Estimating a Person's Age based on MRI Brain Scans

Thomas Nash Viire Talts Sarah Wiegreffe

May 25, 2015

Contents

1 Introduction

Until the 20th century people older than 60 years made up less than 5% of the global population. In 2014, the percentage of the population over sixty years old was nearly 12 and it is still growing. By the year 2050 it is expected to rise to 21% (GAWI, 2014). Ageing of a population brings along ageing-associated health problems. Age is a major risk factor for most common neurodegenerative diseases. Different brain diseases cause 50% of years of life lived with disability (Olesen & Leonardi, 2003). Evaluating the state of the brain and comparing it to normal data allows the risks of age-related diseases to be assessed and the prevention process to be started earlier, even before the cognitive or other disorders have been revealed. For this, exact understanding of what morphological changes take place in the ageing brain is needed.

Some have stated that normalized whole-brain volume (nWBV) diminishes with age (Marcus et al. 2007). Several studies have shown that global grey matter volume decreases with increasing age (Good et al., 2001; Sowell et al., 2004).

Also, white matter loss has been consistently reported as following the same trend (Sowell et al., 2004). On the contrary, Good (et al., 2001) report that no global white matter volume decline with age was detected in their research, but rather local areas face relative accelerated loss and preservation.

Changes in the brain do not become apparant to the same extent in all brain regions (Trollor and Valenzuela, 2001). Reviewing the research shows that the most affected region is the prefrontal cortex. But reductions in the volume also appear in the striatum, temporal lobe, cerebellar vermis, cerebellar hemispheres, and hippocampus (Raz, 2004).

Most studies trying to indicate brain tissue changes use cross-sectional data (Raz et al, 2005; Sowell et al., 2004). But there is also a possibility that cross-sectional brain tissue volume data analyses do not necessarily depict the real brain tissue loss taking place. It can only be deduced because the assessment is not done over time and differences between the data of separate age-groups may be the result of cohort differences (Sowell et al., 2004). There are only a few longitudinal studies which analyze age effects on brain volumes. Many of these use only two assessments with different amounts of time between them (Sowell et al., 2004). In a longitudinal study with 116 subjects (aged 59-85) no changes caused by ageing were detected in global brain volume. However, this does not prove the stability of brain volume over time because there was only a year between the two assessments (Resnick et al., 2000).

On a contrasting note, a longitudinal study (together with a cross-sectional study) in which the mean $\pm$ SD follow-up interval was 5.27 ± 0.30 years concluded that the caudate, the cerebellum, the hippocampus and the association cortices shrunk significantly with increasing age. The authors also found even bigger brain tissue shrinkage compared to the cross-sectional estimates (Raz et al., 2005).

Changes in the brain associated with ageing are quite individual. The biggest variations in changes have been found in in the cerebellum, the prefrontal white matter, the fusiform gyrus, the visual cortex and the inferior temporal cortex. Differences in brain tissue shrinkage with increasing age have been found between genders (Good et al., 2001; Xu et al., 2000). The maintenance of cognitive function in old age is positively correlated to the number of years spent learning (education). Volumes of brain structures correlate with a person's age, gender, level of education, socioeconomic status, ethnicity, and whether or not (s)he has Alzheimer's Disease. The possibility to create a model that could predict characteristics such as these simply by analyzing a subject's brain structure volumes via magnetic resonance imaging (MRI) scans has been suggested (Long and Holder, 2013). Our project will follow this lead.

The aim of our project is to create a regression model to predict a person's age based on his magnetic resonance imaging (MRI) scans, by assessing changes in different brain area volumes with increasing age.

2 Methods

2.1 Subjects

For our data, we used the Open Access Series of Imaging Studies (OASIS), which is a publicly available series of magnetic resonance imaging data sets. The OASIS cross-sectional dataset contains data of 416 subjects aged 18 to 96 with three or four individual T1-weighted MRI scans for each subject, obtained in single imaging sessions. All images are in 16-bit big-endian Analyze 7.5 format (Marcus et al., 2007). Each MRI scan file, of which there are 3 or 4 for each subject, has a structure of size 176x208x176 composed of voxels, representing a segmented brain. Each voxel is assigned a value of 1 through 3 representing spinal fluid, grey matter, and white matter, respectively.

In addition to the MRI scans, the datasets contain data on each subject's age, gender, handedness, socio-economic status, education and medical status. In addition, derived anatomic volumes such as estimated total intracranial volume (eTIV) (mm3), Atlas scaling factor (ASF) and normalized whole brain volume (nWBV) are given.

The subjects are graded by Clinical Dementia Rating (CDR) of 0-2- 0 showing nondementia, 0.5 very mild dementia, and 1 mild dementia (Marcus et al., 2007). 100 of the subjects in OASIS datasets have very mild to moderate Alzheimer's Disease (AD) as distinguished by their CDR score. In brains affected by AD, loss of grey matter volume, especially severe in the hippocampus, and the enlargement of ventricles is evident (Nestor et al., 2008; Wenk, 2003). For this reason, including the data of these subjects in the analysis would interfere with the detection of age-related changes. So we excluded subjects both with mild and moderate AD and for later analyses we used the dataset with 316 subjects (197 women and 119 men), all with CDR scores of 0 (Figure 1). The subjects are all right-handed, aged between 18-94.

Although gender affects the localization of more accelerated changes of the brain (Good et al., 2001; Xu et al., 2000), we included both men and women in the sample. We also did not take into account subjects' education nor their socio-economic status because of the limited scope of this project. We wished to focus uniquely on brain area volumes.

2.2 Data validation

Although normalized whole brain volume (expressed as the percentage of all voxels in the atlas-masked image that are labeled as grey or white matter by the automated tissue segmentation process) was already given in the OASIS dataset (Marcus et al., 2007), we computed nWBV ourselves in order to validate our voxel-counting method for measuring brain region volumes. This was done by totaling the number of voxels of each brain scan classified as grey or white matter, and calculating this value as a proportion of total voxels (where spinal fluid is included). Comparing nWBV calculated by us to the value that was given in the dataset showed that we had obtained 99.9% accuracy for all subjects.

		FEMALE		MALE	
Age group	**Total n**	**n**	**Mean**	**n**	**Mean**
<20	19	9	18.44	10	18.6
20s	119	68	22.56	51	23.16
30s	16	5	33.4	11	33.36
40s	31	21	45.29	10	46.2
50s	33	22	54.32	11	54.45
60s	25	18	64.72	7	65.29
70s	35	25	73.4	10	73.3
80s	30	22	83.82	8	84.75
90s	8	7	91.14	1	90
Total	316	197		119	

Figure 1: Breakdown of the dataset by age and gender.

Using this data, we created a regression plot showing the relationship between age and nWBV of the subjects in our dataset (see Figure 7). The significance of this will be addressed in the results section.

2.3 Morphometric analysis of brain structure

For detailed anatomical labelling of brain scans, the Talairach Atlas was used. The Talairach Atlas is a 3-dimensional coordinate system of the human brain, which can be used to map the location of brain structures independent from individual differences in the size and overall shape of the brain. The Talairach-Tournoux (TT) coordinate system is defined using landmarks inside the brain and therefore can only be determined from an MRI scan. The landmarks used in the TT coordinate system are the anterior and posterior commisura (AC and PC), and the coordinate axes are defined according to

- the origin of the TT coordinate system is in the AC
- the y-axis increases towards the front of the brain, along the line connecting PC and AC
- the z-axis increases towards the top of the brain
- the x-axis increases towards the right side of the brain

(Toennies, 2012). Talairach.nii is a NIfTI image that contains the Talairach label data. The label data are stored as text in the extension section of the image. By referring to the atlas it is possible to identify tissue at a specific location. In the atlas every (x,y,z) voxel has a number and this number corresponds to a hierarchy of labels. The Talairach Atlas has 5 levels of hierarchy.

The OASIS files and the atlas are both centered using the same methods and each coordinate is the same size, but they have different dimensions. The atlas is 141x172x110 voxels and the OASIS files are 176x208x176 (Figure 2). For labeling the OASIS brain scans using Talairach, we had to first align the two.

Both the atlas and the individual scans are referred to in the common Talairach space. This space has an origin, i.e. physical coordinates (0,0,0), at a point called the Anterior Commissure point. The arrays of the atlas and scans are of different size but organized so that this origin point occurs at the "middle" voxel in both arrays. So the alignment simply matches the "middle" voxel of both arrays. This middle voxel is the voxel occurring in the middle of each direction. For example for the atlas of size 141x172x110, the middle voxel occurs at the array position 71x86x55. One small issue is that for directions that have an even number of sites (such as the last two directions in the former example) there are two indices that can be considered the middle, 86 and 87, 55 and 56 for the last two directions respectively, so an ideal alignment should be centered at 86.5 and 55.5. We divided each dimension by two and rounded up to get the midpoint (71 x 86 x 55 for the Talairach, 88 x 104 x 88 for the OASIS). So instead of ranging from 1-176 on the x-axis on the OASIS, the centered Talairach ranged from 18-158 (Figure 3).

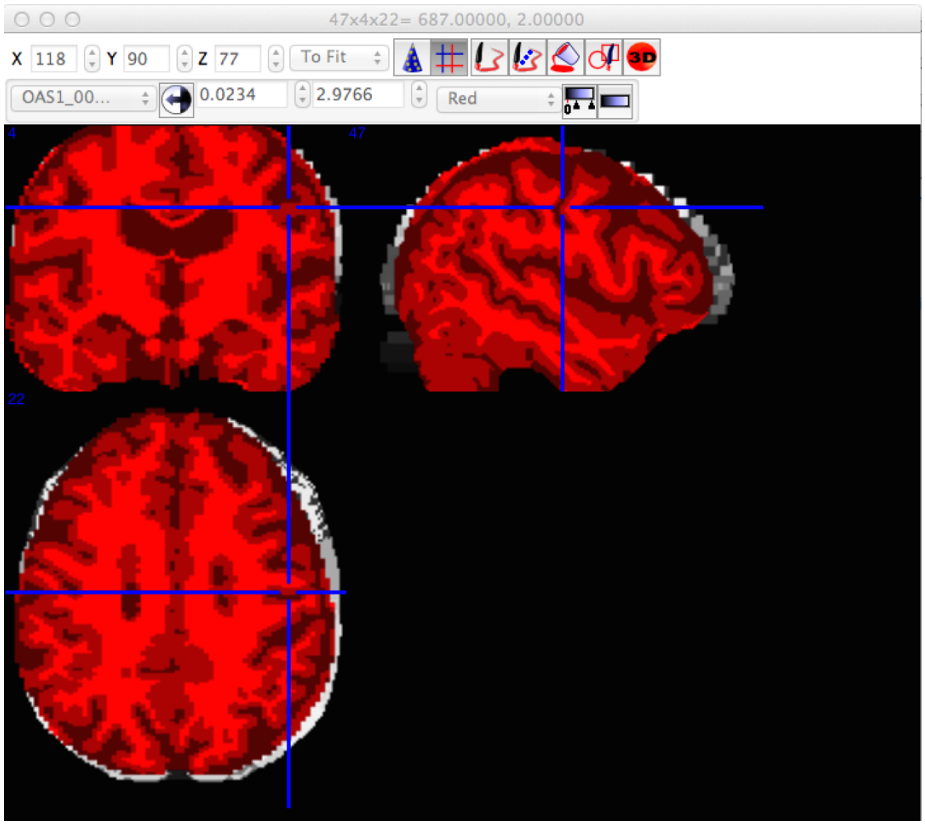

Figure 2: The Talairach atlas overlayed on an OASIS brain scan. The difference in dimensions can be seen.

After adjusting each of the three dimensions to find out what coordinates in the OASIS are covered by the Talairach, we could establish a starting point in the OASIS space at (18, 19, 34) and an endpoint at (158, 190, 143) and then start at the Talairach starting point of (1,1,1) and iterate through all points to assign a brain region label to the subjects' scans. The approach was to start at this point, then to get all neighboring points on the current slice, the one above, and the one below. We checked each -1, +0, and +1 to the current x,y,z coordinate and all possible combinations of such to get the 27 surrounding points, which make up a 3x3x3 cube centered around the current voxel. Then, taking the regions reported in the Talairach from these 27 points (sometimes there are less, for example the first point (1,1,1) has no slices below it), and finding the most frequent one, the current position was set to that region.

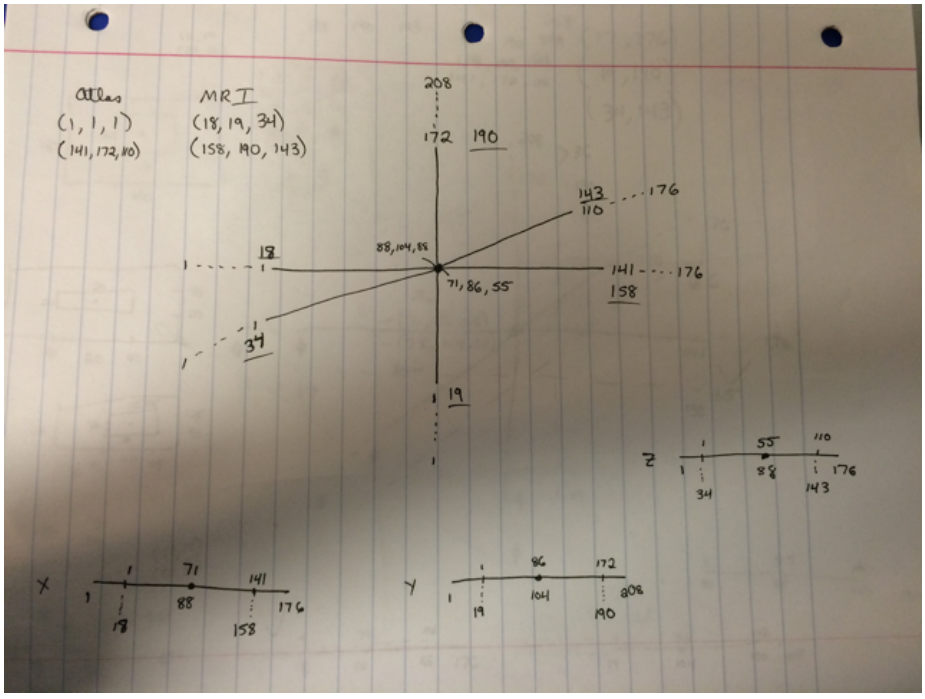

Figure 3: A visual of the alignment of an OASIS brain scan and the Talairach atlas.

2.4 Building the Model

2.4.1 Hierarchies of predictors

We used the volumes of different areas of the brain that we measured as predictors in our model. The Talairach atlas has a hierarchy of labels, three of which were effective for our purposes.

```
121     Left Cerebrum.Temporal Lobe.Fusiform Gyrus.*.*
122     Right Cerebrum.Temporal Lobe.Fusiform Gyrus.*.*
123     Left Cerebrum.Temporal Lobe.Fusiform Gyrus.Gray Matter.Brodmann area 20
124     Right Cerebrum.Temporal Lobe.Fusiform Gyrus.Gray Matter.Brodmann area 20
128     Left Cerebrum.Temporal Lobe.Fusiform Gyrus.White Matter.*
131     Right Cerebrum.Temporal Lobe.Fusiform Gyrus.White Matter.*
132     Right Cerebrum.Limbic Lobe.Fusiform Gyrus.*.*
135     Right Cerebrum.Limbic Lobe.Fusiform Gyrus.Gray Matter.Brodmann area 20
136     Left Cerebrum.Temporal Lobe.Fusiform Gyrus.Gray Matter.Brodmann area 36
137     Left Cerebrum.Limbic Lobe.Fusiform Gyrus.White Matter.*
138     Right Cerebrum.Limbic Lobe.Fusiform Gyrus.White Matter.*
```

Figure 4: A sample of the hierarchy of labels used in the Talairach atlas.

The "full volume" model uses predictors at the third level of the hierarchy. Referring to Figure 4, a small sample of the Talairach labels, "full volume" means grouping on a specific path:

Left Cerebrum.Temporal Lobe.Fusiform Gyrus - 121,123,128,136
Right Cerebrum.Temporal Lobe.Fusiform Gyrus - 122,124,131
Left Cerebrum.Limbic Lobe.Fusiform Gyrus - 137
Right Cerebrum.Limbic Lobe.Fusiform Gyrus - 132,135,138

The "local volume" model also groups at the third level, but groups by region in general:

Fusiform Gyrus - 121,122,123,124,128,131,132,135,136,137,138

Finally, the "Brodmann" model groups by Brodmann volumes, the fifth level of the hierarchy and the most specific:

Brodmann area 20 - 123,124,135
Brodmann area 36 - 136

2.4.2 Using LASSO Regression

Because of the high number of predictors present in our models, we used a "least absolute shrinkage and selection operator", the LASSO method. This regularized version of least-squares regression penalizes for having too many predictors, helping to avoid overfitting our model. It minimizes the usual sum of squared errors, with a bound on the sum of the absolute values of the coefficients (Tibshirani, 1996). After using LASSO, we were able to achieve models with similar MSEs using many fewer predictors (see Figures 4 and 5).

Standard Regression Models

	R-squared Value	Num. Predictors (of 191)
Full Volume	0.6511	191
Local Volume	0.8747	56
Brodmann Regions	0.8633	43

Figure 5: Statistics on regression models built. Compare to Figure 6.

LASSO Regression Models

	R-squared Value	Num. Predictors
Full Volume	0.8530	14 of 191
Local Volume	0.8021	13 of 56
Brodmann Regions	0.8392	9 of 43

Figure 6: Statistics on LASSO models built. Compare to Figure 5.

3 Results

Figure 7 depicts the relationship between age and nWBV. There appears to be a relatively strong correlation between increasing age and decreasing normalized whole brain volume.

Our initial result regression models built using the three hierarchies of predictors are represented in Figures 8, 9, and 10. After using the LASSO method, our results improved. Our three final models are represented in Figures 11, 12, and 13. These plots illustrate actual vs. predicted age, computed on a training set size of 250 and a test set size of 66. Actual age is represented by the blue line, and predicted age by the red.

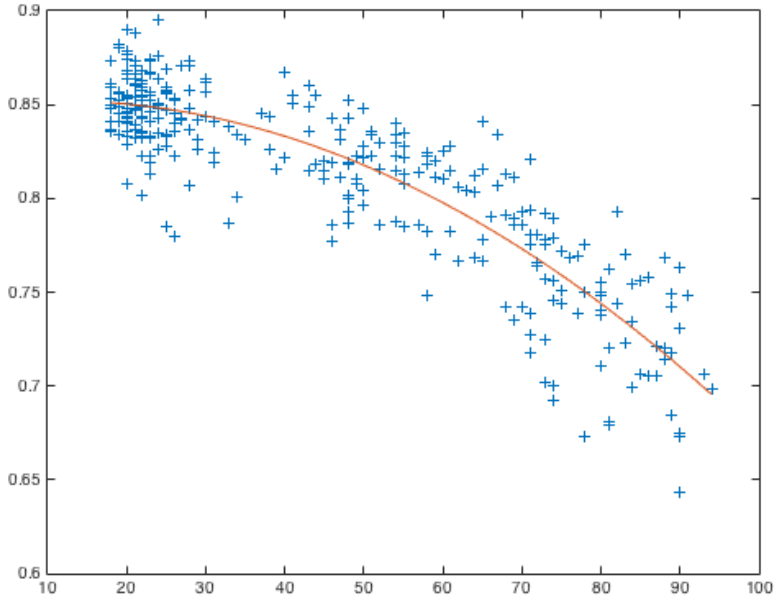

Figure 7: Polynomial regression plot showing age versus normalized whole brain volume (%). Each point represents a unique subject from a single scanning session.

In the "full volume" model, the top statistically significant areas of volume change related to aging are in the:

Left Cerebrum.Sub-lobar.Transverse Temporal Gyrus
Right Cerebrum.Limbic Lobe.Fusiform Gyrus
Right Cerebrum.Limbic Lobe.Anterior Cingulate
Left Cerebrum.Temporal Lobe.Subcallosal Gyrus
Left Cerebrum.Frontal Lobe.Inferior Parietal Lobule

So, it appears that side of the brain plays some significance, and that the lobes of the cerebrum are very significant. This model is illustrated in Figure 11, and uses 14 of 191 statistically significant brain area paths as predictors, with an R-squared value of 0.8530.

In the third level of hierarchy several areas which show statistically significant volume loss were found (the "local volume" model). The most statistically significant predictors for the model are the declive of vermis, insula, precentral gyrus, caudate, and cingulate gyrus. This model, "local volume", uses 13 of 56

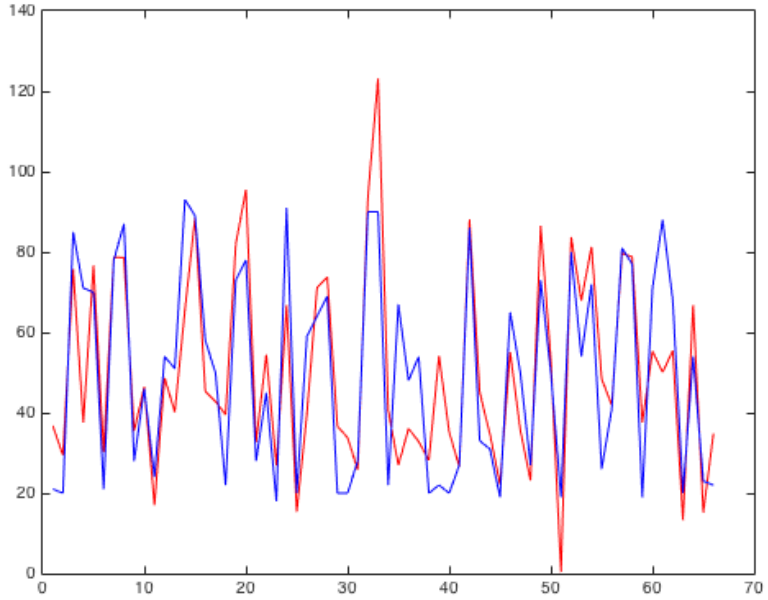

Figure 8: Regression plot showing actual vs. predicted age for the "Full Volume" model using the exact path to the third level hierarchy. (R-squared: 0.6511, Predictors: 191)

brain areas as statistically significant predictors and has an R-squared value of 0.8021 (see Figure 12).

Several Brodmann areas which showed statistically significant volume loss were found using the "Brodmann" model. The biggest changes were observed in Brodmann areas 13, 30, 6, 25, and 40. This model uses 9 of 43 Brodmann areas as statistically significant predictors and has an R-squared value of 0.8392 (see Figure 13).

4 Discussion

One of the major findings in our study is the decreasing nWBV caused by increasing age. The finding is in consistence with other studies (Marcus et al., 2007) As a limitation of the study we did not investigate the changing trends of global white matter volume and some further research is needed.

Another major finding in our study is the computational method of applying an atlas to brain scans to measure volumes of all regions of the brain.

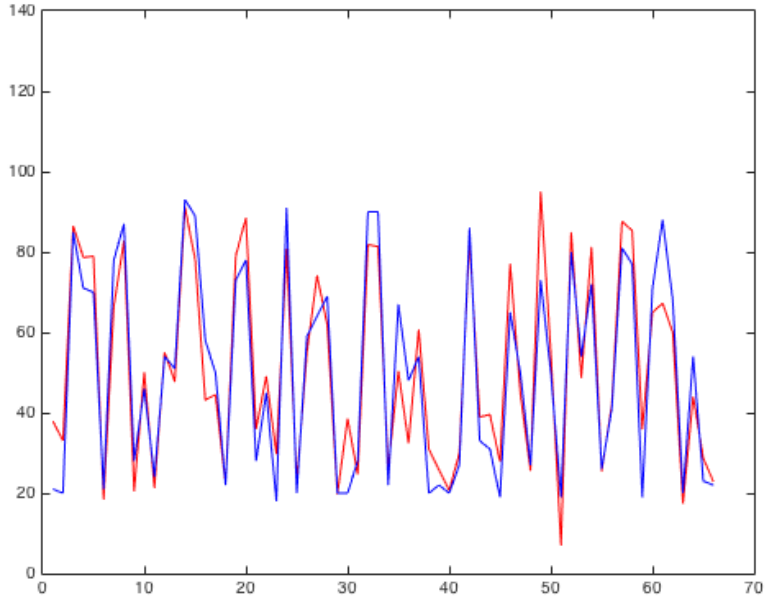

Figure 9: Regression plot showing actual vs. predicted age for the "Local Volume" model using only the third level hierarchy. (R-squared: 0.8747, Predictors: 56)

It is difficult to draw conclusions on what areas of the brain, or which level of hierarchy, over our three models are the most significant in predicting a subject's age. Further research is needed.

5 Conclusions

Normalized whole brain volume decreases with increasing age.

Changes in the volume of several brain regions correlate with increasing age.

A person's age is predictable by his MRI scan.

The effectiveness of detecting a person's age by his MRI scan depends on which predictors and which brain regions' level of hierarchy is used.

6 References

Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, Snyder AZ. (2004). A unified approach for morphometric and functional data analy-

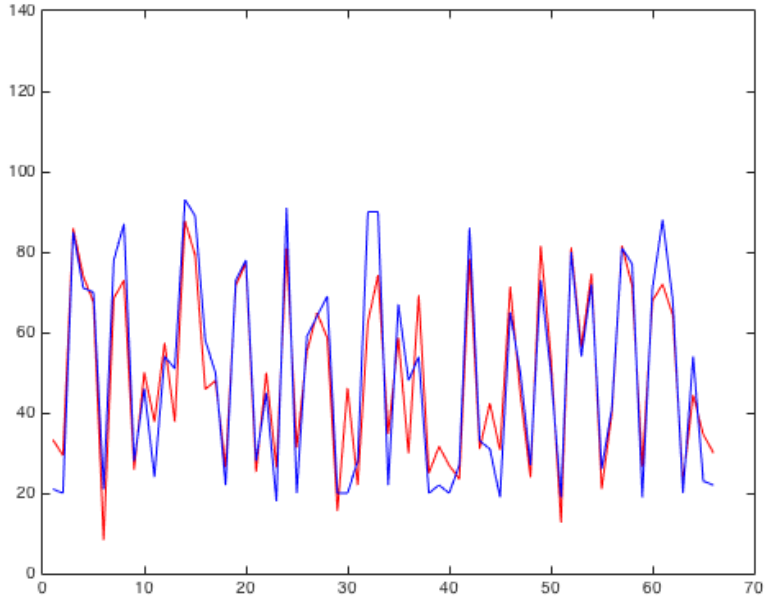

Figure 10: Regression plot showing actual vs. predicted age for the "Brodmann" model using the fifth level hierarchy, the Brodmann areas. (R-squared: 0.8633, Predictors: 43)

sis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage, 23: 724 – 738.

Global Age Watch Index. Inside report. (2014). Available at: http://www.helpage.org/global-agewatch/reports/global-agewatch-index-2014-insight-report-summary-and-methodology/ Date accessed: 1 May. 2015.

Good CD, Johnsrude IS, Ashburner J, Henson RNA, Friston K J, Frackowiack RSJ. (2001). A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. Neuroimage, 14: 21-36.

Long S, Holder L. Discovery of Discriminating Neural Regions for MRI Classification. AAAI Workshops, North America, jun. 2013. Available at: ¡https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/view/7154¿. Date accessed: 23 May. 2015.

Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. (2007). "Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI

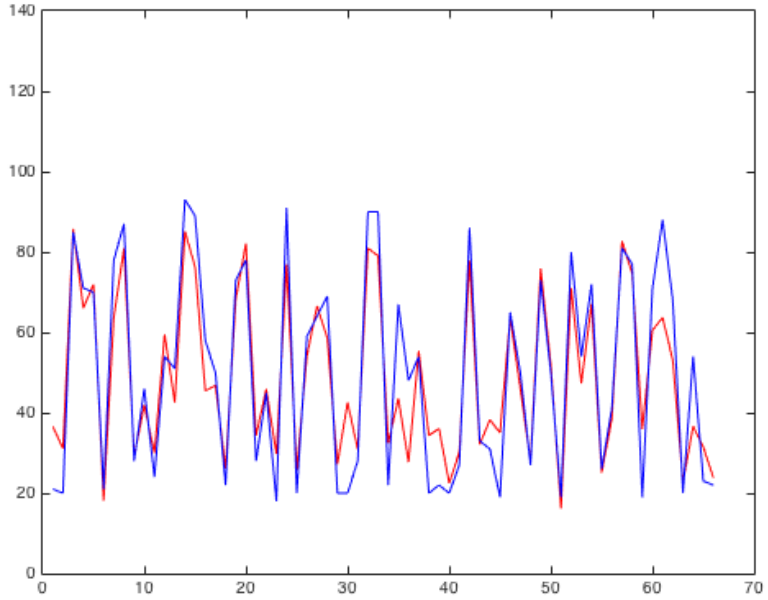

Figure 11: Final Model 1: LASSO normalized regression plot showing actual vs. predicted age for the "Full Volume" model using the exact path to the third level hierarchy. (R-squared: 0.8530, Predictors: 14 of 191)

Data in Young, Middle Aged, Nondemented, and Demented Older Adults". Journal of Cognitive Neuroscience, 19:9: 1498–1507 .

Nestor S, Rupsingh R, Borrie M, Smith M, Accomazzi V, Wells J, Fogarty J, Bartha, R. (2008). Ventricular enlargement as a possible measure of alzheimer's disease progression validated using the alzheimer's disease neuroimaging initiative database. Brain 131(9):2443.

Olesen, J. and Leonardi, M. (2003). The burden of brain diseases in Europe. European Journal of Neurology, 10: 471–477.

Raz N. (2004). The ageing brain: structural changes and their implications for cognitive ageing. In: Dixon R, Bäckman L, Nilsson L, eds. New frontiers in cognitive ageing. Oxford: Oxford University Press.

Raz N, Lindenberger U, Rodrigue KM, Kennedy KM, Head D, Williamson A, Dahle C, Gerstorf D, Acker JD. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cereb Cortex 15:1676–1689.

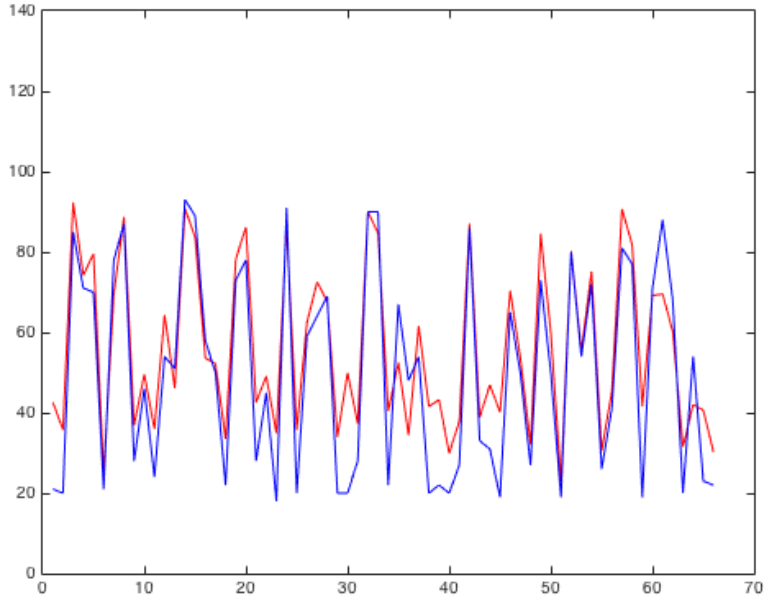

Figure 12: Final Model 2: LASSO normalized regression plot showing actual vs. predicted age for the "Local Volume" model using only the third level hierarchy. (R-squared: 0.8021, Predictors: 13 of 56)

Resnick SM, Goldszal AF, Davadzikos C, Golski S, Kraut MA, Metter EJ, Brian RN, Zonderman AB. (2000). One-year Age Changes in MRI Brain Volumes in Older Adults. Cerebral Cortex, 10: 464-472.

Sowell ER, Thompson PM, Toga AW. (2004). Mapping Changes in the Human Cortex throughout the Span of Life. Neuroscientist, 10: 372-392.

Staff RT , Murray AD, Deary IJ, Whalley LJ. (2004). What provides cerebral reserve? Brain, 127(5): 1191 - 1199.

Talairach.org. Available at: http://www.talairach.org/about.html Date accessed: 1 May. 2015.

Tibshirani R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, 58(1): 267-288.

Toennies KD. (2012). Guide to Medical Image Analysis: Methods and Algorithms. Springer. p. 326.

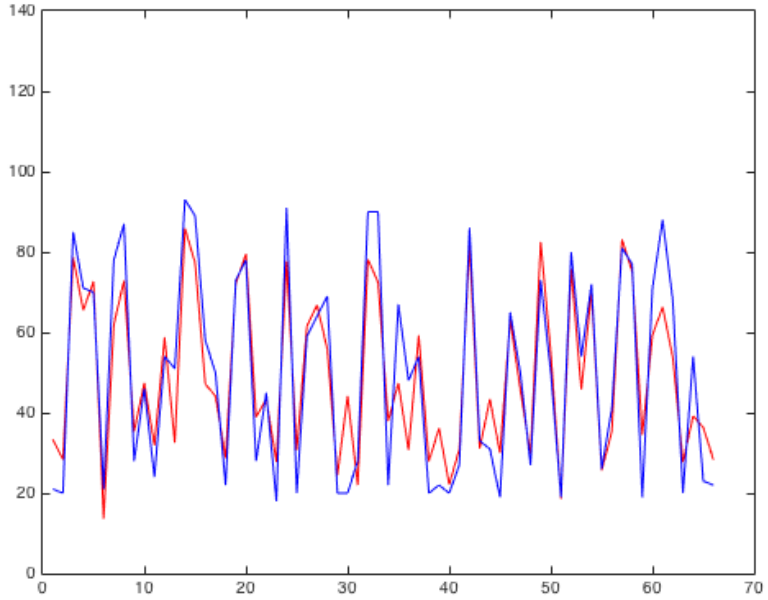

Figure 13: Final Model 3: LASSO normalized regression plot showing actual vs. predicted age for the "Brodmann" model using the fifth level hierarchy, the Brodmann areas. (R-squared: 0.8392, Predictors: 9 of 43)

Trollor JN, Valenzuela MJ. (2001). Brain ageing in the new millennium. Aust. NZ J Psychiatry 35:788 – 805.

Wenk GL. (2003). Neuropathologic Changes in Alzheimer's Disease. The Journal of Clinical Psychiatry, 64 Suppl 9:7–10.

Xu J, Kobayashi S, Yamaguchi S, Iijima K, Okada K, Yamashita K. (2000). Gender effects on age-related changes in brain structure. AJNR Am J Neuroradiol, 21(1):112-8.