Deep-Learning-Based Age Estimation From Brain MR Images

Nyeli Kratz

Abstract—**Brain age estimation could aid in the diagnosis of neurodegenerative diseases. This paper presents an algorithm which predicts age based on brain MR images. A description of the data set is presented, followed by an overview of prior work done on brain MRI segmentation and age estimation. The paper then discusses our algorithm which uses multi-atlas segmentation followed by a deep learning network to predict age. Evaluation experiments are discussed followed by a summary of the contributions by each team member. The team was able to create an algorithm that predicts age based on brain region volume with a final testing RMSE of 15.73 years.**

I. INTRODUCTION

Over the past few decades, enormous progress has been made in imaging brain injury and anatomy through MRI. Prior research has shown structural changes in human brain MRIs with chronological age [1]. Based on these structural changes, estimation of age from brain MRIs has attracted more attention in recent years. Brain age estimation could aid in the early diagnosis of neurogenerative diseases such as Alzheimer's, Parkinson's, and Multiple Sclerosis [2]. Traditionally, segmentation of brain MR images is done manually, which results in some variation due to personal opinion. Additionally, manual brain MRI segmentation requires considerable time and effort and is thus impractical for large data sets [2]. Therefore, the team developed an algorithm which uses registration to apply the manual segmentations of a few volumes to many other volumes. In this paper, I present our fully-automated algorithm which segments 3D brain MR images and reports estimated age based on the segmentation and other features extracted from the MRI.

II. BACKGROUND AND THEORY

A. Dataset Description

We have been provided with a set of atlases containing 40 volumes in the form of T1-weighted images which were registered to the MN1152 template with a resolution of 0.8mm and preprocessed with N4 inhomogeneity correction and white matter peak normalization. We have also been provided with the corresponding manual delineations of these volumes segmented in 207 subregions represented by different integer labels. Figure 1 shows the T1-weighted axial view of the MR image next to its corresponding manual delineation. We have also been provided with a training set of 500 other volumes in the form of T1-weighted MR images registered to MN1152 template with a resolution of 1 mm along with a .csv file containing the age of the patient for each MRI ranging from 19 to 86 years old which corresponds to each of the training volumes.

B. Prior Work on Brain MRI Segmentation

Various models have been proposed for the segmentation of brain MR images including Markov random field (MRF) models, multimodal segmentation, deep learning methods, and multi-atlas segmentation. Unlike other imaging applications where convolutional neural networks (CNNs) have largely outperformed other segmentation methods with the increase in computer processing capabilities in the past few decades, CNNs have remained comparatively less successful in brain MRI segmentation. This is likely due to the fact that manual delineation of whole-brain volumes is very time consuming and requires significant expertise, leading to a small training set being available for this data. Multi-atlas has often outperformed deep-learning-based approaches in brain MRI segmentation [3].
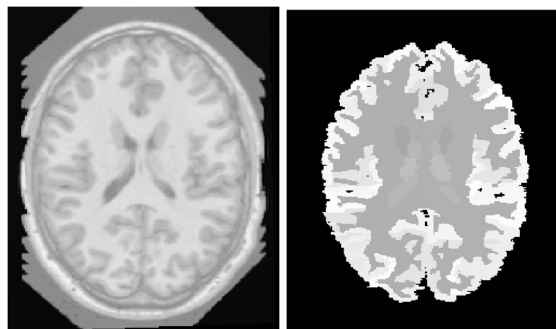

Fig. 1. T1-weighted axial view of brain MRI raw data (left) and manually delineated data (right).

C. Prior Work on Brain Age Estimation

In regards to deep learning algorithms, both classification and regression models have been explored extensively for brain age estimation based on MRI data with significant success. Shallow learning methods have been used such as Gaussian processes regression (GPR) and Support vector regression (SVR) [4]. Other strategies such as Hidden Markov Models and Random Forests have also been used with significant success. The team moved forward with a neural network regression method of brain age estimation based on the success of [5] and also due to the team's experience building similar models.

III. METHODS

A. Skull-stripping Method

The team developed a novel method for skull-stripping which achieved satisfactory results on about 70% of the training

data. This method binarizes the image based on a low starting threshold intensity and then calculates the number of connected components in the image. This initial number of connected components is always one due to the starting threshold intensity being only slightly higher than the background intensity. The threshold intensity then increments until there are two different connected components in the image. It is assumed that one of these connected components represents the skull while the other represents the brain because in most images in the data set, a dark volume of CSF separates the brain from the skull and is thus the first part of the anatomy to fall below the binarizing threshold.

The connected components were labelled using bwlabeln which allowed the outermost white voxel group to be identified as the skull and removed from the image. A spherical structuring element is then created and used to dilate the brain mask to avoid losing brain image information in images where the intensities of the edges of the brain are very similar to that of the CSF.

This method relies on the assumption that the CSF is darker than all other tissue inside of the skull and separates the brain from the skull at all points in 3D space. For some images in the training set this was not the case, and thus our method performed very poorly for some images. For this reason, the HD-BET package was used to strip the skulls from the images which were fed into the registration.

B. Registration

Splitting the atlases into random sets prior to multi-atlas segmentation has been shown to outperform using all available atlases for each segmentation [6]. Therefore, of the 40 provided label atlases, six random atlases were chosen for multi-atlas segmentation. These six atlases were registered to the skull-stripped image using an affine transformation. Six atlases were chosen in order to manage the tradeoff between computation time and segmentation results. The affine transformation is a time-consuming process and the registration required about 13 minutes with 6 affine transformations. It was decided that any extra accuracy afforded by additional atlas registration was not worth the cost of computational time.
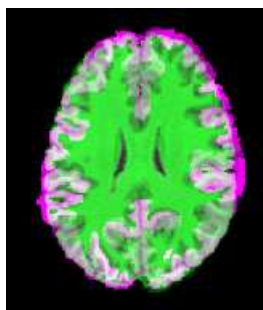

Fig. 2. Skull-stripped MR Image overlaid on atlas registered to the image using an affine transformation.

Alternatively, the team considered utilizing the voxelmorph package for registration which would have required significant training time, but would have registered images significantly faster after it was trained. Additionally, the voxelmorph package is based on diffeomorphic rather than affine registration, which has been shown to be more effective in inter-subject registration [7]. Ultimately, the team decided not to pursue the voxelmorph package due to time constraints and a lack of experience with similar models. It was decided that a good method for label fusion could make up for any inaccuracies in affine registration.

C. Label Fusion

Majority voting is a simple strategy and has shown to be very effective in brain multi-atlas segmentation [8]. For this reason, majority voting from each of the 6 atlases was utilized to label each voxel of the input image, resulting in a segmented volume. Figure 3 shows an example output segmentation from this algorithm.
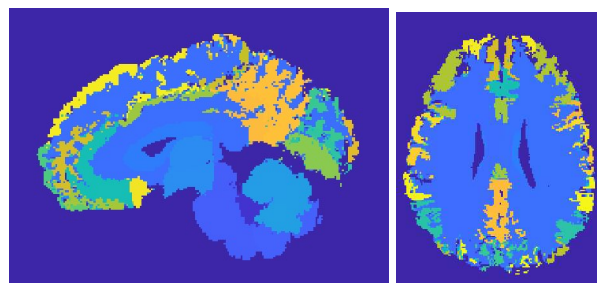

Fig. 3. Sagittal (left) and axial (right) segmentation example produced by this algorithm.

The team also considered a weighted voting label fusion method where the atlases which are most similar to the image get more votes than those which are less similar. The team briefly implemented this strategy using the sum of squared difference as a similarity metric and giving the most similar atlas 5 votes, the next most similar 4 votes, the next most similar 3 votes and the other atlases only one vote each for each voxel. In our initial experiments, the segmentation results did not improve enough to justify the cost of computation time. Had the team been able to reduce the computation time of the registration, this strategy may have been more feasible. With more time, the team would have adjusted the weighting and pursued this approach more thoroughly.

D. Age Estimation

The age estimation network is based on prior research which shows that brain age is associated with the volume of specific regions of the brain [1]. For instance, because the hippocampus is related to memorization, one might expect to see a decreased hippocampus volume in older brains.

Gray matter and white matter volume were extracted from the T1-weighted MRI because these features have proven useful in brain age estimation in previous works [5]. These features were extracted directly from the raw MRI data based on tissue intensity to avoid the deformation which is introduced by the segmentation method. The gray matter and white matter volume, along with volumes of each of the 207 regions labeled by the segmentation portion of the algorithm, were fed into a neural network along with their associated age to create a network

which predicts age. By processing scalar volumes as opposed to 3D images, the training speed of the algorithm is greatly increased. However, in this process, 3D information is lost such as label location and shape, which could decrease the accuracy of the network.

The final network architecture consists of five layers. First a fully connected later with 20 nodes to reduce the number of features, then a batch normalization, followed by a RELU activation layer, then a fully connected layer, and finally an output layer. The network was trained on 100 segmentations due to time limitations associated with the segmentation portion of the algorithm and 23 other segmentations were held out for validation.

IV. EXPERIMENTS AND REUSLTS

A. Skull-stripping Method

To test the novel skull-stripping method, the team compared the results of brain masks generated using our method to those generated by HD-BET, a popular and well-established skull-stripping package. For 70% of the training data, the dice coefficient between the team's mask and the HD-BET mask was greater than 0.90, indicating very good overlap. However, for the remaining 30% of the training data, the dice coefficient was as low as zero, indicating zero overlap. Our novel skull-stripping method works very well in cases where our assumptions hold and very poorly in cases which they do not.

B. Segmentation

To test the segmentation method, the team used our affine registration and majority voting label fusion method to segment a volume with an associated manual delineation. We then calculated the dice coefficient between the segmentation and the ground truth manual delineation. For the two provided test segmentations, the average dice coefficient between our segmentation and the ground truth was 0.42 and 0.44, indicating fairly poor overlap in most regions of the segmentation. Figure 4 shows an example of a segmentation produced by our method compared to its associated ground truth manual delineation.
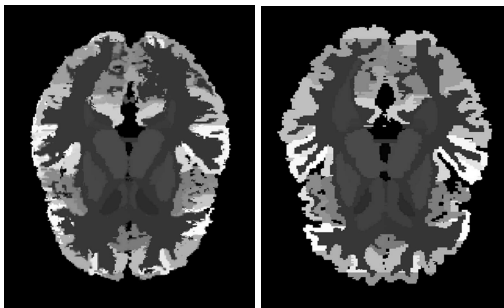

Fig. 4. Segmentation produced by our algorithm (left) compared to the ground-truth manual segmentation (right).

The age prediction network was initially evaluated based on training error vs. iteration, as shown in figure 5. The

final training RMSE was about 14 years and stabilized after about 500 iterations.
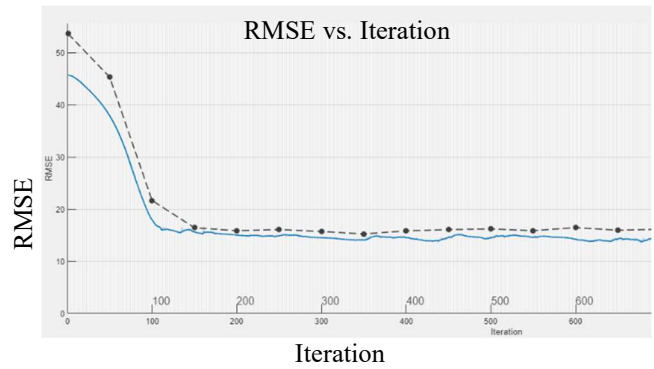

Fig. 5. RMSE vs. Iteration for age prediction network training data.

To further evaluate the age prediction method, the team held out 23 segmentations with their associated gray matter and white matter volume from the training set and then ran these images through the age prediction algorithm. The average RMSE of these held out test images was 6.4 years. The fact that the testing RMSE is lower than the training RMSE indicates that the model is not overfitted, however this test set is too small to draw significant conclusions. Figure 6 shows a plot of actual age vs. predicted age for this validation data.
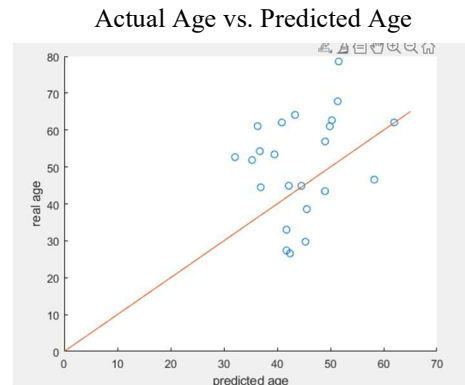

Fig. 6. RMSE vs. Iteration for age prediction network training data.

V. DISCUSSION

A. Skull-stripping Method

The skull-stripping method developed by the team relied on the assumption that the CSF was darker than all other tissues and that the CSF consistently separated the brain from the skull in 3D space. This assumption allowed the team to develop a very fast algorithm that is quite effective for most images in the training set, but this assumption does not hold for all images in the set. For this reason, the skull-stripping method developed by the team was only effective for about 70% of the training data set. For 70% of the training data set, the dice coefficient was above 0.90, however because the dice coefficient was very low on the remaining 30% of the data, the

overall average dice coefficient was only about 0.65 across all training data. Ultimately, a threshold-based skull stripping method is not robust to data with different intensity ranges.

B. Segmentation

The average dice coefficient of our segmentations was about 0.40, which was lower than the team would have liked. Our segmentation algorithm could have been improved significantly by a diffeomorphic registration method as opposed to the affine registration that we used. Additionally, the computation time for the segmentation portion of our algorithm was longer than we would have liked. With more time, the team would have pursued a deep-learning-based diffeomorphic registration package such as voxelmorph or ANTS, which would have likely proven both faster and more accurate than the current affine registration method.

C. Age Estimation

The team aimed to create a network with an RMSE of less than 10 years, as this is moderately successful when compared to other recent segmentation-based brain age prediction algorithms [9]. However, the actual RMSE of our network was 14 years across the training data. The testing RMSE of the network was only 6.4 years. However, the small held out testing set of only 23 images makes the team hesitant to conclude that this algorithm can consistently predict age with an RMSE of 6.4 years.

The accuracy of the age estimation portion of the algorithm is highly dependent upon the accuracy of the segmentation and therefore was impacted by our relatively poor segmentation results. Additionally, training the network on a larger set would have likely improved the accuracy of the network. However, the team was able to create a fully automated algorithm which is able to segment 15 brain MRIs and make an age prediction based on the data within the time constraint of 24 hours.

D. Demo Day Results

The team's skull-stripping algorithm was not compatible with the 5 provided segmentation images, so the team used HD-BET as a skull-stripping method for the segmentation images. However, the skull-stripping method was successful for 9 of the 15 provided age images and because our skull-stripping method is significantly faster than HD-BET, this method was used to skull-strip 9 of the 15 age prediction images.

Our algorithm was able to segment and predict ages for 15 images as well as segment an additional 5 images within the allotted 24-hour time constraint. On the final testing set, our average segmentation dice score was 0.3642, which was relatively low compared to other teams, but expected based on our initial results. Our RMSE age error was 15.73 years which was similar to our training error and thus also expected. Overall, the team ranked 5th out of seven teams, which we are happy with considering that we are one of only two undergraduate

teams and made up of only three members. Future work to improve our multi-atlas segmentation method includes implementing a diffeomorphic registration method and pursuing weighted voting for label fusion. An improvement in our segmentation method would likely also increase the accuracy of our age prediction network.

VI. TEAM MEMBER CONTRIBUTIONS

I was primarily responsible for project 2 task A and project 1 task B. I created a relatively successful skull stripping method as well as a registration and label fusion segmentation method for project 2. I also created a method to identify the EDV and ESV contours in project 1 and estimate ejection fraction.

Table 1 indicates the primary responsibilities of each team member. Tasks are assigned as indicated with an "X" in the table, however team members contributed to other parts of the project when needed.

TABLE I.

Team Member	Project 1 Task A	Project 1 Task B	Project 2 Task A	Project 2 Task B
Nyeli Kratz		X	X	
Sharon Reitsma	X			
Jiayin Qu				X

REFERENCES

[1] A.M. Fjell, K.B. Walhovd, "Structural Brain Changes in Aging: Courses, Causes and Cognitive Consequences," *Rev Neurosci.,* vol. 21, no. 3, pp. 187-221. 2010.

[2] H. Sajedi, N. Pardakhti. "Age Prediction Based on Brain MRI Image: A Survey." *J Med Syst.* 2019 Jul 11; vol. 43, no. 279.

[3] J. Wu, X. Tang. "Brain Segmentation Based on Multi-Atlas Guided 3D Fully Convolutional Network Ensembles." arXiv preprint arXiv: 1901.01381, 2019.

[4] S. Valizadeh, J. Hänggi, S. Mérillat, L. Jäncke. "Age Prediction on the Basis of Brain Anatomical Measures." *Hum Brain Mapp.* (2017) vol.38, pp. 997-1008.

[5] B.A. Jonsson, G. Bjornsdottir, T.E. Thorgeirsson, et al. "Brain Age Prediction Using Deep Learning Uncovers Associated Sequence Variants. *Nat Commun* (2019) vol. 10, no. 5409.

[6] P. Aljabar, R.A.Heckemann, A. Hammers, J.V. Hajnal, D. Rueckert, "Multi-Atlas Based Segmentation of Brain Images: Atlas Selection and its Effect on Accuracy." *NeuroImage.* (2009) vol. 46, no. 3, pp. 729-738.

[7] M. Lorenzi, N. Ayache, G.B. Frisoni, X. Pennec. "LCC-Demons: A Robust and Accurate Symmetric Diffeomorphic Registration Algorithm," NeuroImage. (2013). vol. 81. pp. 470-483.

[8] G. Balakrishnan, A. Shao, M. Sabuncu, J. Guttag, A. Dalca. "VoxelMorph: A Leaning Framework for Deformable Medical Image Registration." IEEE TMI: Transactions on Medical Imaging. 2019.

[9] H. Jiang, N. Lu, K. Chen, L. Yao, K. Li, J. Zhang, X. Guo. "Predicting Brain Age of Healthy Adults Based on Structural MRI Parcellation Using Convolutional Neural Networks" Front. Neurol. (January 2020) vol. 10, no. 134