(1):70–80
- Fischl B, Salat DH, van der Kouwe AJW, Makris N, Ségonne F, Quinn BT, Dale AM (2004) Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 23:S69–84
- Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale AM, Dickerson B, Fischl B (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32(1):180–194
- Hutton C, De Vita E, Ashburner J, Deichmann R, Turner R (2008) Voxel-based cortical thickness measurements in MRI. *Neuroimage* 40(4):1701–1710
- Hutton C, Draganski B, Ashburner J, Weiskopf N (2009) A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage* 48(2–8):371–380
- Jones SE, Buchbinder BR, Aharon I (2000) Three-dimensional mapping of cortical thickness using Laplace's equation. *Hum Brain Mapp* 11(1):12–32
- Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JAD, Bogovic JA, Hua J, Chen M, Jarso S, Smith SA, Joel S, Mori S, Pekar JJ, Barker PB, Prince JL, van Zijl PCM (2011) Multi-parametric neuroimaging reproducibility: a 3-T resource study. *Neuroimage* 54(4):2854–2866
- Li Q, Pardoe H, Lichter R, Werden E, Raffelt A, Cumming T, Brodtmann A (2015) Cortical thickness estimation in longitudinal stroke studies: a comparison of 3 measurement methods. *NeuroImage Clin* 8:526–535
- Lüsebrink F, Wollrab A, Speck O (2013) Cortical thickness determination of the human brain using high resolution 3 T and 7 T MRI data. *Neuroimage* 70:122–131
- Malone IB, Cash D, Ridgway GR, MacManus DG, Ourselin S, Fox NC, Schott JM (2013) MIRIAD-Public release of a multiple time point Alzheimer's MR imaging dataset. *Neuroimage* 70:33–36
- Marquez J, Bloch I, Bousquet T, Schmitt T, Grangeat C (2005) Shape-based averaging for craniofacial anthropometry (abstract). In: *IEEE sixth Mexican international conference on computer science*, pp 314–319
- Mateos MJ, Gastelum-Strozzi A, Barrios FA, Bribiesca E, Alcauter S, Marquez-Flores JA (2020) A novel voxel-based method to estimate the cortical sulci width and its application to compare patients with Alzheimer's disease to controls. *Neuroimage* 207:116343
- Price CC, Wood MF, Leonard CM, Towler S, Ward J, Montijo H, Kellison I, Bowers D, Monk T, Newcomer JC, Schmalfuss I (2011) Entorhinal cortex volume in older adults: reliability and validity considerations for three published measurement protocols. *J Int Neuropsychol Soc* 16(5):846–855

- Righart R, Schmidt P, Dahnke R, Biberacher V, Beer A, Buck B, Hemmer B, Kirschke JS, Zimmer C, Gaser C, Mühlau M (2017) Volume versus surface-based cortical thickness measurements: A comparison study with healthy controls and multiple sclerosis patients. *PLoS One* 12(7):e0179590
- Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH, van der Kouwe A, Jenkins BG, Dale AM, Fischl B (2002) Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58(5):695–701
- Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RSR, Busa E, Morris JC, Dale AM, Fischl B (2004) Thinning of the cerebral cortex in aging. *Cereb Cortex* 14(7):721–730
- Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B (2004) A hybrid approach to the skull-stripping problem in MRI. *Neuroimage* 22(3):1060–1075
- Seiger R, Ganger S, Kranz GS, Hahn A, Lanzenberger R (2018) Cortical thickness estimation of FreeSurfer and the CAT12 toolbox in patients with Alzheimer's disease and healthy controls. *J Neuroimaging* 28(5):515–523
- Talairach J, Tournoux P (1988) *Coplanar stereotaxic atlas of the human brain*. Thieme Medical Publishers, New York
- Yotter RA, Dahnke R, Thompson PM, Gaser C (2011a) Topological correction of brain surface meshes using spherical harmonics. *Hum Brain Mapp* 32(7):1109–1124
- Yotter RA, Thompson PM, Gaser C (2011b) Algorithm to improve the reparameterization of spherical mappings of brain surface meshes. *J Neuroimaging* 21(2):e134–147

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.