

1

2 ANTsX: A dynamic ecosystem for 3 quantitative biological and medical imaging

4 Nicholas J. Tustison^{1,9}, Philip A. Cook², Andrew J. Holbrook³, Hans J. Johnson⁴, John
5 Muschelli⁵, Gabriel A. Devenyi⁶, Jeffrey T. Duda², Sandhitsu R. Das², Nicholas C. Cullen⁷,
6 Daniel L. Gillen⁸, Michael A. Yassa⁹, James R. Stone¹, James C. Gee², Brian B. Avants¹ for
7 the Alzheimer's Disease Neuroimaging Initiative[†]

8 ¹Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

9 ²Department of Radiology, University of Pennsylvania, Philadelphia, PA

10 ³Department of Biostatistics, University of California, Los Angeles, CA

11 ⁴Department of Electrical and Computer Engineering, University of Iowa, Philadelphia, PA

12 ⁵School of Public Health, Johns Hopkins University, Baltimore, MD

13 ⁶Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, QC

14 ⁷Lund University, Scania, SE

15 ⁸Department of Statistics, University of California, Irvine, CA

16 ⁹Department of Neurobiology and Behavior, University of California, Irvine, CA

17 Corresponding author:

18 Nicholas J. Tustison, DSc

19 Department of Radiology and Medical Imaging

20 University of Virginia

21 ntustison@virginia.edu

22

23 [†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did
24 not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/AD_NI_Acknowledgement_List.pdf

²⁷ **Abstract**

²⁸ The Advanced Normalizations Tools ecosystem, known as ANTsX, consists of multiple open-
²⁹ source software libraries which house top-performing algorithms used worldwide by scientific
³⁰ and research communities for processing and analyzing biological and medical imaging data.
³¹ The base software library, ANTs, is built upon, and contributes to, the NIH-sponsored
³² Insight Toolkit. Founded in 2008 with the highly regarded Symmetric Normalization image
³³ registration framework, the ANTs library has since grown to include additional functionality.
³⁴ Recent enhancements include statistical, visualization, and deep learning capabilities through
³⁵ interfacing with both the R statistical project (ANTsR) and Python (ANTsPy). Additionally,
³⁶ the corresponding deep learning extensions ANTsRNet and ANTsPyNet (built on the popular
³⁷ TensorFlow/Keras libraries) contain several popular network architectures and trained models
³⁸ for specific applications. One such comprehensive application is a deep learning analog
³⁹ for generating cortical thickness data from structural T1-weighted brain MRI. Not only
⁴⁰ does this significantly improve computational efficiency and provide comparable-to-superior
⁴¹ accuracy [over multiple criteria relative to](#) the existing ANTs pipelines but it also illustrates
⁴² the importance of the comprehensive ANTsX approach as a framework for medical image
⁴³ analysis.

⁴⁴ The ANTsX ecosystem: A brief overview

⁴⁵ Image registration origins

⁴⁶ The Advanced Normalization Tools (ANTs) is a state-of-the-art, open-source software toolkit
⁴⁷ for image registration, segmentation, and other functionality for comprehensive biological and
⁴⁸ medical image analysis. Historically, ANTs is rooted in advanced image registration techniques
⁴⁹ which have been at the forefront of the field due to seminal contributions that date back to
⁵⁰ the original elastic matching method of Bajcsy and co-investigators¹⁻³. Various independent
⁵¹ platforms have been used to evaluate ANTs tools since their early development. In a landmark
⁵² paper⁴, the authors reported an extensive evaluation using multiple neuroimaging datasets
⁵³ analyzed by fourteen different registration tools, including the Symmetric Normalization
⁵⁴ (SyN) algorithm⁵, and found that “ART, SyN, IRTK, and SPM’s DARTEL Toolbox gave
⁵⁵ the best results according to overlap and distance measures, with ART and SyN delivering
⁵⁶ the most consistently high accuracy across subjects and label sets.” **Participation in other**
⁵⁷ **independent competitions**^{6,7} provided additional evidence of the utility of ANTs registration
⁵⁸ and other tools. Despite the extremely significant potential of deep learning for image
⁵⁹ registration algorithmic development⁸, ANTs registration tools continue to find application
⁶⁰ in the various biomedical imaging research communities.

⁶¹ Current developments

⁶² Since its inception, though, ANTs has expanded significantly beyond its image registration
⁶³ origins. Other core contributions include template building⁹, segmentation¹⁰, image prepro-
⁶⁴ cessing (e.g., bias correction¹¹ and denoising¹²), joint label fusion^{13,14}, and brain cortical
⁶⁵ thickness estimation^{15,16} (cf Table 1). Additionally, ANTs has been integrated into multiple,
⁶⁶ publicly available workflows such as fMRIprep¹⁷ and the Spinal Cord Toolbox¹⁸. Frequently
⁶⁷ used ANTs pipelines, such as cortical thickness estimation¹⁶, have been integrated into Docker
⁶⁸ containers and packaged as Brain Imaging Data Structure (BIDS)¹⁹ and FlyWheel applica-
⁶⁹ tions (i.e., “gears”). It has also been independently ported for various platforms including
⁷⁰ Neurodebian²⁰ (Debian OS), Neuroconductor²¹ (the R statistical project), and Nipype²²

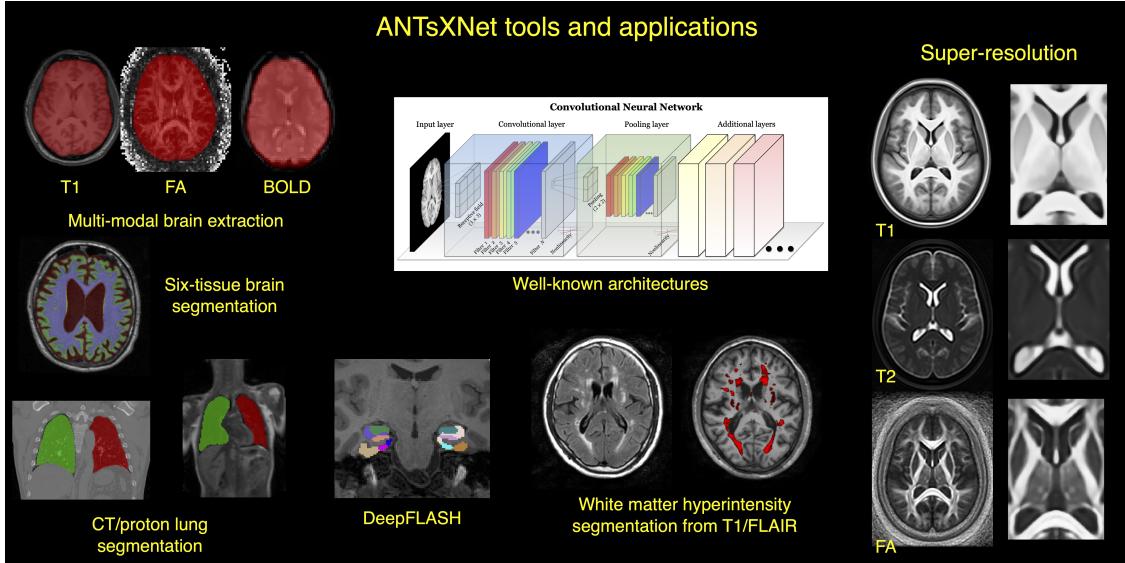


Figure 1: An illustration of the tools and applications available as part of the ANTsRNet and ANTsPyNet deep learning toolkits. Both libraries take advantage of ANTs functionality through their respective language interfaces—ANTsR (R) and ANTsPy (Python). Building on the Keras/TensorFlow language, both libraries standardize popular network architectures within the ANTs ecosystem and are cross-compatible. These networks are used to train models and weights for such applications as brain extraction which are then disseminated to the public.

71 (Python). Even competing softwares, such as FreeSurfer²³, have incorporated well-performing
 72 and complementary ANTs components^{11,12} into their own libraries. Finally, according to
 73 GitHub, recent unique “clones” have averaged 34 per day with the total number of clones
 74 being approximately twice that many. 50 unique contributors to the ANTs library have made
 75 a total of over 4500 commits. Additional insights into usage can be viewed at the ANTs
 76 GitHub website.

77 Over the course of its development, ANTs has been extended to complementary frameworks
 78 resulting in the Python- and R-based ANTsPy and ANTsR toolkits, respectively. These ANTs-
 79 based interfaces with extremely popular, high-level, open-source programming platforms
 80 have significantly increased the user base of ANTs and facilitated research workflows which
 81 leverage the advantages of these high-level programming languages. The rapidly rising
 82 popularity of deep learning motivated further recent enhancement of ANTs and its extensions.
 83 Despite the existence of an abundance of online innovation and code for deep learning
 84 algorithms, much of it is disorganized and lacks a uniformity in structure and external data

Functionality	Citations
SyN registration ⁵	2616
bias field correction ¹⁶	2188
ANTs registration evaluation ⁶	2013
joint label fusion ¹⁸	669
template generation ¹⁴	423
cortical thickness: implementation ²⁰	321
MAP-MRF segmentation ¹⁵	319
ITK integration ¹²	250
cortical thickness: theory ¹⁹	180

Table 1: The significance of core ANTs tools in terms of their number of citations (from October 17, 2020).

85 interfaces which would facilitate greater uptake. With this in mind, ANTsR spawned the deep
 86 learning ANTsRNet package which is a growing Keras/TensorFlow-based library of popular
 87 deep learning architectures and applications specifically geared towards medical imaging.
 88 Analogously, ANTsPyNet is an additional ANTsX complement to ANTsPy. Both, which we
 89 collectively refer to as “ANTsXNet”, are co-developed so as to ensure cross-compatibility
 90 such that training performed in one library is readily accessible by the other library. In
 91 addition to a variety of popular network architectures (which are implemented in both 2-D
 92 and 3-D), ANTsXNet contains a host of functionality for medical image analysis that have
 93 been developed in-house and collected from other open-source projects. For example, an
 94 extremely popular ANTsXNet application is a multi-modal brain extraction tool that uses
 95 different variants of the popular U-net²⁴ architecture for segmenting the brain in multiple
 96 modalities. These modalities include conventional T1-weighted structural MRI as well as
 97 T2-weighted MRI, FLAIR, fractional anisotropy and BOLD. Demographic specialization also
 98 includes infant T1-weighted and/or T2-weighted MRI. Additionally, we have included other
 99 models and weights into our libraries such as a recent BrainAGE estimation model²⁵, based
 100 on > 14,000 individuals; HippMapp3r²⁶, a hippocampal segmentation tool; the winning entry
 101 of the MICCAI 2017 white matter hyperintensity segmentation competition²⁷; MRI super
 102 resolution using deep-projection networks²⁸; and NoBrainer, a T1-weighted brain extraction
 103 approach based on FreeSurfer (see Figure 1).

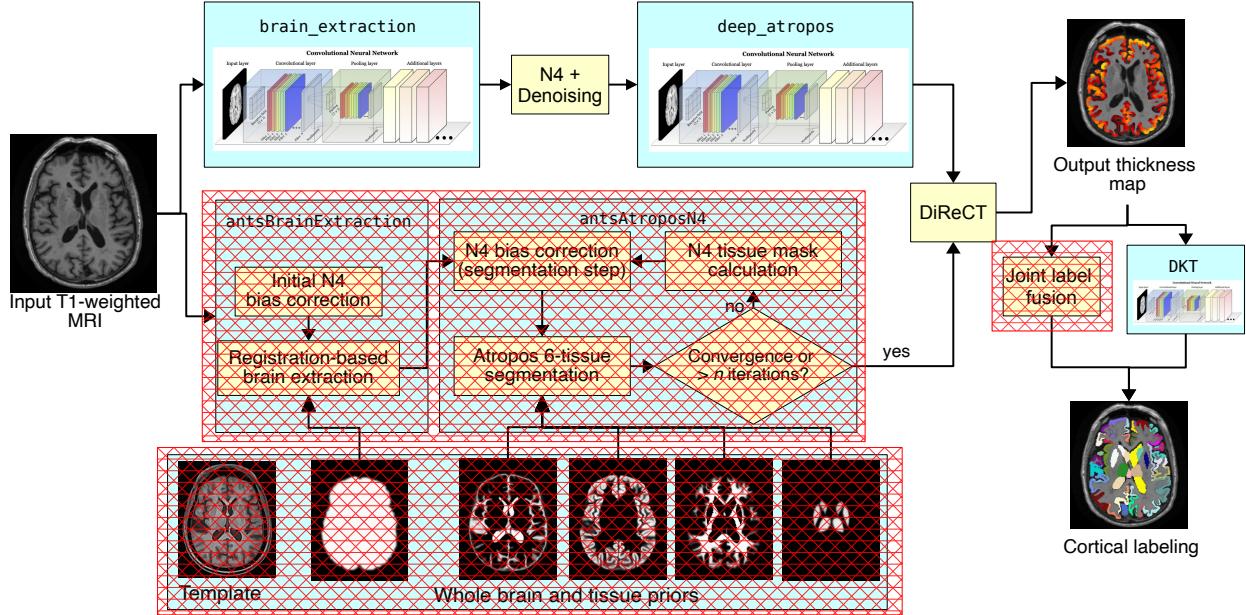


Figure 2: Illustration of the ANTsXNet cortical thickness pipeline and the relationship to its traditional ANTs analog. The hash-designated sections denote pipeline steps which have been obviated by the deep learning approach. These include template-based brain extraction, template-based n -tissue segmentation, and joint label fusion for cortical labeling.

104 The ANTsXNet cortical thickness pipeline

105 The most recent ANTsX innovation involves the development of deep learning analogs of
 106 our popular ANTs cortical thickness cross-sectional¹⁶ and longitudinal²⁹ pipelines within
 107 the ANTsXNet framework for, amongst other potential benefits, increased computational
 108 efficiency. Figure 2, adapted from our previous work¹⁶, illustrates some of the major changes
 109 associated with the single-subject pipeline. It should be noted that this improvement in
 110 efficiency is principally a result of eliminating deformable image registration from the pipeline—
 111 a step which has historically been used to propagate prior, population-based information
 112 (e.g., tissue maps) to individual subjects for such tasks as brain extraction³⁰ and tissue
 113 segmentation¹⁰ which is now configured within the neural networks.

114 These structural processing pipelines are currently available as open-source within the
 115 ANTsXNet libraries which underwent a thorough evaluation using both cross-sectional and
 116 longitudinal data and discussed within the context of our previous evaluations^{16,29}. Note
 117 that related work has been recently reported by external groups^{31,32}. Fortunately, these

¹¹⁸ overlapping contributions provide a context for comparison to motivate the utility of the
¹¹⁹ ANTsX ecosystem.

¹²⁰ Results

¹²¹ The original ANTs cortical thickness pipeline

¹²² The original ANTs cortical thickness pipeline¹⁶ consists of the following steps:

- ¹²³ • preprocessing: denoising¹² and bias correction³³;
- ¹²⁴ • brain extraction³⁰;
- ¹²⁵ • brain segmentation with spatial tissue priors¹⁰ comprising the
 - ¹²⁶ – cerebrospinal fluid (CSF),
 - ¹²⁷ – gray matter (GM),
 - ¹²⁸ – white matter (WM),
 - ¹²⁹ – deep gray matter,
 - ¹³⁰ – cerebellum, and
 - ¹³¹ – brain stem; and
- ¹³² • cortical thickness estimation¹⁵.

¹³³ Our recent longitudinal variant incorporates an additional step involving the construction of
¹³⁴ a single subject template (SST)⁹ coupled with the generation of tissue spatial priors of the
¹³⁵ SST for use with the processing of the individual time points as described above.

¹³⁶ Although the resulting thickness maps are conducive to voxel-based³⁴ and related analyses³⁵,
¹³⁷ here we employ the well-known Desikan-Killiany-Tourville (DKT)³⁶ labeling protocol (31
¹³⁸ labels per hemisphere) to parcellate the cortex for averaging thickness values regionally. This
¹³⁹ allows us to 1) be consistent in our evaluation strategy for comparison with our previous
¹⁴⁰ work^{16,29} and 2) leverage an additional deep learning-based substitution within the proposed
¹⁴¹ pipeline.

¹⁴² Overview of cortical thickness via ANTsXNet

¹⁴³ Note that the entire analysis/evaluation framework, from preprocessing to statistical analysis,
¹⁴⁴ is made possible through the ANTsX ecosystem and simplified through the open-source R
¹⁴⁵ and Python platforms. Preprocessing, image registration, and cortical thickness estimation
¹⁴⁶ are all available through the ANTsPy and ANTsR libraries whereas the deep learning steps
¹⁴⁷ are performed through networks constructed and trained via ANTsRNet/ANTsPyNet with
¹⁴⁸ data augmentation strategies and other utilities built from ANTsR/ANTsPy functionality.

¹⁴⁹ The brain extraction, brain segmentation, and DKT parcellation deep learning components
¹⁵⁰ were trained using data derived from our previous work¹⁶. Specifically, the IXI³⁷, MMRR³⁸,
¹⁵¹ NKI³⁹, and OASIS⁴⁰ data sets, and the corresponding derived data, comprising over 1200
¹⁵² subjects from age 4 to 94, were used for network training. Brain extraction employs a
¹⁵³ traditional 3-D U-net network²⁴ with whole brain, template-based data augmentation⁴¹
¹⁵⁴ whereas brain segmentation and DKT parcellation are processed via 3-D U-net networks
¹⁵⁵ with attention gating⁴² on image octant-based batches. We emphasize that a single model
¹⁵⁶ (as opposed to ensemble approaches where multiple models are used to produce the final
¹⁵⁷ solution²⁷) was created for each of these steps and was used for all the experiments described
¹⁵⁸ below.

¹⁵⁹ Cross-sectional performance evaluation

¹⁶⁰ Due to the absence of ground-truth, we utilize the evaluation strategy from our previous
¹⁶¹ work¹⁶ where we used cross-validation to build and compare age prediction models from
¹⁶² data derived from both the proposed ANTsXNet pipeline and the established ANTs pipeline.
¹⁶³ Specifically, we use “age” as a well-known and widely-available demographic correlate of
¹⁶⁴ cortical thickness⁴³ and quantify the predictive capabilities of corresponding random forest
¹⁶⁵ classifiers⁴⁴ of the form:

$$AGE \sim VOLUME + GENDER + \sum_{i=1}^{62} T(DKT_i) \quad (1)$$

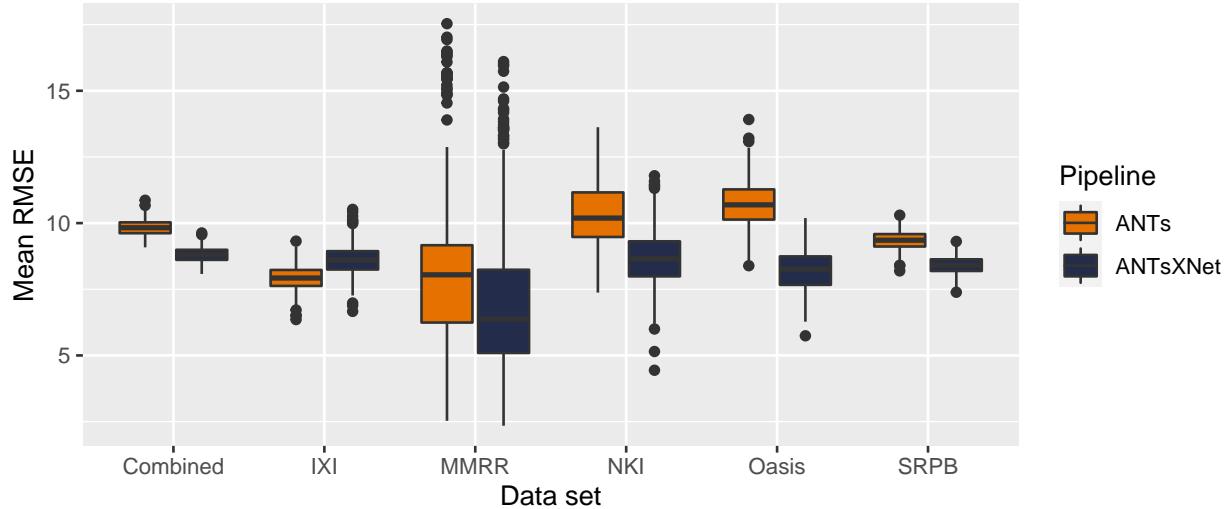


Figure 3: Distribution of mean RMSE values (500 permutations) for age prediction across the different data sets between the traditional ANTs and deep learning-based ANTsXNet pipelines. Total mean values are as follows: Combined—9.3 years (ANTs) and 8.2 years (ANTsXNet); IXI—7.9 years (ANTs) and 8.6 years (ANTsXNet); MMRR—7.9 years (ANTs) and 7.6 years (ANTsXNet); NKI—8.7 years (ANTs) and 7.9 years (ANTsXNet); OASIS—9.2 years (ANTs) and 8.0 years (ANTsXNet); and SRPB—9.2 years (ANTs) and 8.1 years (ANTsXNet).

with covariates *GENDER* and *VOLUME* (i.e., total intracranial volume). $T(DKT_i)$ is the average thickness value in the i^{th} DKT region. Root mean square error (RMSE) between the actual and predicted ages are the quantity used for comparative evaluation. As we have explained previously¹⁶, we find these evaluation measures to be much more useful than other commonly applied criteria as they are closer to assessing the actual utility of these thickness measurements as biomarkers for disease⁴⁵ or growth. For example, in recent work³¹ the authors employ correlation with FreeSurfer thickness values as the primary evaluation for assessing relative performance with ANTs cortical thickness¹⁶. This evaluation, unfortunately, is fundamentally flawed in that it is a prime example of a type of circularity analysis⁴⁶ whereby data selection is driven by the same criteria used to evaluate performance. Specifically, the underlying DeepSCAN network used for the tissue segmentation step employs training based on FreeSurfer results which directly influences thickness values as thickness/segmentation are highly correlated and vary characteristically between software packages. Relative performance with ANTs thickness (which does not use FreeSurfer for training) is then assessed by determining correlations with FreeSurfer thickness values. Almost as problematic is their

¹⁸¹ use of repeatability, which they confusingly label as “robustness,” as an additional ranking
¹⁸² criterion. Repeatability evaluations should be contextualized within considerations such
¹⁸³ as the bias-variance tradeoff and quantified using relevant metrics, such as the intra-class
¹⁸⁴ correlation coefficient which takes into account both inter- and intra-observer variability.

¹⁸⁵ In addition to the training data listed above, to ensure generalizability, we also compared
¹⁸⁶ performance using the SRPB data set⁴⁷ comprising over 1600 participants from 12 sites. Note
¹⁸⁷ that we recognize that we are processing a portion of the evaluation data through certain
¹⁸⁸ components of the proposed deep learning-based pipeline that were used to train the same
¹⁸⁹ pipeline components. Although this does not provide evidence for generalizability (which is
¹⁹⁰ why we include the much larger SRPB data set), it is still interesting to examine the results
¹⁹¹ since, in this case, the deep learning training can be considered a type of noise reduction on
¹⁹² the final results. It should be noted that training did not use age prediction (or any other
¹⁹³ evaluation or related measure) as a criterion to be optimized during network model training
¹⁹⁴ (i.e., circular analysis⁴⁶).

¹⁹⁵ The results are shown in Figure 3 where we used cross-validation with 500 permutations
¹⁹⁶ per model per data set (including a “combined” set) and an 80/20 training/testing split.
¹⁹⁷ The ANTsXNet deep learning pipeline outperformed the classical pipeline¹⁶ in terms of age
¹⁹⁸ prediction in all data sets except for IXI. This also includes the cross-validation iteration
¹⁹⁹ where all data sets were combined. Importance plots ranking the cortical thickness regions
²⁰⁰ and the other covariates of Equation (1) are shown in Figure 4. Rankings employ “MeanDe-
²⁰¹ creaseAccuracy” which quantifies the decrease in model accuracy based on the exclusion of a
²⁰² specific random forest regressor. Additionally, repeatability assessment on the MMRR data
²⁰³ set yielded ICC values (“average random rater”) of 0.99 for both pipelines.

²⁰⁴ Longitudinal performance evaluation

²⁰⁵ Given the excellent performance and superior computational efficiency of the proposed
²⁰⁶ ANTsXNet pipeline for cross-sectional data, we evaluated its performance on longitudinal
²⁰⁷ data using the longitudinally-specific evaluation strategy and data we employed with the
²⁰⁸ introduction of the longitudinal version of the ANTs cortical thickness pipeline²⁹. We also

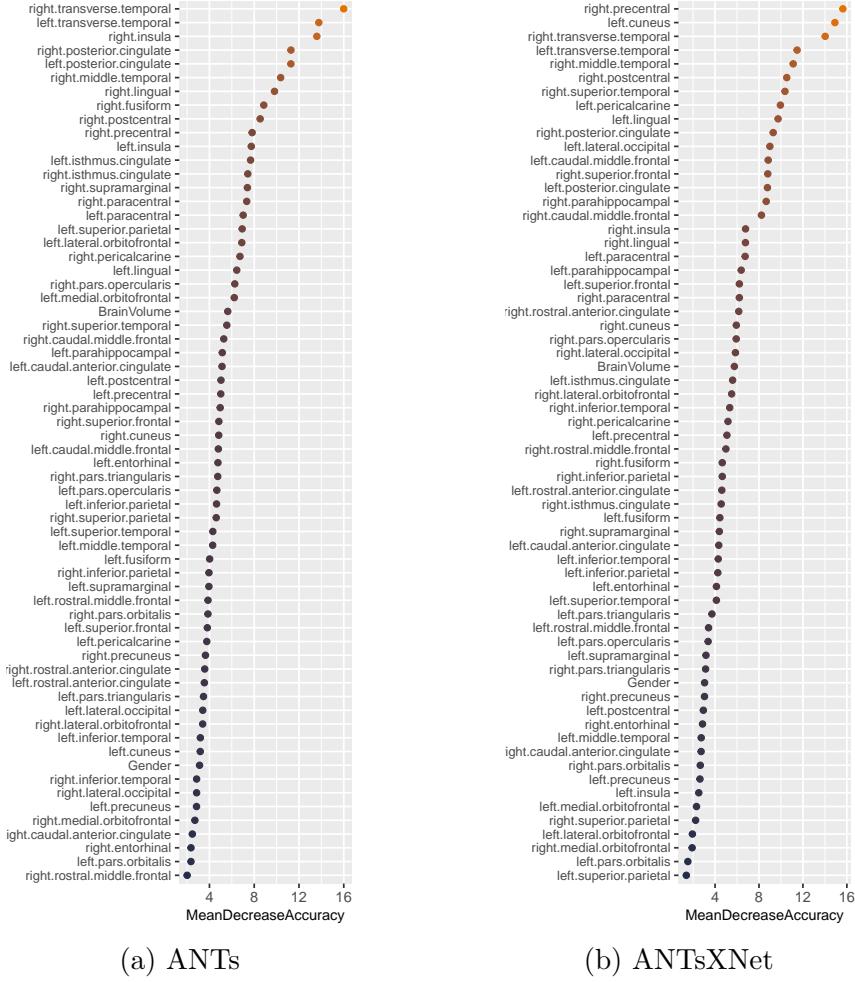


Figure 4: Importance plots for the SRPB data set using “MeanDecreaseAccuracy” for the random forest regressors (i.e., cortical thickness regions, gender, and brain volume specified by Equation (1).

evaluated an ANTsXNet-based pipeline tailored specifically for longitudinal data. In this variant, an SST is generated and processed using the previously described ANTsXNet cross-sectional pipeline which yields tissue spatial priors. These spatial priors are used in our traditional brain segmentation approach¹⁰. The computational efficiency of this variant is also significantly improved due to the elimination of the costly SST prior generation which uses multiple registrations combined with joint label fusion¹⁴.

The ADNI-1 data used for our previous longitudinal performance evaluation²⁹ consisted of over 600 subjects (197 cognitive normals, 324 LMCI subjects, and 142 AD subjects) with one or more follow-up image acquisition sessions every 6 months (up to 36 months) for a

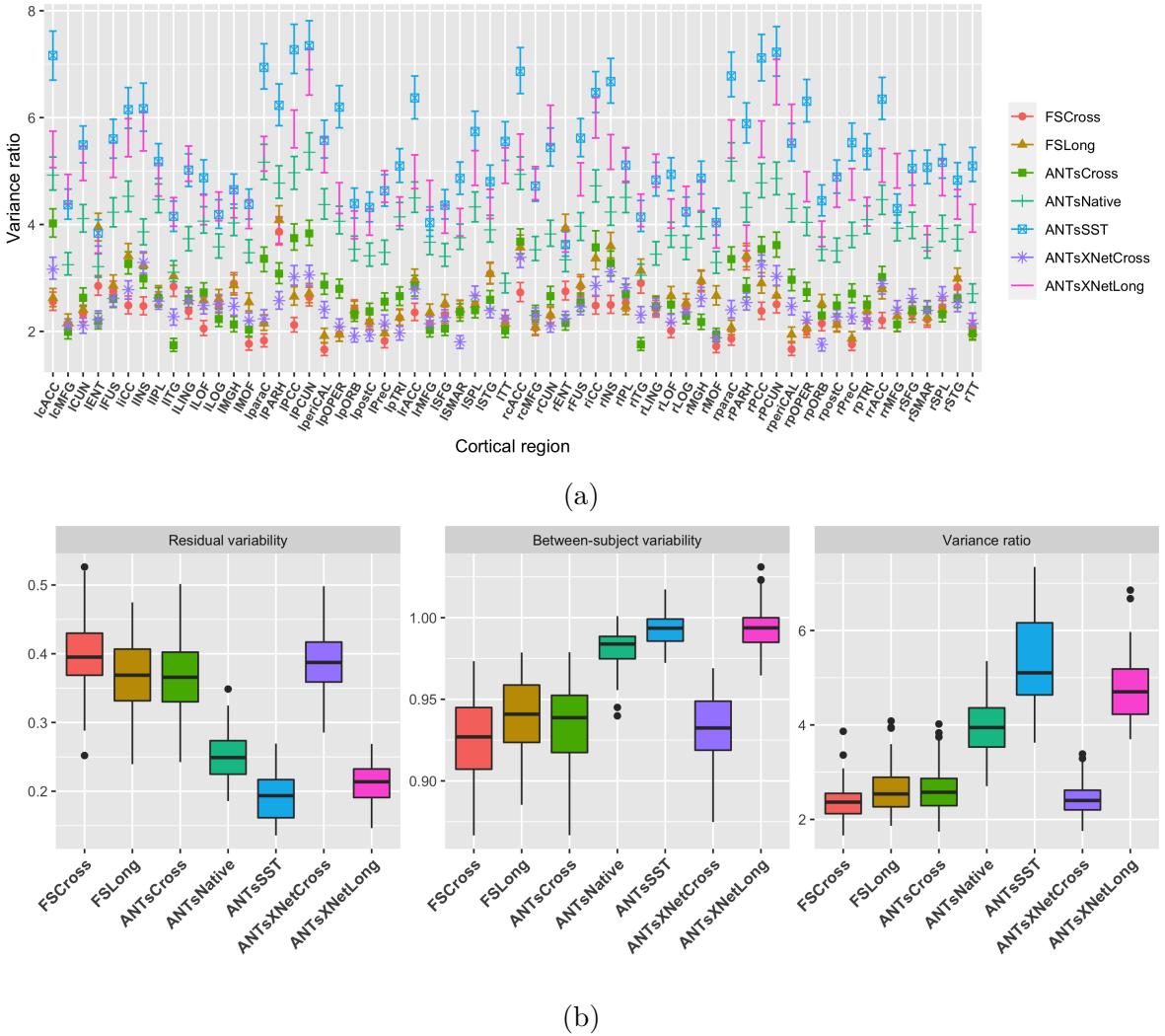


Figure 5: Performance over longitudinal data as determined by the variance ratio. (a) Region-specific 95% confidence intervals of the variance ratio showing the superior performance of the longitudinally tailored ANTsX-based pipelines, including ANTsSST and ANTsXNetLong. (b) Residual variability, between subject, and variance ratio values per pipeline over all DKT regions.

total of over 2500 images. In addition to the ANTsXNet pipelines (“[ANTsXNetCross](#)” and “[ANTsXNetLong](#)”) for the current evaluation, our previous work included the FreeSurfer²³ cross-sectional (“FSCross”) and longitudinal (“FSLong”) streams, the ANTs cross-sectional pipeline (“ANTsCross”) in addition to two longitudinal ANTs-based variants (“ANTsNative” and “ANTsSST”). Two evaluation measurements, one unsupervised and one supervised, were used to assess comparative performance between all five pipelines. We add the results of the ANTsXNet pipeline [cross-sectional and longitudinal](#) evaluations in relation to these other

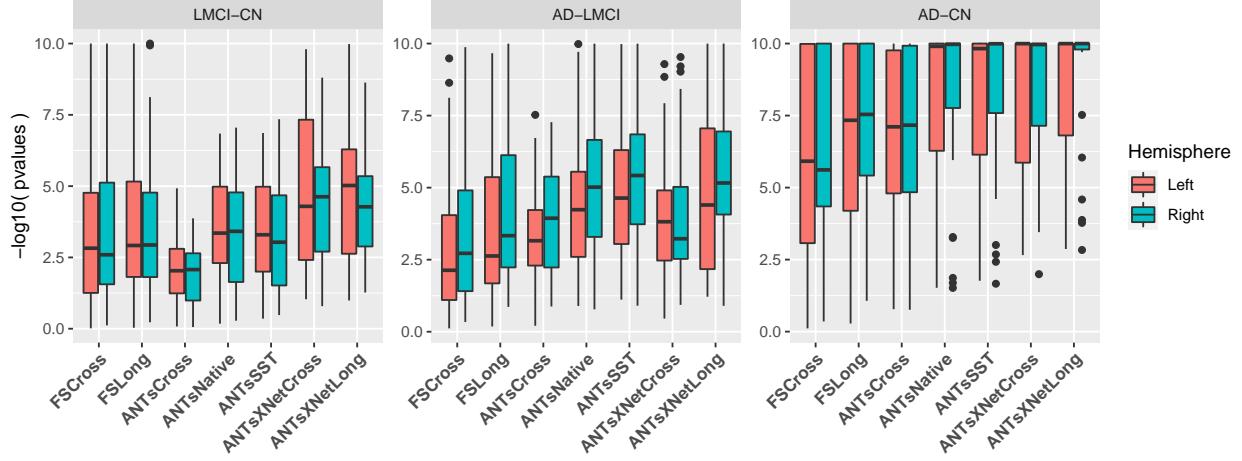


Figure 6: Measures for the both the supervised and unsupervised evaluation strategies, respectively given in (a) and (b). (a) Log p-values for diagnostic differentiation of LMCI-CN, AD-LMCI, and AD-CN subjects for all pipelines over all DKT regions. (b) Residual variability, between subject, and variance ratio values per pipeline over all DKT regions.

225 pipelines to provide a comprehensive overview of relative performance.

First, the supervised evaluation employed Tukey post-hoc analyses with false discovery rate (FDR) adjustment to test the significance of the LMCI-CN, AD-LMCI, and AD-CN diagnostic contrasts. This is provided by the following [linear mixed-effects](#) LME model

$$\begin{aligned} \Delta Y \sim & Y_{bl} + AGE_{bl} + ICV_{bl} + APOE_{bl} + GENDER + DIAGNOSIS_{bl} \\ & + VISIT : DIAGNOSIS_{bl} + (1|ID) + (1|SITE). \end{aligned} \quad (2)$$

226 Here, ΔY is the change in thickness of the k^{th} DKT region from baseline (bl) thickness
 227 Y_{bl} with random intercepts for both the individual subject (ID) and the acquisition site.
 228 The subject-specific covariates AGE , $APOE$ status, $GENDER$, $DIAGNOSIS$, ICV , and
 229 $VISIT$ were taken directly from the ADNIMERGE package.

Second, LME⁴⁸ modeling was used to quantify between-subject and residual variabilities, the ratio of which provides an estimate of the effectiveness of a given biomarker for distinguishing between subpopulations. In order to assess this criteria while accounting for changes that

may occur through the passage of time, we used the following Bayesian LME model:

$$\begin{aligned} Y_{ij}^k &\sim N(\alpha_i^k + \beta_i^k t, \sigma_k^2) \\ \alpha_i^k &\sim N(\alpha_0^k, \tau_k^2) \quad \beta_i^k \sim N(\beta_0^k, \rho_k^2) \\ \alpha_0^k, \beta_0^k &\sim N(0, 10) \quad \sigma_k, \tau_k, \rho_k \sim \text{Cauchy}^+(0, 5) \end{aligned} \tag{3}$$

where Y_{ij}^k denotes the i^{th} individual's cortical thickness measurement corresponding to the k^{th} region of interest at the time point indexed by j and specification of variance priors to half-Cauchy distributions reflects commonly accepted best practice in the context of hierarchical models⁴⁹. The ratio of interest, r^k , per region of the between-subject variability, τ_k , and residual variability, σ_k is

$$r^k = \frac{\tau_k}{\sigma_k}, k = 1, \dots, 62 \tag{4}$$

where the posterior distribution of r_k was summarized via the posterior median.
Results for both longitudinal evaluation scenarios are shown in Figure 6. Log p-values are provided in Figure 6(a) which demonstrate excellent LMCI-CN and AD-CN differentiation and comparable AD-LMCI differentiation relative to the other pipelines. Figure 6(b) shows significantly better performance for the longitudinal ANTsXNet pipeline where, in a longitudinal setting, we prefer to see lower values for residual variability and higher values for between-subject variability, leading to a larger variance ratio. In contrast, cross-sectional ANTsXNet performs remarkably poorly for these measures.

Discussion

The ANTsX software ecosystem provides a comprehensive framework for quantitative biological and medical imaging. Although ANTs, the original core of ANTsX, is still at the forefront of image registration technology, it has moved significantly beyond its image registration origins. This expansion is not confined to technical contributions (of which there are many) but also consists of facilitating access to a wide range of users who can use ANTsX tools

²⁴⁴ (whether through bash, Python, or R scripting) to construct tailored pipelines for their own
²⁴⁵ studies or to take advantage of our pre-fabricated pipelines. And given the open-source
²⁴⁶ nature of the ANTsX software, usage is not limited, for example, to academic institutions—a
²⁴⁷ common constraint characteristic of other packages.

²⁴⁸ One of our most widely used pipelines is the estimation of cortical thickness from neuroimag-
²⁴⁹ ing. This is understandable given the widespread usage of regional cortical thickness as a
²⁵⁰ biomarker for developmental or pathological trajectories of the brain. In this work, we used
²⁵¹ this well-vetted ANTs tool to provide training data for producing alternative variants which
²⁵² leverage deep learning for improved computational efficiency and also provides superior perfor-
²⁵³ mance with respect to previously proposed evaluation measures for both cross-sectional¹⁶ and
²⁵⁴ longitudinal scenarios²⁹. In addition to providing the tools which generated the original train-
²⁵⁵ ing data for the proposed ANTsXNet pipeline, the ANTsX ecosystem provides a full-featured
²⁵⁶ platform for the additional steps such as preprocessing (ANTsR/ANTsPy); data augmenta-
²⁵⁷ tion (ANTsR/ANTsPy); network construction and training (ANTsRNet/ANTsPyNet); and
²⁵⁸ visualization and statistical analysis of the results (ANTsR/ANTsPy).

²⁵⁹ It is the comprehensiveness of ANTsX that provides significant advantages over much of the
²⁶⁰ deep learning work that is currently taking place in medical imaging. In other words, various
²⁶¹ steps in the deep learning training processing (e.g., data augmentation, preprocessing) can all
²⁶² be performed within the same ecosystem where such important details as header information
²⁶³ for image geometry are treated the same. In contrast, related work³¹ described and evaluated
²⁶⁴ a similar thickness measurement pipeline. However, due to the lack of a complete processing
²⁶⁵ and analysis framework, training data was generated using the FreeSurfer stream, deep
²⁶⁶ learning-based brain segmentation employed DeepSCAN⁵⁰ (in-house software), and cortical
²⁶⁷ thickness estimation¹⁵ was generated using the ANTs toolkit. For the reader interested in
²⁶⁸ reproducing the authors' results, they are primarily prevented from doing so due, as far as
²⁶⁹ we can tell, to the lack of the public availability of the DeepSCAN software. However, in
²⁷⁰ addition, the interested reader must also ensure the consistency of the input/output interface
²⁷¹ between packages (a task for which the Nipype development team is quite familiar.)

272 In terms of future work, the recent surge and utility of deep learning in medical image analysis
273 has significantly guided the areas of active ANTsX development. As demonstrated in this
274 work with our widely used cortical thickness pipelines, there are many potential benefits
275 of deep learning analogs to existing ANTs tools as well as the development of new ones.
276 Performance is [mostly](#) comparable-to-superior relative to existing pipelines depending on
277 the evaluation metric. [Specifically, the ANTsXNet cross-sectional pipeline does well for the](#)
278 [age prediction performance framework and in terms of the ICC.](#) Additionally, this pipeline
279 performs relatively well for longitudinal ADNI data for disease differentiation but not so
280 much in terms of the generic variance ratio criterion. However, for such longitudinal-specific
281 studies, the ANTsXNet longitudinal variant performs well for both performance measures.
282 We see possible additional longitudinal extensions incorporating subject ID and months as
283 additional network inputs.

284 Methods

285 Software, average DKT regional thickness values for all data sets, and the scripts to perform
286 both the analysis and obtain thickness values for a single subject ([cross-sectionally or](#)
287 [longitudinally](#)) are provided as open-source. Specifically, all the ANTsX libraries are hosted
288 on GitHub (<https://github.com/ANTsX>). The cross-sectional data and analysis code are
289 available as .csv files and R scripts at the GitHub repository dedicated to this paper (<https://github.com/ntustison/PaperANTsX>) whereas the longitudinal data and evaluation scripts
290 are organized with the repository associated with our previous work²⁹ (<https://github.com/ntustison/CrossLong>).

293 Implementation

```
294 import ants
295 import antspynet
296
297
298 # ANTsPy/ANTsPyNet processing for subject IXI002-Guys-0828-T1
299 t1_file = "IXI002-Guys-0828-T1.nii.gz"
300 t1 = ants.image_read(t1_file)
301
302 # Atropos six-tissue segmentation
303 atropos = antspynet.deep_atropos(t1, do_preprocessing=True, verbose=True)
```

```

304
305 # Kelly Kapowski cortical thickness (combine Atropos WM and deep GM)
306 kk_segmentation = atropos['segmentation_image']
307 kk_segmentation[kk_segmentation == 4] = 3
308 kk_gray_matter = atropos['probability_images'][2]
309 kk_white_matter = atropos['probability_images'][3] + atropos['probability_images'][4]
310 kk = ants.kelly_kapowski(s=kk_segmentation, g=kk_gray_matter, w=kk_white_matter,
311                         its=45, r=0.025, m=1.5, x=0, verbose=1)
312
313 # Desikan-Killiany-Tourville labeling
314 dkt = antspynet.desikan_killiany_tourville_labeling(t1, do_preprocessing=True, verbose=True)
315
316 # DKT label propagation throughout the cortex
317 dkt_cortical_mask = ants.threshold_image(dkt, 1000, 3000, 1, 0)
318 dkt = dkt_cortical_mask * dkt
319 kk_mask = ants.threshold_image(kk, 0, 0, 0, 1)
320 dkt_propagated = ants.iMath(kk_mask, "PropagateLabelsThroughMask", kk_mask * dkt)
321
322 # Get average regional thickness values
323 kkRegional_stats = ants.label_stats(kk, dkt_propagated)

```

Listing 1: ANTsPy/ANTsPyNet command calls for a single IXI subject in the evaluation study for the cross-sectional pipeline.

In Listing 1, we show the ANTsPy/ANTsPyNet code snippet for cross-sectional processing a single subject which starts with reading the T1-weighted MRI input image, through the generation of the Atropos-style six-tissue segmentation and probability images, application of `ants.kelly_kapowski` (i.e., DiReCT), DKT cortical parcellation, subsequent label propagation through the cortex, and, finally, regional cortical thickness tabulation. The cross-sectional and longitudinal pipelines are encapsulated in the ANTsPyNet functions `antspynet.cortical_thickness` and `antspynet.longitudinal_cortical_thickness`, respectively. Note that there are precise, line-by-line R-based analogs available through ANTsR/ANTsRNet.

Both the `ants.deep_atropos` and `antspynet.desikan_killiany_tourville_labeling` functions perform brain extraction using the `antspynet.brain_extraction` function. Internally, `antspynet.brain_extraction` contains the requisite code to build the network and assign the appropriate hyperparameters. The model weights are automatically downloaded from the online hosting site <https://figshare.com> (see the function `get_pretrained_network` in ANTsPyNet or `getPretrainedNetwork` in ANTsRNet for links to all models and weights) and loaded to the constructed network. `antspynet.brain_extraction` performs a quick translation transformation to a specific template (also downloaded automatically) using the centers of intensity mass, a common alignment initialization strategy. This is to ensure

proper gross orientation. Following brain extraction, preprocessing for the other two deep learning components includes `ants.denoise_image` and `ants.n4_bias_correction` and an affine-based reorientation to a version of the MNI template⁵¹.

We recognize the presence of some redundancy due to the repeated application of certain preprocessing steps. Thus, each function has a `do_preprocessing` option to eliminate this redundancy for knowledgeable users but, for simplicity in presentation purposes, we do not provide this modified pipeline here. Although it should be noted that the time difference is minimal considering the longer time required by `ants.kelly_kapowski`. `ants.deep_atropos` returns the segmentation image as well as the posterior probability maps for each tissue type listed previously. `antspynet.desikan_killiany_tourville_labeling` returns only the segmentation label image which includes not only the 62 cortical labels but the remaining labels as well. The label numbers and corresponding structure names are given in the program description/help. Because the DKT parcellation will, in general, not exactly coincide with the non-zero voxels of the resulting cortical thickness maps, we perform a label propagation step to ensure the entire cortex, and only the non-zero thickness values in the cortex, are included in the tabulated regional values.

As mentioned previously, the longitudinal version, `antspynet.longitudinal_cortical_thickness`, adds an SST generation step which can either be provided as a program input or it can be constructed from spatial normalization of all time points to a specified template. `ants.deep_atropos` is applied to the SST yielding spatial tissues priors which are then used as input to `ants.atropos` for each time point. `ants.kelly_kapowski` is applied to the result to generate the desired cortical thickness maps.

Computational time on a CPU-only platform is approximately 1 hour primarily due to `ants.kelly_kapowski` processing. Other preprocessing steps, i.e., bias correction and denoising, are on the order of a couple minutes. This total time should be compared with 4 – 5 hours using the traditional pipeline employing the `quick` registration option or 10 – 15 hours with the more comprehensive registration parameters employed). As mentioned previously, elimination of the registration-based propagation of prior probability images to individual

371 subjects is the principal source of reduced computational time. For ROI-based analyses, this
372 is in addition to the elimination of the optional generation of a population-specific template.
373 Additionally, the use of `antspynet.desikan_killiany_tourville_labeling`, for cortical
374 labeling (which completes in less than five minutes) eliminates the need for joint label fusion
375 which requires multiple pairwise registrations for each subject in addition to the fusion
376 algorithm itself.

377 Training details

378 Training differed slightly between models and so we provide details for each of these com-
379 ponents below. For all training, we used ANTsRNet scripts and custom batch generators.
380 Although the network construction and other functionality is available in both ANTsPyNet
381 and ANTsRNet (as is model weights compatibility), we have not written such custom batch
382 generators for the former (although this is on our to-do list). In terms of hardware, all
383 training was done on a DGX (GPUs: 4X Tesla V100, system memory: 256 GB LRDIMM
384 DDR4).

385 **T1-weighted brain extraction.** A whole-image 3-D U-net model²⁴ was used in conjunction
386 with multiple training sessions employing a Dice loss function followed by categorical cross
387 entropy. Training data was derived from the same multi-site data described previously
388 processed through our registration-based approach³⁰. A center-of-mass-based transformation
389 to a standard template was used to standardize such parameters as orientation and voxel size.
390 However, to account for possible different header orientations of input data, a template-based
391 data augmentation scheme was used⁴¹ whereby forward and inverse transforms are used
392 to randomly warp batch images between members of the training population (followed by
393 reorientation to the standard template). A digital random coin flipping for possible histogram
394 matching⁵² between source and target images further increased data augmentation. The
395 output of the network is a probabilistic mask of the brain. Although not detailed here,
396 training for brain extraction in other modalities was performed similarly.

397 **Deep Atropos.** Dealing with 3-D data presents unique barriers for training that are often
398 unique to medical imaging. Various strategies are employed such as minimizing the number

399 of layers and/or the number of filters at the base layer of the U-net architecture (as we do
400 for brain extraction). However, we found this to be too limiting for capturing certain brain
401 structures such as the cortex. 2-D and 2.5-D approaches are often used with varying levels of
402 success but we also found better performance using full 3-D information. This led us to try
403 randomly selected 3-D patches of various sizes. However, for both the six-tissue segmentations
404 and DKT parcellations, we found that an octant-based patch strategy yielded the desired
405 results. Specifically, after a brain extracted affine normalization to the MNI template, the
406 normalized image is cropped to a size of [160, 190, 160]. Overlapping octant patches of size
407 [112, 112, 112] were extracted from each image and trained using a batch size of 12 such
408 octant patches with weighted categorical cross entropy as the loss function. As we point out
409 in our earlier work¹⁶, obtaining proper brain segmentation is perhaps the most critical step
410 to estimating thickness values that have the greatest utility as a potential biomarker. In fact,
411 the first and last authors (NT and BA, respectively) spent much time during the original
412 ANTs pipeline development¹⁶ trying to get the segmentation correct which required manually
413 looking at many images and manually adjusting where necessary. This fine-tuning is often
414 omitted or not considered when other groups^{31,53,54} use components of our cortical thickness
415 pipeline which can be potentially problematic⁵⁵. Fine-tuning for this particular workflow was
416 also performed between the first and last authors using manual variation of the weights in the
417 weighted categorical cross entropy. Specifically, the weights of each tissue type was altered in
418 order to produce segmentations which most resemble the traditional Atropos segmentations.
419 Ultimately, we settled on a weight vector of (0.05, 1.5, 1, 3, 4, 3, 3) for the CSF, GM, WM,
420 Deep GM, brain stem, and cerebellum, respectively. Other hyperparameters can be directly
421 inferred from explicit specification in the actual code. As mentioned previously, training
422 data was derived from application of the ANTs Atropos segmentation¹⁰ during the course of
423 our previous work¹⁶. Data augmentation included small affine and deformable perturbations
424 using `antspynet.randomly_transform_image_data` and random contralateral flips.

425 **Desikan-Killiany-Tourville parcellation.** Preprocessing for the DKT parcellation train-
426 ing was similar to the Deep Atropos training. However, the number of labels and the
427 complexity of the parcellation required deviation from other training steps. First, labeling

428 was split into an inner set and an outer set. Subsequent training was performed separately
429 for both of these sets. For the cortical labels, a set of corresponding input prior probability
430 maps were constructed from the training data (and are also available and automatically
431 downloaded, when needed, from <https://figshare.com>). Training occurred over multiple
432 sessions where, initially, categorical cross entropy was used and then subsequently refined
433 using a Dice loss function. Whole-brain training was performed on a brain-cropped template
434 size of [96, 112, 96]. Inner label training was performed similarly to our brain extraction
435 training where the number of layers at the base layer was reduced to eight. Training also
436 occurred over multiple sessions where, initially, categorical cross entropy was used and then
437 subsequently refined using a Dice loss function. Other hyperparameters can be directly
438 inferred from explicit specification in the actual code. Training data was derived from
439 application of joint label fusion¹³ during the course of our previous work¹⁶. When call-
440 ing `antspynet.desikan_killiany_tourville_labeling`, inner labels are estimated first
441 followed by the outer, cortical labels.

⁴⁴² **Acknowledgments**

⁴⁴³ Data collection and sharing for this project was funded by the Alzheimer's Disease Neu-
⁴⁴⁴ roimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD
⁴⁴⁵ ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the
⁴⁴⁶ National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering,
⁴⁴⁷ and through generous contributions from the following: AbbVie, Alzheimer's Association;
⁴⁴⁸ Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-
⁴⁴⁹ Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.;
⁴⁵⁰ Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company
⁴⁵¹ Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy
⁴⁵² Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development
⁴⁵³ LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx
⁴⁵⁴ Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Pira-
⁴⁵⁵ mal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The
⁴⁵⁶ Canadian Institutes of Health Research is providing funds to support ADNI clinical sites
⁴⁵⁷ in Canada. Private sector contributions are facilitated by the Foundation for the National
⁴⁵⁸ Institutes of Health (www.fnih.org). The grantee organization is the Northern California
⁴⁵⁹ Institute for Research and Education, and the study is coordinated by the Alzheimer's
⁴⁶⁰ Therapeutic Research Institute at the University of Southern California. ADNI data are
⁴⁶¹ disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

462 **References**

- 463 1. Bajcsy, R. & Broit, C. Matching of deformed images. in *Sixth International Conference on*
464 *Pattern Recognition (ICPR'82)* 351–353 (1982).
- 465 2. Bajcsy, R. & Kovacic, S. Multiresolution elastic matching. *Computer Vision, Graphics,*
466 *and Image Processing* **46**, 1–21 (1989).
- 467 3. Gee, J., Sundaram, T., Hasegawa, I., Uematsu, H. & Hatabu, H. Characterization of
468 regional pulmonary mechanics from serial magnetic resonance imaging data. *Acad Radiol* **10**,
469 1147–52 (2003).
- 470 4. Klein, A. *et al.* Evaluation of 14 nonlinear deformation algorithms applied to human brain
471 MRI registration. *Neuroimage* **46**, 786–802 (2009).
- 472 5. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic
473 image registration with cross-correlation: Evaluating automated labeling of elderly and
474 neurodegenerative brain. *Med Image Anal* **12**, 26–41 (2008).
- 475 6. Murphy, K. *et al.* Evaluation of registration methods on thoracic CT: The EMPIRE10
476 challenge. *IEEE Trans Med Imaging* **30**, 1901–20 (2011).
- 477 7. Menze, B., Reyes, M. & Van Leemput, K. The multimodal brain tumor image segmentation
478 benchmark (BRATS). *IEEE Trans Med Imaging* (2014) doi:[10.1109/TMI.2014.2377694](https://doi.org/10.1109/TMI.2014.2377694).
- 479 8. Tustison, N. J., Avants, B. B. & Gee, J. C. Learning image-based spatial transformations
480 via convolutional neural networks: A review. *Magn Reson Imaging* **64**, 142–153 (2019).
- 481 9. Avants, B. B. *et al.* The optimal template effect in hippocampus studies of diseased
482 populations. *Neuroimage* **49**, 2457–66 (2010).
- 483 10. Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A. & Gee, J. C. An open source multivariate
484 framework for n -tissue segmentation with evaluation on public data. *Neuroinformatics* **9**,
485 381–400 (2011).

- 486 11. Tustison, N. J. & Gee, J. C. N4ITK: Nick's N3 ITK implementation for MRI bias field
487 correction. *The Insight Journal* (2009).
- 488 12. Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L. & Robles, M. Adaptive
489 non-local means denoising of MR images with spatially varying noise levels. *J Magn Reson*
490 *Imaging* **31**, 192–203 (2010).
- 491 13. Wang, H. & Yushkevich, P. A. Multi-atlas segmentation with joint label fusion and
492 corrective learning—an open source implementation. *Front Neuroinform* **7**, 27 (2013).
- 493 14. Wang, H. *et al.* Multi-atlas segmentation with joint label fusion. *IEEE Trans Pattern*
494 *Anal Mach Intell* **35**, 611–23 (2013).
- 495 15. Das, S. R., Avants, B. B., Grossman, M. & Gee, J. C. Registration based cortical thickness
496 measurement. *Neuroimage* **45**, 867–79 (2009).
- 497 16. Tustison, N. J. *et al.* Large-scale evaluation of ANTs and FreeSurfer cortical thickness
498 measurements. *Neuroimage* **99**, 166–79 (2014).
- 499 17. Esteban, O. *et al.* FMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat*
500 *Methods* **16**, 111–116 (2019).
- 501 18. De Leener, B. *et al.* SCT: Spinal cord toolbox, an open-source software for processing
502 spinal cord MRI data. *Neuroimage* **145**, 24–43 (2017).
- 503 19. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and
504 describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044 (2016).
- 505 20. Halchenko, Y. O. & Hanke, M. Open is not enough. Let's take the next step: An
506 integrated, community-driven computing platform for neuroscience. *Front Neuroinform* **6**, 22
507 (2012).
- 508 21. Muschelli, J. *et al.* Neuroconductor: An R platform for medical imaging analysis.
509 *Biostatistics* **20**, 218–239 (2019).

- 510 22. Gorgolewski, K. *et al.* Nipype: A flexible, lightweight and extensible neuroimaging data
511 processing framework in python. *Front Neuroinform* **5**, 13 (2011).
- 512 23. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–81 (2012).
- 513 24. Falk, T. *et al.* U-net: Deep learning for cell counting, detection, and morphometry. *Nat
514 Methods* **16**, 67–70 (2019).
- 515 25. Bashyam, V. M. *et al.* MRI signatures of brain age and disease over the lifespan based
516 on a deep brain network and 14,468 individuals worldwide. *Brain* **143**, 2312–2324 (2020).
- 517 26. Goubran, M. *et al.* Hippocampal segmentation for brains with extensive atrophy using
518 three-dimensional convolutional neural networks. *Hum Brain Mapp* **41**, 291–308 (2020).
- 519 27. Li, H. *et al.* Fully convolutional network ensembles for white matter hyperintensities
520 segmentation in mr images. *Neuroimage* **183**, 650–665 (2018).
- 521 28. Haris, M., Shakhnarovich, G. & Ukita, N. Deep back-projection networks for super-
522 resolution. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*
523 1664–1673 (2018). doi:[10.1109/CVPR.2018.00179](https://doi.org/10.1109/CVPR.2018.00179).
- 524 29. Tustison, N. J. *et al.* Longitudinal mapping of cortical thickness measurements: An
525 Alzheimer's Disease Neuroimaging Initiative-based evaluation study. *J Alzheimers Dis* (2019)
526 doi:[10.3233/JAD-190283](https://doi.org/10.3233/JAD-190283).
- 527 30. Avants, B. B., Klein, A., Tustison, N. J., Woo, J. & Gee, J. C. Evaluation of open-access,
528 automated brain extraction methods on multi-site multi-disorder data. in *16th annual meeting
529 for the organization of human brain mapping* (2010).
- 530 31. Rebsamen, M., Rummel, C., Reyes, M., Wiest, R. & McKinley, R. Direct cortical
531 thickness estimation using deep learning-based anatomy segmentation and cortex parcellation.
532 *Hum Brain Mapp* (2020) doi:[10.1002/hbm.25159](https://doi.org/10.1002/hbm.25159).
- 533 32. Henschel, L. *et al.* FastSurfer - a fast and accurate deep learning based neuroimaging
534 pipeline. *Neuroimage* **219**, 117012 (2020).

- 535 33. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging*
536 **29**, 1310–20 (2010).
- 537 34. Ashburner, J. & Friston, K. J. Voxel-based morphometry—the methods. *Neuroimage* **11**,
538 805–21 (2000).
- 539 35. Avants, B. *et al.* Eigenanatomy improves detection power for longitudinal cortical change.
540 *Med Image Comput Comput Assist Interv* **15**, 206–13 (2012).
- 541 36. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical
542 labeling protocol. *Front Neurosci* **6**, 171 (2012).
- 543 37. <https://brain-development.org/ixi-dataset/>.
- 544 38. Landman, B. A. *et al.* Multi-parametric neuroimaging reproducibility: A 3-T resource
545 study. *Neuroimage* **54**, 2854–66 (2011).
- 546 39. http://fcon_1000.projects.nitrc.org/indi/pro/nki.html.
- 547 40. <https://www.oasis-brains.org>.
- 548 41. Tustison, N. J. *et al.* Convolutional neural networks with template-based data augmenta-
549 tion for functional lung image quantification. *Acad Radiol* **26**, 412–423 (2019).
- 550 42. Schlemper, J. *et al.* Attention gated networks: Learning to leverage salient regions in
551 medical images. *Med Image Anal* **53**, 197–207 (2019).
- 552 43. Lemaitre, H. *et al.* Normal age-related brain morphometric changes: Nonuniformity
553 across cortical thickness, surface area and gray matter volume? *Neurobiol Aging* **33**, 617.e1–9
554 (2012).
- 555 44. Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).
- 556 45. Holbrook, A. J. *et al.* Anterolateral entorhinal cortex thickness as a new biomarker for
557 early detection of Alzheimer’s disease. *Alzheimer’s & Dementia: Diagnosis, Assessment &*
558 *Disease Monitoring* **12**, e12068 (2020).

- 559 46. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis
560 in systems neuroscience: The dangers of double dipping. *Nat Neurosci* **12**, 535–40 (2009).
- 561 47. <https://bicr-resource.atr.jp/srpbs1600/>.
- 562 48. Verbeke, G. Linear mixed models for longitudinal data. in *Linear mixed models in practice*
563 63–153 (Springer, 1997).
- 564 49. Gelman, A. & others. Prior distributions for variance parameters in hierarchical models
565 (comment on article by Browne and Draper). *Bayesian analysis* **1**, 515–534 (2006).
- 566 50. McKinley, R. *et al.* Few-shot brain segmentation from weakly labeled data with deep
567 heteroscedastic multi-task networks. *CoRR* **abs/1904.02436**, (2019).
- 568 51. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almlí, C. & Collins, D. L. Unbiased
569 nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*
570 **S102**, (2009).
- 571 52. Nyúl, L. G. & Udupa, J. K. On standardizing the MR image intensity scale. *Magn Reson*
572 *Med* **42**, 1072–81 (1999).
- 573 53. Clarkson, M. J. *et al.* A comparison of voxel and surface based cortical thickness
574 estimation methods. *Neuroimage* **57**, 856–65 (2011).
- 575 54. Schwarz, C. G. *et al.* A large-scale comparison of cortical thickness and volume methods
576 for measuring alzheimer’s disease severity. *Neuroimage Clin* **11**, 802–812 (2016).
- 577 55. Tustison, N. J. *et al.* Instrumentation bias in the use and evaluation of scientific software:
578 Recommendations for reproducible practices in the computational sciences. *Front Neurosci*
579 **7**, 162 (2013).

⁵⁸⁰ **Author contributions**

⁵⁸¹ Conception and design N.T., A.H., M.Y., J.S., B.A. Analysis and interpretation N.T., A.H.,
⁵⁸² D.G., M.Y., J.S. B.A. Creation of new software N.T., P.C., H.J., J.M., G.D., J.D., S.D., N.C.,
⁵⁸³ J.G., B.A. Drafting of manuscript N.T., A.H., P.C., H.J., J.M., G.D., J.G., B.A.

⁵⁸⁴ **Competing interests**

⁵⁸⁵ The authors declare no competing interests.