# Lesser of Two Evils Improves Learning in the Context of Cortical Thickness Estimation Models - Choose Wisely

Filip Rusak[1,2](✉), Rodrigo Santa Cruz[2], Elliot Smith[3], Jurgen Fripp[2], Clinton Fookes[1], Pierrick Bourgeat[2], and Andrew P. Bradley[1]

[1] Queensland University of Technology, Brisbane, QLD, Australia
[2] CSIRO, Herston, QLD, Australia
filip.rusak@csiro.au
[3] Maxwell Plus, Brisbane, QLD, Australia

**Abstract.** Cortical thickness (CTh) is an important biomarker commonly used in clinical studies for a range of neurodegenerative and neurological conditions. In such studies, CTh estimation software packages are employed to estimate CTh from T1-weighted (T1-w) brain MRI scans. Since commonly used software packages (e.g. FreeSurfer) are time-consuming, the fast-inference Machine Learning (ML) CTh estimation solutions have gained much popularity. Recently, several ML regression-based solutions offering morphological properties (CTh, volume and curvature) estimation have emerged but typically achieved lower accuracy compared to mainstream alternatives. One of the reasons for such performance of the ML-based CTh estimation models is the inaccurate automatic labels typically used for their training. In this paper, we investigate the impact of automatic labels selection on the performance of the current state-of-the-art ML regression-based CTh estimation method - HerstonNet. We train two models on pairs of brain MRIs and FreeSurfer/DL+DiReCT automatic CTh measurements to investigate the benefits of using DL+DiReCT instead of, the more frequently used, FreeSurfer CTh measurements on the learning capability of a modified version of HerstonNet. Then, we evaluate the performance of the two trained models on three test sets with scans coming from four publicly available datasets. We show that HerstonNet trained on DL+DiReCT labels overall achieves a 13.3% higher Intraclass Correlation Coefficient (ICC) on a test set composed of ADNI and AIBL scans, 19.4% on OASIS-3 and 17.1% on SIMON dataset compared to the same model trained on FreeSurfer derived measurements. The results suggest that DL+DiReCT provides automatic labels more suitable for CTh estimation model training than FreeSurfer.

**Keywords:** Weak labels · Cortical thickness definition · Cortical thickness estimation · Model learning optimisation

This work was funded in part through an Australian Department of Industry, Energy and Resources CRC-P project between CSIRO, Maxwell Plus and I-Med Radiology Network.

H. V. Nguyen et al. (Eds.): DALI 2022, LNCS 13567, pp. 33–42, 2022.
https://doi.org/10.1007/978-3-031-17027-0_4

## 1 Introduction

Cortical thickness (CTh) is an important biomarker for the diagnosis and prognosis of neurodegenerative diseases, such as Alzheimer's disease (AD) [14]. Estimating and tracking CTh changes in a living brain may reveal insights into disease trajectory, quantitatively evaluate treatment effects, and enable correlations between brain regions and age, cognitive deterioration, genotype, or medication [1]. Despite the importance of CTh as a biomarker, a generally accepted gold standard for in-vivo CTh measurements currently does not exist [15]. An accepted gold standard does not exist as there is no standardised definition of CTh estimates and even if there was one, post-mortem histology measurements are unreliable [15]. In-vivo CTh measurements can only be estimated from human brain scans acquired using neuroimaging techniques such as magnetic resonance imaging (MRI) [9]. Manual CTh estimation from MRI scans is laborious, subjective and requires a high level of expertise which makes it infeasible in practice [7]. Therefore robust software tools such as FreeSurfer [6] are typically utilised for automatic CTh estimation from brain MRIs [21]. However, such tools are also time-consuming (FreeSurfer - up to 10 h per scan) since their CTh estimation relies on throughputs such as segmentation maps, partial volume maps, and triangular meshes that need to be constructed before estimation takes place [20]. Therefore, such tools are not adequate for clinical applications requiring timely results [20]. Recently, a couple of studies approached the problem of time-consuming automatic CTh, volume and curvature estimations by proposing Deep Learning (DL)-based solutions that reduce estimation time from hours to seconds at the expense of estimation accuracy [18,20]. Currently, DL-based methods [18,20] are trained and tested against FreeSurfer measurements that are considered to be the best approximation of the gold-standard ground truth, so called silver-standard. By doing so, these DL-based methods learn FreeSurfer-specific CTh definition bias as well as FreeSurfer software-specific biases coming from the method design and its implementation. While biases cannot be completely avoided due to a non-existing true bias-free gold-standard CTh measurements, the choice of CTh measurement for training may impact DL-model learning capabilities.

In this paper, we investigate the impact of the choice of automatic labels (CTh measurements) on the training of the state-of-the-art DL-based CTh estimation method - HerstonNet [20]. Firstly, we modify the original HerstonNet solution to decouple CTh from the other morphological estimations (volume and curvature). Then we train the decoupled HerstonNet solution on CTh measurements derived by FreeSurfer and DL+DiReCT [17] to ensure different CTh definition and method biases. Finally, we evaluate the trained models on three subsets of brain MRIs from four datasets ADNI, AIBL, OASIS and SIMON. The contributions of this paper are the following: i) insights into the impact of bias labels (CTh estimations) choice on the training of HerstonNet and ii) comparison in performance (intraclass correlation and test-retest) between HerstonNet trained with FreeSurfer (silver standard) and DL+DiReCT-derived CTh estimations on three datasets.

## 2 Methods

**Data & Pre-processing.** In this work, we used T1-weighted (T1-w) brain MRI scans from four datasets: Alzheimer's Disease Neuroimaging Initiative (ADNI)[1] [11,23] Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL) [19], Open Access Series of Imaging Studies (OASIS) [13] and Single Individual volunteer for Multiple Observations across Networks (SIMON) [5]. The models under test were trained, validated and tested on a subset composed of 9310 MRIs from both ADNI and AIBL datasets. The subset was split into training, validation and test sets, roughly in the 60:15:25 ratio, respectively, with no overlap between subsets to avoid data leakage. Further insights into the subset, data split, demographics and pathologies are detailed in Table 1. The MRIs taken from ADNI and AIBL datasets were pre-processed by correcting the bias field in the brain region of interest (ROI) [22]. Further, 9310 bias-field corrected MRIs from ADNI and AIBL datasets, together with 2720 MRIs from OASIS-3 and 96 MRIs from SIMON datasets, were rigidly registered to MNI-space (181 × 217 × 181 voxels) and z-score intensity normalised with the mean value computed in the brain ROI. The MRIs from the OASIS-3 and SIMON datasets were used for testing only.

**Automatic Cortical Thickness Measurements.** In this work, we employed two CTh estimation tools, FreeSurfer cross-sectional pipeline [8] and DL+DiReCT [17]. FreeSurfer cross-sectional pipeline relies on the construction of white matter (WM) and grey matter (GM) surfaces to map morphometric mea-

**Table 1.** Insights into the train, validation and test subsets of ADNI and AIBL datasets, data split, demographics and pathology across subsets. $S$ annotates the number of subjects while $N$ stands for the number of data points. The column *Other* comprises subjects/data points with under-represented or unavailable pathology.

| | Healthy Control (HC) | | Mild Cognitive Impairment (MCI) | | Alzheimer's Disease (AD) | | Other | |
|---|---|---|---|---|---|---|---|---|
| | *S* | *N* | *S* | *N* | *S* | *N* | *S* | *N* |
| Train | 441 | 1127 | 996 | 1972 | 537 | 922 | 1013 | 1611 |
| Validation | 113 | 284 | 243 | 487 | 157 | 268 | 242 | 352 |
| Test | 190 | 482 | 391 | 736 | 252 | 420 | 421 | 649 |
| Overall | **744** | **1893** | **1630** | **3195** | **946** | **1610** | **1676** | **2612** |
| | *Mean Age (± STD)* | *% Female* | *Mean Age (± STD)* | *% Female* | *Mean Age (± STD)* | *% Female* | *Mean Age (± STD)* | *% Female* |
| Train | 74.79 ± 6.42 | 55.01 | 73.98 ± 6.90 | 58.87 | 76.30 ± 5.72 | 69.20 | 72.74 ± 7.25 | 49.65 |
| Validation | 74.02 ± 7.35 | 72.18 | 74.03 ± 6.16 | 56.88 | 76.74 ± 6.23 | 71.27 | 73.27 ± 7.55 | 44.68 |
| Test | 73.58 ± 6.37 | 49.38 | 73.85 ± 6.23 | 59.24 | 76.16 ± 6.27 | 77.62 | 72.65 ± 7.22 | 43.33 |
| Overall | **74.37 ± 6.57** | **56.15** | **73.96 ± 6.64** | **58.65** | **76.34 ± 5.95** | **71.74** | **72.80 ± 7.28** | **47.31** |

[1] Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org [11,23].

surements on the reconstructed surface. Once WM and GM surfaces are reconstructed, FreeSurfer estimates CTh as an average minimum distance between vertices on GM and WM surfaces and vice versa. DL+DiReCT estimates CTh by segmenting neuroanatomy using a DL-based model called DeepSCAN [16] followed by diffeomorphic registration-based CTh (DiReCT) [3] measurements. DeepSCAN segments T1-w brain MRIs into GM and WM segmentation as well as parcellation, while DiReCT obtains CTh from MRIs and corresponding segmentations. DiReCT defines CTh as a distance measure between corresponding cerebrospinal fluid (CSF)/GM and GM/WM interfaces, where continuous one-to-one mapping is ensured by diffeomorphic registration [3,17].

**HerstonNet, Modifications and Training.** In this section, we focus on HerstonNet [20], the state-of-the-art regression-based neural network for efficient brain morphometry analysis, and modifications we made to decouple the CTh estimation from the rest of the morphometry measurements. HerstonNet is a 3D ResNet-based neural network that learns rich features directly from MRI. Throughout the multi-scale regression scheme, HerstonNet predicts morphometric measures from feature maps of various resolutions, which robustly leverages the network optimisation to avoid poor quality minima and lower the prediction variance. Santa Cruz *et al.* trained HerstonNet on pairs of images, and FreeSurfer derived CTh, volume and curvature. Further, the authors employed a data-augmentation strategy by applying Gaussian noise injection, translations (up to 15 voxels), and rotations (up to 30°) on input brain MRIs. After 170 epochs of training, they apply Stochastic Weight Averaging (SWA) optimisation technique [10] to improve the generalisation of the model. In our experiments, we follow the architecture, training strategy, and data splits as described in [20] with a couple of major modifications. Instead of predicting the CTh, volume, and curvature, we modify the output size and restrict predictions to CTh only. Further, we skip the SWA step to emphasise the model generalisability difference between models trained on different automatic CTh measurements. We trained the modified version of HerstonNet according to [20] with the difference in the number of epochs. Instead of training modified HerstonNet models for 170 epochs, we stopped training after 140 h (142 epochs) when both models' losses plateaued. Both models were optimised by minimising the mean squared error (MSE) on batches of six samples by employing the Adam optimiser with a learning rate $10^{-4}$. We also followed the augmentation strategy detailed in [20].

## 3 Experiments and Results

**Visualisation of CTh Measurement Difference.** To better understand the difference between FreeSurfer and DL+DiReCT measurements, we mapped the region-wise mean absolute difference and standard deviation (STD) to the brain template meshes (Fig. 1). We considered 34 ROIs per hemisphere defined by the Desikan-Killiany atlas [4]. The mean absolute difference and the STD (in mm) were computed on the training set that comprises 5582 MRI scans from ADNI and AIBL datasets. According to Fig. 1, in the parietal lobe, we measured

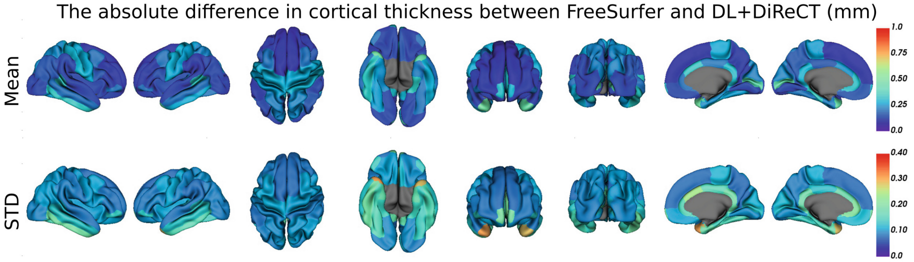
The absolute difference in cortical thickness between FreeSurfer and DL+DiReCT (mm)
Mean
STD
1.0
0.75
0.50
0.25
0.0
0.40
0.30
0.20
0.10
0.0

**Fig. 1.** The absolute CTh difference between FreeSurfer and DL+DiReCT computed on 5582 scans randomly selected from ADNI and AIBL datasets.

a mean absolute difference around 0.25 mm, while in the frontal and occipital lobes we measured minimal difference between the CTh estimations. In the temporal lobe, the methods predominantly differ in inferior temporal gyrus, banks of the superior temporal sulcus, fusiform gyrus, entorhinal cortex and temporal pole, where the difference in CTh estimations reaches 0.5 mm. Overall, the differences between the estimated CTh are mainly symmetric on both hemispheres. Given that the average CTh spans across regions in the [2.5 to 3] mm interval [9], the difference of up to 0.5 mm between the CTh estimation methods is significant. We confirmed the significance of CTh estimations difference in all 34 brain ROI on both hemispheres by performing a t-test followed by Bonferroni correction. In the context of model training, validation and testing, such a significant difference between FreeSurfer and DL+DiReCT CTh estimations imply that models should not be trained with FreeSurfer while being validated and tested with DL+DiReCT CTh estimations, and vice versa.

**FreeSurfer-Trained vs. DL+DiReCT-Trained HerstonNet Model Estimations.** To evaluate the impact of labels (CTh measurements) derived by FreeSurfer and DL+DiReCT on model learning and performance, we train two modified HerstonNet models, FreeSurfer-trained and DL+DiReCT-trained HerstonNet. FreeSurfer-trained HerstonNet was trained on pairs of brain MRIs and corresponding FreeSurfer CTh estimations, while the DL+DiReCT-trained HerstonNet was trained on pairs of brain MRIs and corresponding DL+DiReCT CTh estimations. Both models were trained on the same train set comprising 5582 MRIs from ADNI and AIBL datasets. Each training sample is represented with an MRI and corresponding CTh estimations of 68 regions (34 regions/hemisphere). Once trained, we tested both models on three different test sets, composed of ADNI+AIBL, OASIS and SIMON MRIs. For model performance evaluation, we used the intraclass correlation coefficient (ICC) as well as a 95% confidence interval. Based on the research reliability guidelines for ICC values reporting [12], we computed the two-way mixed effects, absolute agreement, and single rater (ICC(2,1)) between the predicted and either FreeSurfer CTh estimations as ground-truth in the case of FreeSurfer-trained HerstonNet and DL+DiReCT CTh estimations as ground-truth in the case of DL+DiReCT-trained HerstonNet model. The negative ICC values indicate a negative correlation, the ICC value of zero indicates no correlation, while the ICC value of one indicates the perfect correlation between predicted and ground truth values. The overall ICC scores on test sets are presented in Table 2. The DL+DiReCT-trained HerstonNet achieved a higher ICC score than the FreeSurfer-trained HerstonNet model on all three datasets. The best performing model (DL+DiReCT-trained HerstonNet) achieved the highest ICC score on the dataset composed of ADNI and AIBL scans. Such an observation is intuitive since the model was trained on MRI scans that belong to either ADNI or AIBL datasets. Nevertheless, DL+DiReCT-trained HerstonNet also achieved a higher ICC score on the other two datasets (OASIS-3 and SIMON), which were not involved in the training. Since both models were not exposed to any images from the OASIS-3 and SIMON datasets, the fact that DL+DiReCT-trained HerstonNet performed better than FreeSurfer-trained HerstonNet is a strong indication of higher generalisability. Overall, DL+DiReCT-trained HerstonNet achieved 13.3% higher ICC score than FreeSurfer-trained HerstonNet on the dataset composed of ADNI and AIBL MRIs, 19.4% higher ICC on the OASIS-3 dataset and 17.1% on SIMON dataset.

**Table 2.** The mean ICC value and standard deviation, computed over all 34 regions on both hemispheres, achieved by both modified HerstonNet models trained on pairs of MRI and either FreeSurfer or DL+DiReCT CTh estimates on three datasets: ADNI+AIBL, OASIS, and SIMON. The difference in achieved is expressed in %. The sign ↑ denotes that the higher metric values suggest better results.

| Test set | Number of MRI scans | Intraclass Correlation Coefficient (ICC) ↑ | | Difference (%) |
|---|---|---|---|---|
| | | HerstonNet (FreeSurfer) | HerstonNet (DL+DiReCT) | |
| ADNI + AIBL | 2282 | $0.767 \pm 0.093$ | $\mathbf{0.9 \pm 0.047}$ | 13.3% |
| OASIS-3 | 2720 | $0.227 \pm 0.125$ | $\mathbf{0.421 \pm 0.194}$ | 19.4% |
| SIMON | 96 | $0.122 \pm 0.12$ | $\mathbf{0.293 \pm 0.15}$ | 17.1% |

We also compared the region-wise achieved ICC scores and 95% confidence intervals of FreeSurfer-trained and DL+DiReCT-trained HerstonNet models. Based on the discussion provided in [2] and following [18,20], we utilise the ICC intervals, commonly used in clinical applications. The ICC intervals are visualised at the bottom of Fig. 2. The comparison of ICC scores and 95% confidence intervals achieved by FreeSurfer-trained and DL+DiReCT-trained HerstonNet are visualised in Fig. 2. According to Fig. 2, DL+DiReCT-trained HerstonNet achieves higher ICC than FreeSurfer-trained HerstonNet in all 34 brain regions on both hemispheres. The mean ICC values achieved by DL+DiReCT-trained HerstonNet in all 34 brain regions on both hemispheres fall into the ICC $\geq 0.75$ (excellent) ICC interval, while ICC values achieved by FreeSurfer-trained HerstonNet mainly fall into $0.6 \geq \text{ICC} < 0.75$ (good) ICC interval. Further, DL+DiReCT-trained HerstonNet overall achieved tighter 95% confidence intervals than the FreeSurfer-trained HerstonNet model. There are six brain regions on both hemispheres where FreeSurfer-trained HerstonNet achieved tighter 95% confidence intervals than DL+DiReCT-trained HerstonNet. The six brain regions

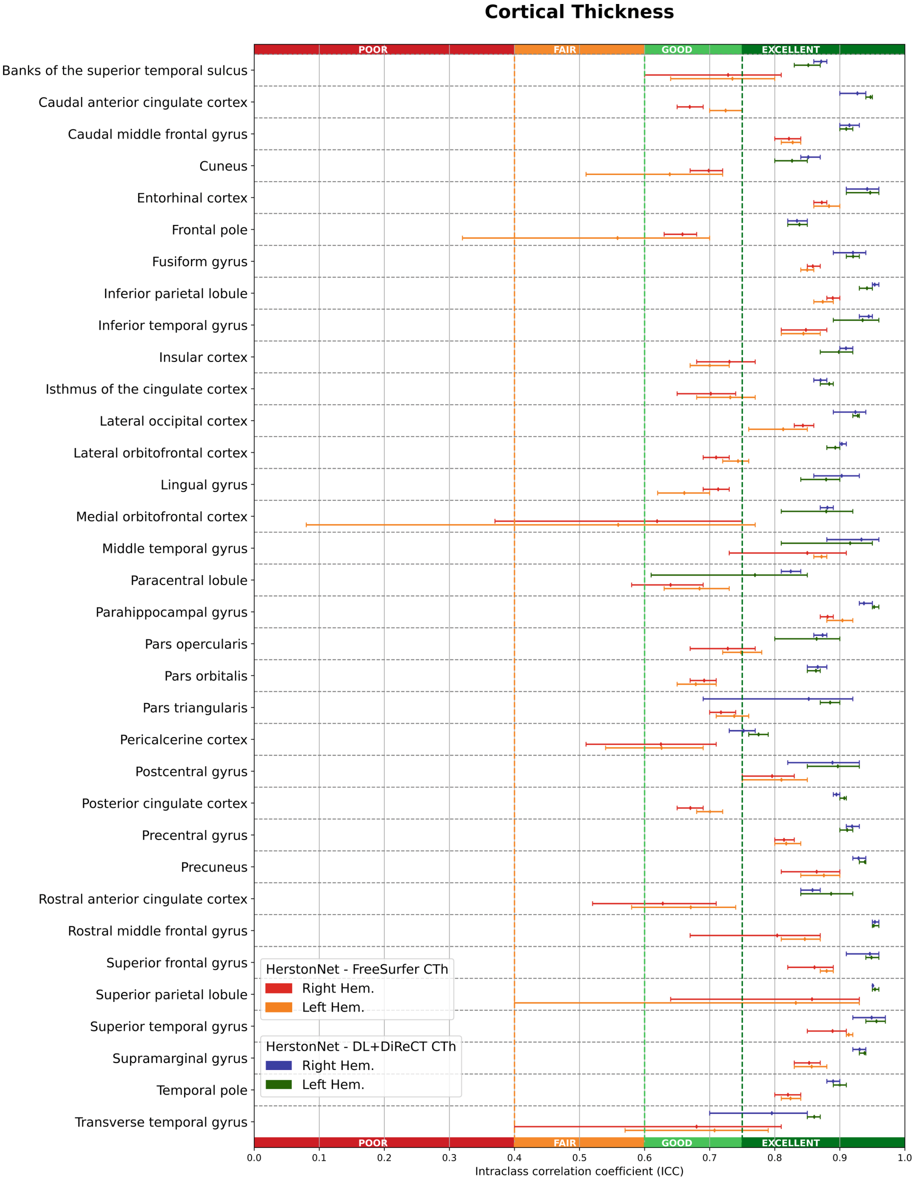

**Fig. 2.** ICC scores with 95% confidence intervals, computed on 2282 MRIs coming from ADNI and AIBL datasets, for FreeSurfer-trained and DL+DiReCT-trained HerstonNet CTh estimations in 34 cortical regions per hemisphere.

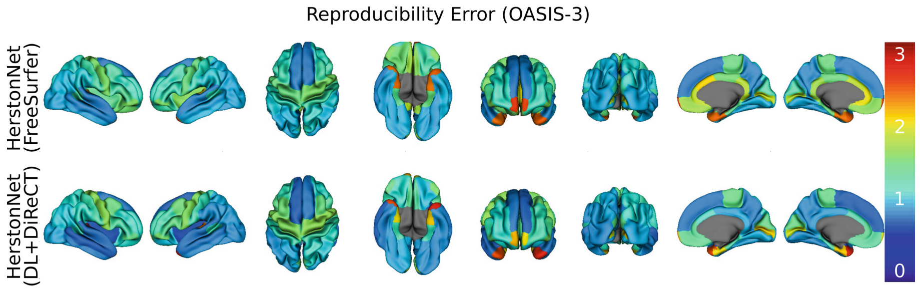

**Fig. 3.** Test-retest - colour coded reproducibility error (%) computed on subset OASIS-3 and mapped on a template mesh.

where DL+DiReCT-trained HerstonNet achieves wider confidence interval than FreeSurfer-trained HerstonNet on the left hemisphere are the entorhinal cortex, pars opercularis, paracentral lobe, the inferior temporal, middle temporal and superior temporal gyrus. The six brain regions where DL+DiReCT-trained HerstonNet achieves wider confidence interval than FreeSurfer-trained HerstonNet on the right hemisphere are the entorhinal cortex, lateral occipital cortex, pars triangularis, fusiform, lingual and postcentral gyrus.

**Reproducibility Evaluation (Test-Retest).** To evaluate the robustness of FreeSurfer-trained and DL+DiReCT-trained HerstonNet models, we computed the region-wise reproducibility error ($\epsilon$), formally defined as follows:

$$\epsilon = \frac{100}{N} \sum_{i=1}^{N} \left( \frac{1}{n_i} \sum_{t=1}^{n_i} \frac{|\mu_{i,t} - \mu_i|}{\mu_i} \right), \mu_i = \frac{1}{n_i} \sum_{t=1}^{n_i} m_{i,t} \qquad (1)$$

where $N$ stands for the number of scanning sessions of the same subject, $n_i$ denotes the number of scans obtained in a session $i$, while $m_{i,t}$ denotes the measurement computed by the algorithm from the $t^{th}$ scan in the session $i$. For the computation of $\epsilon$ we used 592 subjects from OASIS-3 dataset, with a total number of 1536 scans acquired in 757 sessions. Once computed, we mapped $\epsilon$ per region on a template mesh (Fig. 3). According to Fig. 3, DL+DiReCT-trained HerstonNet achieved slightly lower or equal $\epsilon$ than FreeSurfer-trained HerstonNet in all regions except temporal pol on both hemispheres. While the difference in $\epsilon$ across 34 brain regions and both hemispheres is not significant, such an outcome suggests higher reliability of DL+DiReCT-trained HerstonNet over FreeSurfer-trained HerstonNet.

## 4 Conclusion

In this paper, we investigated the benefits of using two different automatic CTh estimations for the training of HerstonNet - the state-of-the-art DL-based model for direct CTh estimation from brain MRIs. The obtained results indicate that training HerstonNet on DL+DiReCT CTh estimations makes the model more generalisable and robust when compared with HerstonNet trained on FreeSurfer CTh. However, more experiments are needed to evaluate whether such a conclusion is generalisable to other automatic CTh estimations, DL-based models and datasets. For future work, we plan to thoroughly investigate the impact of several automatic CTh estimations on the training of DL models as well as the main drivers behind the learning acceleration.

## References

1. Aganj, I., Sapiro, G., Parikshak, N., Madsen, S.K., Thompson, P.M.: Measurement of cortical thickness from MRI by minimum line integrals on soft-classified tissue. Human Brain Mapp. **30**(10), 3188–3199 (2009)
2. Cicchetti, D.V.: Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assessm. **6**(4), 284 (1994)
3. Das, S.R., Avants, B.B., Grossman, M., Gee, J.C.: Registration based cortical thickness measurement. Neuroimage **45**(3), 867–879 (2009)
4. Desikan, R.S.,et al.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into Gyral based regions of interest. Neuroimage **31**(3), 968–980 (2006)
5. Duchesne, S., et al.: Structural and functional multi-platform MRI series of a single human volunteer over more than fifteen years. Sci. Data **6**(1), 1–9 (2019)
6. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
7. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. **97**(20), 11050–11055 (2000)
8. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis: Ii: inflation, flattening, and a surface-based coordinate system. Neuroimage **9**(2), 195–207 (1999)
9. Hutton, C., De Vita, E., Ashburner, J., Deichmann, R., Turner, R.: Voxel-based cortical thickness measurements in MRI. Neuroimage **40**(4), 1701–1710 (2008)
10. Izmailov, P., Podoprikhin, D., Garipov, T., Vetrov, D., Wilson, A.G.: Averaging weights leads to wider optima and better generalization. arXiv preprint arXiv:1803.05407 (2018)
11. Jack, C.R., Jr., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magnet. Resonan. Imag. Off. J. Int. Soc. Magnet. Resonan. Med. **27**(4), 685–691 (2008)
12. Koo, T.K., Li, M.Y.: A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropract. Med. **15**(2), 155–163 (2016)
13. LaMontagne, P.J., et al.: Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. MedRxiv (2019)
14. Li, Q., Pardoe, H., Lichter, R., Werden, E., Raffelt, A., Cumming, T., Brodtmann, A.: Cortical thickness estimation in longitudinal stroke studies: a comparison of 3 measurement methods. NeuroImage Clin. **8**, 526–535 (2015)
15. Lüsebrink, F., Wollrab, A., Speck, O.: Cortical thickness determination of the human brain using high resolution 3 t and 7 t MRI data. Neuroimage **70**, 122–131 (2013)
16. McKinley, R., et al.: Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks. Sci. Rep. **11**(1), 1–11 (2021)

17. Rebsamen, M., Rummel, C., Reyes, M., Wiest, R., McKinley, R.: Direct cortical thickness estimation using deep learning-based anatomy segmentation and cortex parcellation. Human Brain Mapp. **41**(17), 4804–4814 (2020)
18. Rebsamen, M., Suter, Y., Wiest, R., Reyes, M., Rummel, C.: Brain morphometry estimation: from hours to seconds using deep learning. Front. Neurol. **11**, 244 (2020)
19. Rowe, C.C., et al.: Amyloid imaging results from the Australian imaging, biomarkers and lifestyle (AIBL) study of aging. Neurobiol. Aging **31**(8), 1275–1283 (2010)
20. Santa Cruz, R., et al.: Going deeper with brain morphometry using neural networks. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 711–715. IEEE (2021)
21. Tustison, N.J., et al.: Large-scale evaluation of ants and freesurfer cortical thickness measurements. Neuroimage **99**, 166–179 (2014)
22. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE Trans. Med. Imaging **18**(10), 885–896 (1999)
23. Weiner, M.W., et al.: The Alzheimer's disease neuroimaging initiative 3: continued innovation for clinical trial improvement. Alzheimer's Dementia **13**(5), 561–571 (2017)