Predicting Age across Human Lifespan Based on Structural Connectivity from Diffusion Tensor Imaging

Cheol E. Han[1;2], Luis R. Peraza[3], John-Paul Taylor[3], Marcus Kaiser[2;3;4]

Email: cheolhan@korea.ac.kr, luis.peraza-rodriguez@newcastle.ac.uk, john-paul.taylor@newcastle.ac.uk, m.kaiser@ncl.ac.uk

[1] Department of Biomedical Engineering, Korea University, Republic of Korea (South Korea)
[2] Department of Brain Cognitive Sciences, Seoul National University, Republic of Korea (South Korea)
[3] Institute of Neuroscience, Newcastle University, UK
[4] School of Computing Science, Newcastle University, UK

Abstract—**Predicting brain maturity using noninvasive magnetic resonance images (MRI) can distinguish different age groups and help to assess neurodevelopmental disorders. However, group-wise differences are often less informative for assessing features of individuals. Here, we propose a simple method to predict the age of an individual subject solely based on structural connectivity data from diffusion tensor imaging (DTI). Our simple predictor computes a weighted sum of connection strengths of an individual, where weights are the importance of that connection for an observed feature—age in this case. The weights are simply determined through correlations between connection strength and age; thus the proposed predictor requires no parameter tuning. We tested this approach using DTI data from 201 healthy subjects aged 4 to 85 years. After determining importance in a training dataset, our predicted ages in the test dataset showed a strong correlation ($r = 0.79$) with real age deviating by, on average, only about 9 years.**

Keywords—brain connectivity; prediction; neural networks; network analysis; ageing; neuroinformatics

I. INTRODUCTION

The study of how different components of the brain, may they be neurons or brain regions, are connected has become an emerging field within the neurosciences [1, 2]. Structural connectivity observes the physical wiring of neural circuits, while functional connectivity links brain regions with similar activity over time. Magnetic Resonance Imaging (MRI) facilitates human studies since it enables us to construct such networks noninvasively. At the macro scale where network nodes represent brain regions, Diffusion Tensor Imaging (DTI) allows us to quantify the number of streamlines as a proxy of connection strength between two regions. Analyzing brain networks can be a tool to understand the interaction between nodes and has already shown a strong correlation between cognitive functioning and global information integration for functional networks [3]. Predicting brain maturity from MR images is beneficial to assess neurodevelopmental disorders. As pediatric disorders including attention deficit/hyperactivity disorder (ADHD) have delays in brain maturity [4], investigating structural alterations in the brains of children with probable disorders can be a good screening strategy. Also in the elderly, it is beneficial to distinguish cognitive disorders from normal ageing. While a link between functional connectivity and brain maturity was reported earlier [5], it

remains unclear whether structural connectivity alone can be a comparatively suitable predictor, as functional and structural network are often related [6]. Taking advantage of the network analysis framework and DTI, classifying structural networks to distinguish different age ranges has previously been performed by looking at sets of features for individual nodes, or single-node motifs [7, 8]. Also network topological changes over age were investigated in structural connectivity networks [9, 10]. Whereas these studies used aggregate features of the network, we employed only raw connectivity data about individual edges, as given by the number of streamlines to predict the age of a subject. Here, we proposed and tested a simple prediction model for the age of a subject using structural connectivity networks based on DTI. Our predictor computed a weighted sum of the structural connectivity matrix for each subject as a raw score, where the weight of each connection was predefined as a correlation coefficient between edge weights and age over all subjects in the training group. Then, we transformed the raw score to a predicted age with either linear or non-linear regression. We evaluated our method over a test group, computing the correlation between predicted ages and the real ages in the test group. We called this simple prediction model "correlation based regression (CBR)".

II. METHODS

A. Subjects and structural networks

We made use of a public DTI-database provided by the Nathan Kline Institute (NKI) [11], which we used in our pervious network study of human development [12]. In this study, we included 201 participants between 4 and 85 years. T1-weighted and diffusion-weighted MR images were obtained with a 3 Tesla scanner (Siemens MAGNETOM TrioTim syngo, Erlangen, Germany). T1 weighted MRI data were obtained with 1mm isovoxel, FoV 256mm, TR=2500ms, and TE=3.5ms. DTI data were recorded with 2mm isovoxel, FoV=256mm, TR=10000ms, TE=91ms, and 64 diffusion directions with b-factor of 1000 s mm^2 and 12 b0 images. Networks consist of nodes and edges, connections between nodes. As node definition, we extracted volume regions of interests (ROIs) in DTI space, using Freesurfer (http://surfer.nmr.-mgh.harvard.edu), which includes 34 cortical regions (Deskian atlas) and 7 subcortical regions (Nucleus accumbens, Amygdala, Caudate, Hippocampus, Pallidum, Putamen, and Thalamus) for each hemisphere, thus leading to 82 ROIs in

total. To quantify edge weights, we used streamline counts using DTI tractography. We performed eddy-current correction (FSL), and deterministic tracking (Diffusion Toolkit, [13]) with 35 degrees angular threshold. With registered ROIs, we then counted the number of streamlines between all pairs of defined ROIs as edge weights using the UCLA Multimodal Connectivity Package (http://ccn.ucla.edu/wiki/index.php). This led to an undirected (symmetric) weighted connectivity matrix S of size $N \times N \times M$, where N stands for 82 nodes, and M stands for 201 subjects. Because we did not count self-recurrent weights, all diagonal elements were zero. For further details of the network construction, see [12].

B. Proposed method: correlation based regression (CBR)

We focused on simple prediction through correlations between edge weights and age. Correlation coefficients, ranging from -1 to 1, give us a weighting measure of how important each connection is for a feature leading to group correlation matrix C. We used Pearson correlation coefficients $C_{ij} = cov(L_{ij}, A)/(\sigma_L \sigma_A)$, where L_{ij} is the vector of edge weights between node i and j for all subjects, A is the vector of age for all subjects, and cov and σ stands for covariance and variance.

In order to predict age, a predictor value P is then given by the sum over all edges of the individual product between the correlation matrix C and the subject's matrix S as follows (k represents subject index);

$$P_k = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} S_{ijk} \times C_{ij} \qquad (1)$$

The predictor value, P was mapped into a predicted age using a simple regression. We tested and compared both linear regression, $\hat{A} = a + bP$, and a non-linear exponential regression, $\hat{A} = \exp(a + bP)$, where a and b are regression parameters, and $\tilde{A}$ is the predicted age.

We also tested different connectivity matrix normalizations: 1) No normalization, L^0, 2) normalization within subjects, L^1, and 3) normalization between subjects, L^2. The last two are defined as:

$$L_{ijk}^1 = L_{ijk}^0 / (1/2 \sum_{j=1}^{N} \sum_{i=1}^{N} L_{ijk}^0) \qquad (2)$$

$$L_{ijk}^2 = L_{ijk}^0 / (1/m \sum_{k=1}^{m} L_{ijk}^0) \qquad (3)$$

Note that in L^2, m is the number of subjects in the training set; once we computed the denominator in a training set, we used it for both training and testing sets. L^1 was computed for each subject separately. L^1 was to test the effects of total number of streamline counts, and L^2 is recommended in most machine learning methods.

C. Comparisons with other prediction methods

Numerous machine-learning approaches could be applied for predicting features such as age. Here, we compared two other popular prediction methods: the support vector regression (SVR) [14], and the partial least square regression (PLSR) [15]. In contrast to our predictor, the methods require to tune parameters. We used ε-SVR (*libsvm v3.17*) with polynomial kernels of degree 1; complexity C and ε should be decided. PLSR projected the connectivity data into a lower dimensional

space, where covariance between the connectivity data and age in the space was maximized; the dimension of projection should be decided. We selected those parameters, using 10-fold cross-validation. Specifically, after dividing the whole dataset into 10 divisions randomly, we constructed a model with 9 divisions and evaluated performance (RMSE) for the other. We repeated for all 10 divisions.

D. Evaluation

We divided the dataset into two groups: a training matrix L_R (100 subjects) and a test matrix L_T (101 subjects), where the former was used to compute C and regression parameters, and the latter to compute P and $\hat{A}$. We reported a mean absolute error (MAE), $1/n \cdot \Sigma_k |A_k - \hat{A}_k|$ [16], and the Pearson correlation between real age, A, and predicted age, $\hat{A}$ for the testing set. To test the performance, we used a bootstrapping approach. L_T and L_R matrices were created by randomly dividing the population matrix L at 50% (half of the population for training or test), and we performed the prediction for each division. We repeated 100 divisions to obtain extreme and average performances of our age prediction. All of analyses were done with Matlab (2013a, Mathworks, Natick, USA).

III. RESULTS

A. Prediction performance of our method

With bootstrapping, we confirmed that the best performance was obtained with non-normalized connectivity matrix, L^0, and an exponential regression (Table 1). For the case, the correlation coefficient distribution between the predicted ages and the real ages for the 100 experiments spanned between 0.666 and 0.858 (mean±standard deviation; 0.791±0.032), while MAE spanned between 7.656 and 11.513 (9.499±0.674). For the three normalization methods, the best performance was obtained without normalization, followed by normalization within subjects, L^1. Lower performance with linear regression compared to the non-linear regression (Table 1) was supported with a non-linear relation between the predictor value P and the real age A (Figure 1A). Figure 1B depicted predicted ages with non-linear regression and non-normalized connectivity matrix for the best case (Table 1).

B. Local changes during ageing

Besides age prediction, there is a specific interest for neuroscientists to know where the local changes occur during ageing. To infer which brain edges were most correlated with age prediction, we averaged the 100 estimated correlation matrices C, and depicted edges whose absolute correlation values were within top 1% and 5% over all edges with real-value correlation coefficients (Figure 2). We found edges with

TABLE I.   PERFORMANCE OF AGE PREDICTION WITH CBR

regression method	connectivity normalization		
	L^0	L^1	L^2
linear regression	0.771±0.028 (10.412±0.620)[a]	0.683±0.047 (11.828±0.794)	0.593±0.044 (14.137±0.821)
exponential regression	**0.791±0.032** **(9.499±0.674)**	0.687±0.049 (11,636±0.807)	0.600±0.046 (14.387±0.909)

[a.] *Mean correlation* ± standard deviation *(mean MAE* ± standard deviation, years) *in the testing set, computed through 100 bootstrapping of the dataset*
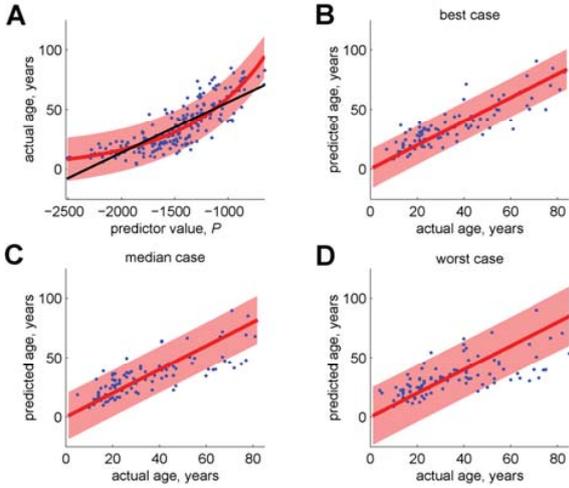
Fig. 1. Age prediction results. A) relationship between the predictor value P and real age A (for whole data set). Thin solid line for linear regression (R^2=0.660) and thick solid line for exponential regression (R^2=704) with 95% confidence interval shown asn shades. B,C,D) relationship between the real age and the predicted age with 95% confidence interval shown as shades for the best case ($r = 0.858$), median case ($r = 0.796$), and worst case ($r = 0.666$) in order (with exponential regression on non-normalized dataset, L^0).

decrease more than edges with increase during ageing. The top 1% edges (Figure 2A) consisted of 16 edges and they all decreased over ageing (median: -0.439, range: [-0.596, -0.389]). In top 5% edges (Figure 2B), during ageing 65 edges decreased (median: -0.299, range: [-0.241, -0.596]), while 17 edges increased (median: 0.261, range: [0.241, 0.381]). Note that the prediction was not led by neither top 1% nor 5% edges, but by the complete set of edges in the connectivity matrix.

C. Effects of the number of features

We also tested the effects of the number of features systematically. We constructed models with different number of edges, and collected predicted ages with the same way (non-normalized data with non-linear regression, 100 bootstrapping). As the number of edges used in the original experiment was 1637 edges (in averaged C), we selected the top strong edges from 100 to 1600 edges. As increasing the number of edges,
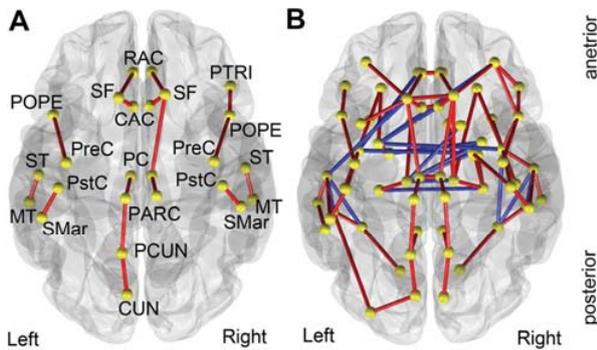

Fig. 2. Age prediction results. A) top 1% edges and B) top 5% edges, where red edges were decreased edges, and blue edges were increased edges during ageing over a transparent brain (the freesurfer's standard subject, visualized by our in-house codes).

the performance became better; Figure 3A, correlation between predicted ages and actual ages increased, and Figure 3B, MAE decreased, showing no over-fitting. Using More than 1200 edges, the average performance reached to the 95% of the performance with all edges.

D. Comparison with other methods

We compared our simple model with other methods: SVR, and PLSR. They require parameter tuning, which were decided where 10-fold cross validation error was the lowest. For SVR, we used following parameter sets [C, ε]: [0.44, 3.6] for L^0, [250, 0.12] for L^1, and [1, 0.1] for L^2. For PLSR, we used 4 dimensions for L^0, 3 for L^1, and 9 for L^2. Table 2 showed that our method ($r = 0.791$) performed better than SVR ($r = 0.772$) and PLSR ($r = 0.724$). As CBR did, these methods performed best with L^0, compared to other connectivity normalization.

IV. DISCUSSION

Our simple predictor estimated age of a subject given structural connectivity data with fair performance. The predicted age was highly correlated with the real age ($r = 0.79$ in the testing set). The top edges correlated with age included the previously reported areas. Regional efficiency, capturing regional information integration, was correlated with age positively in temporal and frontal regions and negatively in parietal and occipital regions previously [9]. Though it was hard to compare directly due to different atlases and different modalities, they also found superior temporal gyri (Figure 2, ST), and superior frontal gyri (SF) as we did. The edges with negative correlations were partially matched with white matter integrity (Fractional Anisotropy, FA) changes over age in a voxel based DTI study [17]: notably, both superior frontal gyri (SF), superior temporal gyri (ST), anterior and posterior cingulate cortices (CAC and PC). The abundance of strong negative correlations compared to the increased correlations is in line with decrease of white matter connectivity over ageing [18, 19] and preferential detachment of structural networks [12].

The concept behind our proposed method is a basic artificial neuron, a weighted sum of inputs (connectivity). We set its learnable weights with correlation coefficients we computed, capturing its influence on age. Statistically, this idea can be connected with the partial least square (PLS) regression [14]. PLS projects the original predictors (edges) into a lower dimensional space, using colinearity of each predictor variable
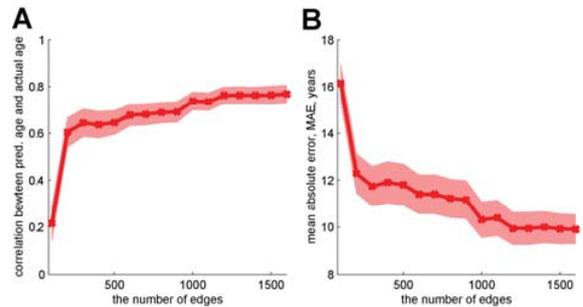

Fig. 3. Effects of the number of edges used in the model. A) correlation coefficients between predicted ages and actual ages, and B) MAE in training sets. We selected the top strong edges from 100 to 1600 edges systematically. Shades represents standard deviation of values over 100 bootstrapping.

TABLE II.   COMPARISON WITH OTHER PREDICTION METHODS

prediction method	connectivity normalization		
	L^0	L^1	L^2
CBR (proposed)	**0.791±0.032** **(9.499±0.674)** [b]	0.687±0.049 (11,636±0.807)	0.600±0.046 (14.387±0.909)
PLSR	**0.724±0.031** **(9.803±0.624)**	0.674±0.039 (10.284±0.675)	0.426 ±0.066 (13.506±0.821)
SVR	**0.772±0.028** **(10.259±0.666)**	0.631±0.045 (16.729±0.929)	0.531 ±0.043 (15.459±0.872)

[b.] *Mean correlation ± standard deviation (mean MAE ± standard deviation, years) in the testing set, computed through 100 bootstrapping of the dataset*

with the feature variable (age) to be fitted. In our method, a simple Pearson correlation coefficient, that represents fitting quality of linear regression, captured colinearity. Then, a weighted sum of a subject's connection strength represented projection of the subject into the predicting score P, where the vector of correlation coefficients was projection mapping, while PLS combines them in a statistically meaningful way.

We observed that non-normalization performed best in general, while between-subject normalization deteriorated its performance most. First of all, we noted that this normalization, L^2, was not standardization with the mean and standard deviation of each edge, but regularization of magnitudes to minimize effects of variability in edges' magnitudes. The vector of correlation coefficients captured the pattern of edges correlated with age. Because correlation coefficients captured the fitting quality (a clear relationship), not the magnitude of influence itself, we concerned that a few edges, whose overall magnitude is relatively larger than other edges, dominantly affected the prediction. So, we regularized edges with large average magnitudes by dividing the group mean of the edge (between-group normalization). Though the between group normalization (L^2) did not affect correlation coefficients, it performed worse than the within-group normalization (L^1), which deteriorated correlation. Thus, we believed that not only the clear relationship captured by the correlation coefficients, but also the magnitude of the edge was important. In other words, edges with larger magnitudes (preponderantly larger variance) had greater roles in predicting brain maturity than edges with smaller magnitudes.

To achieve the better prediction model, one may employ other MRI modality like cortical thickness [16], other features [20], or other prediction techniques [16, 20]. However, our proposed prediction model based on raw connectivity is advantageous: 1) simple to use (no need of parameter tuning), 2) with fair performance ($r = 0.79$ in the testing set), and 3) informative (providing local changes in relationship between brain regions). Our proposed method can be improved in various ways. First, we may use partial correlation coefficients to adjust group biases, or Spearman correlation coefficient to handle non-normality of connection strengths. Second, separating males and females may improve the performance as they may follow different developmental trajectories [9, 12].

V.   CONCLUSION

Using a simple predictor based on correlation between each edge weight and age, we were able to predict age of a subject given structural connectivity data with fair performance. The proposed method is applicable to any network with other modalities including functional connectivity. Since it does not require any parameter tuning or pre-conditions of applicability, it could be a first prediction model to be tested. We note that it can be improved in various ways still, including the use of partial correlation coefficients.

REFERENCES

[1] O. Sporns, D.R. Chialvo, M. Kaiser, and C.C. Hilgetag, "Organization, development and function of complex brain networks," Trends Cogn Sci, vol. 8, pp. 418-25, 2004.

[2] M. Kaiser, "A tutorial in connectome analysis: topological and spatial features of brain networks," Neuroimage, vol. 57, pp. 892-907, 2011.

[3] M.P. van den Heuvel, C.J. Stam, R.S. Kahn, and H.E. Hulshoff Pol, "Efficiency of functional brain networks and intellectual performance," Journal of Neuroscience, vol. 29, pp. 7619-7624, 2009.

[4] P. Shaw, et al., "Development of cortical surface area and gyrification in Attention-Deficit/Hyperactivity Disorder," Biol. Psych., vol. 72, No. 3, pp. 191-7, 2012.

[5] N.U.F. Dosenbach, et al., "Prediction of individual brain maturity using fMRI," Science, vol. 329, pp. 1358-1361, 2010.

[6] C.J. Honey, et al., "Predicting human resting-state functional connectivity from structural connectivity," PNAS, vol. 106, no. 6, pp. 2035-2040, 2009.

[7] L.F. Costa, M. Kaiser, abd C. Hilgetag, "Beyond the average: detecting global singular nodes from local features in complex networks," EPL (Europhysics Letters), vol. 87, p. 18008, 2009.

[8] C. Echtermeyer, et al., "Integrating temporal and spatial scales: human structural network motifs across age and region-of-interest size," Front Neuroinformatics, vol. 5, no. 10, Jul 2011.

[9] G. Gong et al., "Age- and Gender-Related Differences in the Cortical Anatomical Network," The Journal of Neuroscience, vol. 29, no. 50, pp. 15684-15693, December 2009.

[10] P. Hagmann, et al., "White matter maturation reshapes structural connectivity in the late developing human brain" PNAS, vol. 107, no. 44, pp. 19067-72, Oct 2010.

[11] K.B. Nooner, et al., "The NKI-Rockland Sample: A model for accelerating the pace of discovery science in psychiatry," Frontiers in Neuroscience, vol. 6, no.152, 2012.

[12] S. Lim, CE. Han, P.J. Uhlhaas, and M. Kaiser, "Preferential detachment during human brain development: Age- and sex-specific structural connectivity in Diffusion Tensor Imaging (DTI)," Cerebral Cortex, online ahead print, 2013.

[13] R. Wang, T. Benner, A. G. Sorensen, and V. J. Wedeen, "Diffusion Toolkit: a software package for diffusion imaging data processing and tractography" In. Proc. Intl. Soc. Mag. Reson. Med., p 3720, 2008.

[14] H. Drucker, C. Burges, L. Kaufman, A.J. Smola, and V.N. Vapnik, "Support vector regression machines", in Advances in Neural Information Processing Systems 9, NIPS 1996, pp. 155–161, 1997.

[15] S. Wold, A. Ruhe, H. Wold, and W. J. Dunn, III, "The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses," SIAM Journal on Scientific and Statistical Computing , vol. 5, no. 3, pp.735-743. 1984.

[16] K. Franke, G. Ziegler, S. Klöppel, C. Gaser, Alzheimer's Disease Neuroimaging Initiative, "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters," Neuroimage. vol. 50, no. 3, pp. 883-92, 2010.

[17] A. Stadlbauer, et al. "Magnetic resonance fiber density mapping of age-related white matter changes," European Journal of Radiology, vol. 81, pp. 4005-4012, 2012.

[18] L.T. Westlye, et al. "Life-span changes of the human brain white matter: diffusion tensor imaging (DTI) and volumetry," Cereb. Cortex, vol. 20, no. 9, pp. 2055-68.Sep 2010.

[19] C. Lebel, et al. "Diffusion tensor imaging of white matter tract evolution over the lifespan," Neuroimage, vol. 60, no. 1, pp. 340-52. Mar 2012.

[20] B. Mwangi, K.M. Hasan, and J.C. Soares, "Prediction of individual subject's age across the human lifespan using diffusion tensor imaging: A machine learning approach," Neuroimage, vol.75, pp. 58-67, 2013.