NeuroImage xxx (xxxx) xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Brain status modeling with non-negative projective dictionary learning

Mingli Zhang [a,*], Christian Desrosiers [b], Yuhong Guo [c], Budhachandra Khundrakpam [a], Noor Al-Sharif [a], Greg Kiar [a], Pedro Valdes-Sosa [d], Jean-Baptiste Poline [a,1], Alan Evans [a,1]

[a] Montreal Neurological Institute, McGill University, Montreal, H3A 2B4, Canada
[b] Department of Software and IT Engineering, École de Technologie supérieure (ETS), Montreal, H3C 1K3, Canada
[c] School of Computer Science, Carleton University, Canada
[d] University of Electronic Science and Technology of China/ Cuban Neuroscience Center, China

ARTICLE INFO

Keywords:
Brain maturity modeling
Brain age prediction
Projective dictionary learning
Cognitive biomarker

ABSTRACT

Accurate prediction of individuals' brain age is critical to establish a baseline for normal brain development. This study proposes to model brain development with a novel non-negative projective dictionary learning (NPDL) approach, which learns a discriminative representation of multi-modal neuroimaging data for predicting brain age. Our approach encodes the variability of subjects in different age groups using separate dictionaries, projecting features into a low-dimensional manifold such that information is preserved only for the corresponding age group. The proposed framework improves upon previous discriminative dictionary learning methods by incorporating orthogonality and non-negativity constraints, which remove representation redundancy and perform implicit feature selection. We study brain development on multi-modal brain imaging data from the PING dataset ($N = 841$, age $= 3 - 21$ years). The proposed analysis uses our NDPL framework to predict the age of subjects based on cortical measures from T1-weighted MRI and connectome from diffusion weighted imaging (DWI). We also investigate the association between age prediction and cognition, and study the influence of gender on prediction accuracy. Experimental results demonstrate the usefulness of NDPL for modeling brain development.

1. Introduction

Brain development is a dynamic and complex process lasting the duration of childhood, adolescence and the early young adulthood. Brain development and aging are highly variable and have a significant impact on cognition and the occurrence of age-related diseases. Modeling this process could therefore be essential to identify subjects with a high risk of deterioration, estimate the progress of cognitive decline over time, and help select optimal treatment. Investigations on brain maturity (or brain age) have been facilitated by the development of advanced magnetic resonance imaging (MRI) methods (Liem et al., 2017; Cole and Franke, 2017; Dosenbach et al., 2010), as well as by large-scale initiatives like the Brain Development Cooperative Group (NIHPD) (Evans Groupet al, 2006) and the Pediatric Imaging, Neurocognition, and Genetics (PING) (Walhovd et al., 2012) study. Cross-sectional and longitudinal neuroimaging studies based on MRI have shown developmental trajectories of gray matter volumes, surface area and cortical thickness, contributing to a better understanding of brain maturation. The prediction of brain maturity has also been studied for the early diagnosis of neuropsychiatric diseases (Cole and Franke, 2017; Franke et al., 2012; Khundrakpam et al., 2015; Ball et al., 2017). However, despite significant research efforts, modeling brain development and aging remain a challenging task.

An effective way to analyze brain development (maturity) is training a statistical model on lifespan data to estimate brain age, and then comparing a subject's chronological age with the one predicted by the model. Differences between chronological and predicted age can suggest a risk of developing a neurodevelopmental disorder, an information which can later be used by clinicians for diagnosis. Voxel-based morphometry (VBM) analysis (Franke et al., 2012) of T1-weighted MRI data is one of the most common approaches for predicting brain maturity. It has been shown that brain maturation and cognitive developmental trajectories are associated with GM volumes and cortical thickness in highly-localized brain regions (Giedd et al., 1999; Shaw et al., 2006). Recently, machine learning has shown a great potential for modeling and predicting brain maturity (Cole and Franke, 2017). For instance, Ball et al. (2017) proposed manifold embedding as a means to learn a

* Corresponding author.
E-mail address: mingli.zhang@mcgill.ca (M. Zhang).
[1] co-last authorship.

https://doi.org/10.1016/j.neuroimage.2019.116226
Received 29 March 2019; Received in revised form 1 September 2019; Accepted 24 September 2019
Available online xxxx

compact representation that captures variability in brain ageing. Likewise, Cole et al. (2017a) used deep learning to predict brain age from raw T1-weighted MRI, achieving a mean absolute error (MAE) of 4.65 years for a cohort of 2001 healthy subjects with mean age of 36.95 ± 18.12 and age range of 18 − 90 years.

In addition to structural MRI, other modalities like resting-state functional MRI (Barnes et al., 2010) and diffusion tensor imaging (Erus et al., 2014) have also been explored for predicting brain age. For instance, connectivity maps extracted from resting-state fMRI were shown to provide unique information on brain development (Fair et al., 2009). Studies employing diffusion tensor imaging (DTI) for the analysis of brain development and aging have shown age-related differences in apparent diffusion coefficient (ADC) and fractional anisotropy (FA) (Pfefferbaum et al., 2000), potentially due to changes in myelin density and structure of brain fibers (Salat et al., 2005). Several studies have further investigated the benefit of combining data from multiple MRI modalities to predict brain maturity (Liem et al., 2017; Brown et al., 2012). Distinct MRI modalities contain complementary features, which can be combined to provide richer information on the brain maturation process. However, determining the best way to aggregate high-dimensional features from separate modalities can be challenging. Hence, combining them directly can lead to over-fitting of the prediction model and may complicate the analysis of results. Although several works have focused on predicting brain maturity (Cole and Franke, 2017; Franke et al., 2012; Khundrakpam et al., 2015; Cole et al., 2017b; Adeli et al., 2018), few of them have analyzed the individual contributions of morphological measures from structural MRI, such as gray matter (GM) volume, cortical thickness and surface area. Further, no study has proposed a joint model combining these measures with other modalities like diffusion MRI.

Discriminative subspace learning is a powerful approach for fusing multi-modal data into a common feature space (Li et al., 2017; Zhang and Desrosiers, 2018a). This approach, which has been used in various brain-related applications such as the detection of Alzheimer's disease (Zhu et al., 2016) and the classification of EEG signals (Zhou et al., 2012), exploits the low rank property of high dimensional data (Zhang et al., 2018a, b; Zhang and Desrosiers, 2018b). The current study proposes a discriminative subspace learning method, called Non-negative Projective Dictionary Learning (NPDL) for predicting brain age. Based on the projective Dictionary Pair Learning (DPL) technique (Zhang et al., 2018a, b Evans; Gu et al. 2014 Feng), our method seeks, for each age group, a feature subspace which can accurately encode examples of a specific age group, but not those of other age groups. In the proposed method, class-specific subspaces are modeled as a pair of dictionaries, one to map image features to a discriminative representation, and another one to reconstruct these features from the representation. Our method differs from unsupervised subspace learning approaches like autoencoders (Hinton and Salakhutdinov, 2006), where input feature vectors are mapped to a compact and information-preserving representation, without considering their class. The major contributions of this work are as follows:

- **Novel framework:** Our framework extends the algorithms of Zhang et al. (Zhang et al., 2018a, b Evans) and Gu et al. (Gu et al. 2014. Feng) in two important ways. While DPL applies simple norm constraints on the dictionary elements, we also impose these elements to be uncorrelated. In addition to helping avoid overfitting, this facilitates the visualization and understanding of features learned in the dictionary, as these features encode orthogonal components of the data. Moreover, unlike DPL, our framework also enforces non-negativity on the projected features. Non-negative constraints have been shown to be useful in numerous optimization problems like matrix factorization (Eskildsen et al., 2015). In our case, these constraints enforce sparsity in the encoding coefficients (see Section 3.3), thereby acting as an implicit feature selection technique. Furthermore, since these coefficients are non-negative, the prediction corresponds to an affine combination of features, which is more amenable to interpretation than a general linear combination.
- **Application:** We evaluate our NPDL framework on the prediction of brain maturity, using data from the PING study (Walhovd et al., 2012), and show a higher accuracy compared to DPL and the model without non-negativity constraints. Our experiments analyze features of gray and white matter which are important for predicting brain age, as well as the influence of gender on the prediction. The association between the predicted brain age and cognition is also investigated with the proposed NPDL model.

Table 1
Demographic details about the subjects applied in this work.

Total number of subjects: (male/female)	841 (433/408)
Total number of scans:	1682
Total number of acquisition sites:	12
Age range	3-21y
Age	12.55 (± 4.99)
Reading	138.19 (± 68.34)
Flanker	8.99 (± 1.49)
Attention	9.43 (± 1.35)
DCCS.	8.79 (± 1.39)

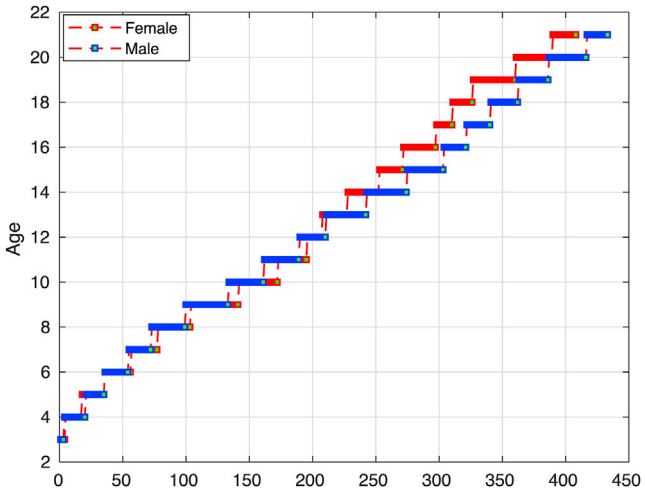


Fig. 1. Distribution of female and male subjects sorted by increasing age.

2. Materials: data and pre-processing

Our study uses 3 T T1-weighted MRI and diffusion tensor image (DTI) data from the PING study.[2] The PING project created a comprehensive, publicly shared, multi-modal data resource for studying standardized assessments of neurocognition, neuroimaging and genetics in a deliberately diverse cohort of typically-developing individuals. Cross-sectional measurements on 1493 individuals, aged between 3 and 21 years, were aggregated from different sites and scanners across the United States. A detailed description of the cohort can be found in (Jernigan et al., 2016; Akshoomoff et al., 2014). Subjects' information related to neurological disorders, history of trauma, preterm birth, autism/bipolar disorder, pregnancy and medical history were obtained from parental questionnaires. Similar proportions of population in each age group and gender participated across the entire age range. Neurocognitive abilities were assessed using the NIH Toolbox Cognition Battery (NTCB[3]), a computerized battery designed for administration across the lifespan (Akshoomoff et al., 2014), which includes eight subtests spanning six

[2] http://pingstudy.ucsd.edu/Data.php.
[3] http://www.nihtoolbox.org/.

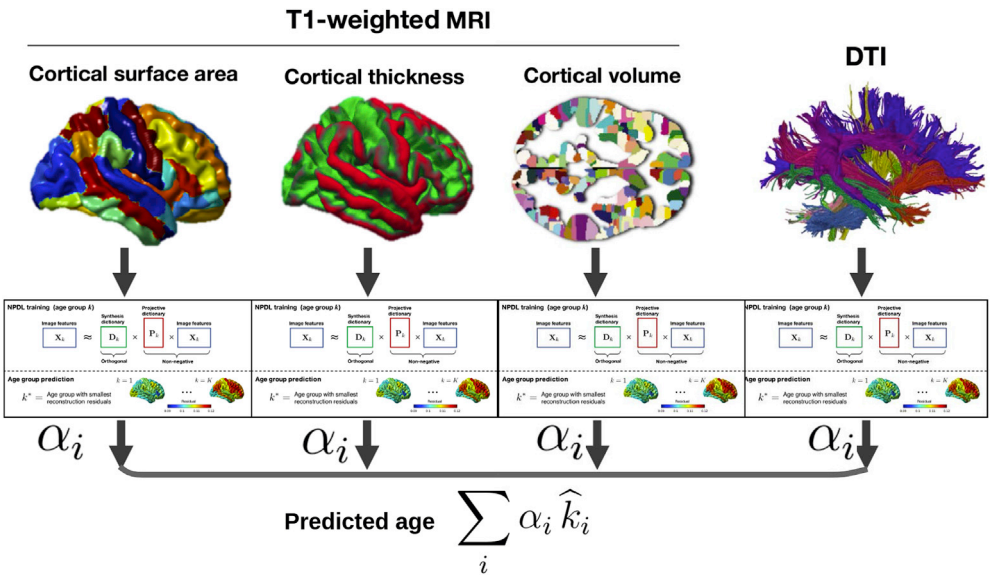


Fig. 2. The proposed non-negative projective dictionary learning (NPDL) method for predicting brain maturity.

domains. In this study, we included the following four measures of executive function: Flanker Inhibitory Control (Flanker), Attention Test (Attention), Dimensional Change Card Sort (DCCS), and language measures of the Oral Reading Recognition tests (Reading). The Demographic details of the subjects and the age group distribution with gender are shown in Table 1 and Fig. 1.

For image acquisition, all institutions participating in the PING study used a standardized structural MRI protocol on experimental, consenting procedures and methods (Jernigan et al., 2016). This protocol included a 3D T1-weighted inversion prepared RF-spoiled gradient echo scan using prospective motion correction (PROMO), for cortical and subcortical segmentation, 3D T2-weighted variable flip angle fast spin echo scan using PROMO, for detection and quantification of white matter lesions and segmentation of CSF.

MRI images were pre-processed using the CIVET[4] pipeline version 2.1.0. T1-weighted images are first non-uniformity corrected with the N3 algorithm (distance 125) (Sled et al., 1998) and a linear transformation (Collins et al., 1994), and then linearly registered to the Talairach-like MNI152 template (established with the ICBM152 dataset). Non-uniformity correction is then repeated using the template mask. A non-linear registration method is then applied to further align the resulting volume to the MNI152 template. Following this, the registered volume is segmented into gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and background, using a neural net classifier (Zijdenbos et al., 2002), and corrected for partial volume effects (Tohka et al., 2004). Inner and outer gray matter (GM) surfaces are then extracted using the Constrained Laplacian-based Automated Segmentation with Proximities (CLASP) algorithm (Kim et al., 2005), and cortical thickness is measured in native space using the linked distance between the two surfaces at 81,924 vertices. Each subject's cortical thickness map was blurred using a 30-mm full width at half maximum surface-based diffusion smoothing kernel to impose a normal distribution on the corticometric data, and to increase the signal to noise ratio. The flowchart of T1 MRI and DWI data pre-processing is shown in Appendix.

Quality control (QC) of these data was implemented by two independent reviewers, which gave a rating of *failed* = 0, *questionable* = 1 or *passed* = 2 based on a low signal to noise ratio, motion artifacts, artifacts of hyperintensities from blood vessels, surface-surface intersections, or poor placement of the gray or white matter (GM and WM) surface.

[4] http://www.bic.mni.mcgill.ca/ServicesSoftware/CIVET.

3. The proposed method

3.1. Input data

The proposed model for predicting brain age is trained with cortical anatomy from T1-weighted MRI and structural connectivity features (structural connectome) from DWI. Three types of cortical surface measures as considered as input features: cortical thickness, surface area and volume. Considering the time complexity and efficiency of the proposed model, these measures are averaged within each of the 31 cortical parcels of the DKT40 parcellation atlas (Desikan et al., 2006), yielding a total of 62 features per type of measure for both brain hemispheres.

Structural connectomes were derived from diffusion weighted imaging using the NDMG pipeline (Kiar et al., 2017), which consists of four steps: registration, tensor estimation, tractography and graph generation. In the registration step, the diffusion weighted tensor imaging (DTI) data is aligned to the MNI152 template. In the tensor estimation stage, the 6-component tensor model of the DIPY package (Garyfallidis et al., 2014) is then applied to get fractional anisotropy maps, which can reflect the structure and arrangement of the tensors. The tractography stage then generates and prunes streamlines. Lastly, in the graph generation stage, the connectivity between all pairs of regions is quantified as a symmetric connectivity matrix. Note that measures of connectomes vary based on the parcellation resolution, and that there is no standard number of parcels for real application (Liem et al., 2017). In this study, we reconstructed connectivity matrices using 68 brain regions. To extract relevant information and make the dimension of DWI-based features consistent with cortical features from structural MRI, we applied singular value decomposition (SVD) on the 68 × 68 DWI connectivity matrices and selected the first 62 singular vectors as structural connectivity features. Most information on connectivity is captured in the first few components, hence using this subset of the 68 singular vectors does not affect results.

3.2. Predictive analysis

The proposed method is summarized in the flowchart of Fig. 2. In the first step, the features of each subject are obtained from measures of cortical thickness, surface area and volume, as well as from the DWI connectivity matrix. Feature vectors corresponding to each feature type are then normalized to have an l_1 norm of one. This normalization step, which is similar to global correction for brain size (Ball et al., 2017), is employed to emphasize the analysis of relative differences across gray

Input: The training features $\mathbf{X}_k$, $k = 1, \ldots, K$ and age group as label;
Output: The analysis dictionary $\mathbf{P}_k \in \mathbf{R}^{M \times S}$ and synthesis dictionary $\mathbf{D}_k \in \mathbf{R}^{S \times M}$

Set $\mathbf{A}_k = \mathbf{P}_k \mathbf{X}_k$ and $\mathbf{Z}_k := 0$, $k = 1, \ldots, K$;

while *not converged* **do**

Find groups of each feature with same label k;
Update $\mathbf{P}_k$, $k = 1, \ldots, K$, using Eq. (4);
Update $\mathbf{D}_k$ using Eq. (5);
Update $\mathbf{A}_k$, by solving Eq. (7);
Update Lagrange multiplier $\mathbf{Z}_k$ using $\mathbf{Z}_k := \mathbf{Z}_k + (\mathbf{A}_k - \mathbf{P}_k \mathbf{X}_k)$;

return $\mathbf{P}_k$ *and* $\mathbf{D}_k$;

matter regions, and their relationship with brain development.

The data is split into training and testing sets using a 10-fold cross-validation strategy. During each fold, we train a NPDL model for each feature type (i.e., cortical surface area, thickness, volume, and singular values of diffusion connectivity matrices) and age group. The reconstruction error measured for the models of different age groups is then used to predict brain age. Furthermore, predictions obtained for each feature type are then combined via linear regression to obtain a final predicted age. Accuracy is assessed with the mean absolute error (MAE), root mean squared error (RMSE) and correlation (R). Finally, the performance measures are averaged across all 10 folds.

3.3. Brain age prediction with non-negative projective dictionary learning (NPDL)

We treat brain age prediction on each feature type as a general classification problem over K classes (age groups). Let $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_k, \cdots, \mathbf{X}_K]$ denote the data (feature) samples, where $\mathbf{X}_k \in \mathbf{R}^{S \times N_k}$ contains the data samples belonging to the k-th age class with S the number of features. The total number of subjects is given by $N = \sum_k N_k$. We first discriminatively model the data by performing separate projective dictionary learning on the subsamples from each class. Specifically, we introduce an analysis dictionary $\mathbf{P}_k \in \mathbf{R}^{M \times S}$ and a synthesis dictionary $\mathbf{D}_k \in \mathbf{R}^{S \times M}$ for each age class k, such that $\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_k, \cdots, \mathbf{P}_K]$ and $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_k, \cdots, \mathbf{D}_K]$, where M is the dimension of the lower dimensional space. We then model data using the following class-specific dictionary pair learning framework:

$$\min_{\mathbf{P},\mathbf{D}} \quad \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{D}_k \mathbf{P}_k \mathbf{X}_k\|_F^2 + \lambda \|\mathbf{P}_k \widehat{\mathbf{X}}_k\|_F^2, \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm and $\widehat{\mathbf{X}}_k$ denotes the complementary data matrix of $\mathbf{X}_k$ in $\mathbf{X}$, i.e. $\widehat{\mathbf{X}}_k = [\mathbf{X}_1, \cdots, \mathbf{X}_{k-1}, \mathbf{X}_{k+1}, \cdots, \mathbf{X}_K]$. In this framework, the regularization term $\|\mathbf{P}_k \widehat{\mathbf{X}}_k\|_F^2$ is used to ensure class-specific dictionary learning by pushing $\mathbf{P}_k \mathbf{X}_j$ towards zero over data samples from any class j such that $j \neq k$. The analysis dictionary $\mathbf{P}_k$ projects the samples $\mathbf{X}_k$ into an encoding coefficient matrix $\mathbf{A}_k = \mathbf{P}_k \mathbf{X}_k$, which is then used to reconstruct $\mathbf{X}_k$ with the synthesis dictionary $\mathbf{D}_k$. Parameter $\lambda > 0$ controls the trade-off between the reconstruction accuracy and regularization terms.

To avoid overfitting, dictionary pair learning methods like (Gu et al. 2014. Feng) normally use a Frobenius norm regularization term over $\mathbf{D}_k$ or a Euclidean norm constraint over each column of this dictionary. Here, we aim to identify a compact synthesis dictionary $\mathbf{D}_k$ with uncorrelated dictionary basis vectors. Toward this goal, we impose an orthogonality constraint over the dictionary: $\mathbf{D}_k^\top \mathbf{D}_k = \mathbf{I}$, for $k = 1, \ldots, K$. Moreover, we further assume the basis vectors of $\mathbf{D}_k$ model the representative latent components of $\mathbf{X}_k$. Hence, $\mathbf{X}_k$ can be taken as an additive combination of these components by enforcing the encoding coefficients $\mathbf{A}_k = \mathbf{P}_k \mathbf{X}_k$ to be non-negative. This leads to the following non-negative projective dictionary learning (NPDL) problem:

$$\begin{aligned} \min_{\mathbf{P},\mathbf{D}} \quad & \sum_{k=1} \|\mathbf{X}_k - \mathbf{D}_k \mathbf{P}_k \mathbf{X}_k\|_F^2 + \lambda \|\mathbf{P}_k \widehat{\mathbf{X}}_k\|_F^2 \\ \text{s.t.} \quad & \mathbf{D}_k^\top \mathbf{D}_k = \mathbf{I}, \quad \mathbf{P}_k \mathbf{X}_k \geq 0, \; k = 1, \ldots, K. \end{aligned} \tag{2}$$

The next section presents an efficient optimization approach to learn dictionary sets $\mathbf{D}$ and $\mathbf{P}$ from training data.

3.4. Training the model

We solve the problem in (2) using an alternating direction method of multipliers (ADMM). Specifically, we introduce an explicit encoding coefficient matrix $\mathbf{A}_k$ with the equality constraint $\mathbf{A}_k = \mathbf{P}_k \mathbf{X}_k$ for each class k, and then incorporate such constraints into the objective with auxiliary variable matrices $\mathbf{Z}_k$. This leads to a minimization of the following augmented Lagrangian function:

$$\begin{aligned} \max_{\mathbf{Z}} \; \min_{\mathbf{P},\mathbf{D},\mathbf{A}} \quad & \sum_{k=1} \|\mathbf{X}_k - \mathbf{D}_k \mathbf{P}_k \mathbf{X}_k\|_F^2 + \lambda \|\mathbf{P}_k \widehat{\mathbf{X}}_k\|_F^2 + \mu \|\mathbf{A}_k - \mathbf{P}_k \mathbf{X}_k + \mathbf{Z}_k\|_F^2 \\ \text{s.t.} \quad & \mathbf{D}_k^\top \mathbf{D}_k = \mathbf{I}, \quad \mathbf{A}_k \geq 0, \; k = 1, \ldots, K. \end{aligned} \tag{3}$$

The ADMM algorithm then performs optimization in an iterative manner. In each iteration, it alternatively updates each variable matrix given the other variable matrices fixed, as follows. With fixed $\{\mathbf{D}, \mathbf{A}, \mathbf{Z}\}$, the minimization over each $\mathbf{P}_k$ is an unconstrained quadratic minimization problem,

$$\min_{\mathbf{P}_k} \; \|\mathbf{X}_k - \mathbf{D}_k \mathbf{P}_k \mathbf{X}_k\|_F^2 + \lambda \|\mathbf{P}_k \widehat{\mathbf{X}}_k\|_F^2 + \mu \|\mathbf{P}_k \mathbf{X}_k - (\mathbf{A}_k + \mathbf{Z}_k)\|_F^2$$

which has the following closed-form solution:

$$\mathbf{P}_k = \left(\mathbf{D}_k^\top \mathbf{X}_k \mathbf{X}_k^\top + \mu(\mathbf{A}_k + \mathbf{Z}_k)\mathbf{X}_k^\top\right)\left((1+\mu)\mathbf{X}_k \mathbf{X}_k^\top + \lambda \widehat{\mathbf{X}}_k \widehat{\mathbf{X}}_k^\top + \gamma \mathbf{I}\right)^{-1} \tag{4}$$

Here, γ is a small constant used to increase invertibility. Following (Gu et al. 2014. Feng), we set this parameter to $\gamma = 10^{-4}$.

Considering $\{\mathbf{A}, \mathbf{P}, \mathbf{Z}\}$ as fixed, updating each $\mathbf{D}_k$ is an orthogonal constrained optimization problem

$$\min_{\mathbf{D}_k} \; \|\mathbf{X}_k - \mathbf{D}_k \mathbf{P}_k \mathbf{X}_k\|_F^2, \; \text{s.t.} \; \mathbf{D}_k^\top \mathbf{D}_k = \mathbf{I}$$

which can be equivalently rewritten into

$$\max_{\mathbf{D}_k} \operatorname{tr}\left(\mathbf{D}_k^\top \mathbf{X}_k \mathbf{X}_k^\top \mathbf{P}_k^\top\right), \; \text{s.t.} \; \mathbf{D}_k^\top \mathbf{D}_k = \mathbf{I}, \tag{5}$$

Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the singular value decomposition (SVD) of $\mathbf{X}_k \mathbf{X}_k^\top \mathbf{P}_k^\top$. Then, problem (5) has the following closed-form solution: $\mathbf{D}_k = \mathbf{U}\mathbf{V}^\top$.

To update each $\mathbf{A}_k$, we fix $\{\mathbf{D}, \mathbf{P}, \mathbf{Z}\}$ and solve a non-negative constrained minimization problem

$$\min_{\mathbf{A}_k} \|\mathbf{A}_k - (\mathbf{P}_k \mathbf{X}_k - \mathbf{Z}_k)\|_F^2, \; \text{s.t.} \; \mathbf{A}_k \geq 0, \tag{6}$$

the solution of which is obtained with the hard-thresholding operator:

$$[\mathbf{A}_k]_{ij} = \max\left([\mathbf{P}_k\mathbf{X}_k - \mathbf{Z}_k]_{ij},\ 0\right). \quad (7)$$

As mentioned before, having non-negative constraints acts as an implicit feature selection technique, where only features for which $[\mathbf{P}_k\mathbf{X}_k - \mathbf{Z}_k]_{ij} \geq 0$ are kept. Note that, for the case where $[\mathbf{Z}_k]_{ij} \approx 0$, this amounts to retaining only features that are positively correlated to the corresponding dictionary atom. Finally, we update the dual variable $\mathbf{Z}_k$ by maximizing the objective of Eq. (3), which can be achieved with $\mathbf{Z}_k := \mathbf{Z}_k + (\mathbf{A}_k - \mathbf{P}_k\mathbf{X}_k)$. The summary of the algorithm is listed in Algorithm 1.

Algorithm 1. Non-negative projective dictionary learning(NPDL)

3.5. Algorithm complexity and convergence

The computational complexity of the proposed dictionary learning framework is as follows. For updating dictionaries $\mathbf{P}_k$ with (4), since the matrix to invert is constant, its inverse can be computed only once in pre-processing. The complexity of this update step is therefore in $\mathcal{O}(KM(S + \max_k\{N_k\}))$. Next, updating each $\mathbf{D}_k$ requires to calculate the SVD of $S\times M$ matrix $\mathbf{X}_k\mathbf{X}_k^\top\mathbf{P}_k^\top$. If the QR decomposition of $\mathbf{X}_k\mathbf{X}_k^\top$ is pre-computed, assuming that $M \leq S$, this step has a total complexity of $\mathcal{O}(KM^3)$. Finally, updating all matrices $\mathbf{A}_k$ and $\mathbf{Z}_k$ can be done in $\mathcal{O}(KM\max_k\{N_k\})$ operations.

The convergence of augmented Lagrangian methods for biconvex problems like the one in this work has been well-studied in the literature (e.g., see (Gu et al. Feng; Gorski et al., 2007)). Since $\mathbf{D}_k^\top\mathbf{D}_k = \mathbf{I}$ and $\mathbf{P}_k\mathbf{X}_k \geq 0$ are both convex sets and the cost function is biconvex, our problem can be formulated in the general biconvex problem definition of Eq. (1) in (Gorski et al., 2007), for which optimizing convex subproblems alternatively is shown to converge. Empirically, we observed a smooth convergence of the method within 10–20 iterations.

3.6. Prediction on test samples

After training, the learned dictionaries can be used to classify new samples by measuring the reconstruction error for each class. We first consider the case of individual feature types (i.e., cortical surface area, thickness, volume and singular vectors of connectivity matrices). Let $\mathbf{x}^i \in \mathbf{R}^{S_i}$ be the features of type i for the sample to classify. We define as $e_k^i = \left\|\mathbf{x}^i - \mathbf{D}_k^i\mathbf{P}_k^i\mathbf{x}^i\right\|_2$ the error of reconstructing $\mathbf{x}^i$ with the dictionaries of class k for feature type i. We then assign the sample to the class whose dictionary gives the lowest error, i.e. $\widehat{k}_i = \text{argmin}_k\ e_k^i$.

To combine the information of multiple feature types, we need to determine their relative importance on the final brain age prediction. To achieve this goal, we use a subset of training examples (our validation set) to learn a regression model where inputs are the predicted ages $\widehat{k}_i$ for each feature type i and the output is the true subject age k_{real}.

$$\min_{\boldsymbol{\alpha}} \left(k_{\text{real}} - \sum_i \alpha_i\,\widehat{k}_i\right)^2, \quad \text{s.t.} \quad \sum_i \alpha_i = 1,\ \alpha_i \geq 0, \forall i. \quad (8)$$

Constraints on regression coefficients α_i enforce the final prediction to be a convex combination of predicted values for each feature type.

4. Experiments and results

We assess the usefulness of our NPDL framework on the task of predicting the brain age of subjects from the PING dataset. Our experiments also analyze the relevance of individual feature types on prediction accuracy, the relationship between predicted age and cognition, as well as the impact of gender on results.

4.1. Brain age prediction

We first evaluate the proposed framework on the task of predicting the chronological age of PING subjects based on the combination of all imaging features. We measure the contribution of orthogonality and non-negativity constraints in our NPDL model by comparing results against those of DPL (Gu et al. Feng), which is equivalent to our model without these constraints, and those of NPDL without non-negativity constraints. The same pre-processing and feature extraction pipelines were applied for all tested methods.

Table 2
Age prediction accuracy in terms of mean RMSE, obtained for subjects of different age ranges.

	Age range (N. of subjects)			
	3–10y (333)	3–14 (526)	3–18y (688)	3–21y (841)
DPL [27]	2.102 ± 0.025	2.803 ± 0.074	3.412 ± 0.063	3.996 ± 0.113
NPDL (w/o non-neg)	1.745 ± 0.041	2.554 ± 0.035	2.965 ± 0.052	3.324 ± 0.026
NPDL (ours)	1.720 ± 0.022	2.549 ± 0.046	2.959 ± 0.030	3.286 ± 0.056

Table 3
Age prediction accuracy in terms of mean MAE, obtained for subjects of different age ranges.

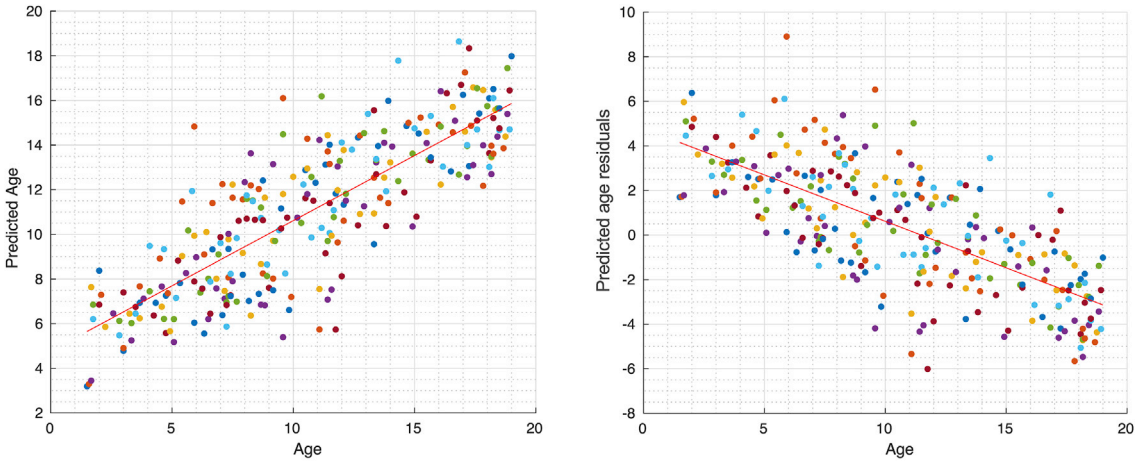
	Age range (N. of subjects)			
	3–10y (333)	3–14y (526)	3–18y (688)	3–21y (841)
DPL [27]	1.629 ± 0.019	2.192 ± 0.062	2.633 ± 0.043	2.985 ± 0.086
NPDL (w/o non-neg)	1.398 ± 0.039	2.035 ± 0.033	2.399 ± 0.049	2.615 ± 0.035
NPDL (ours)	1.370 ± 0.020	2.027 ± 0.037	2.385 ± 0.039	2.614 ± 0.036

Table 4
Mean correlation coefficient (R) between real and predicted age, for different age ranges.

	Age range (N. of subjects)			
	3–10y (333)	3–14y (526)	3–18y (688)	3–21y (841)
DPL [27]	0.616 ± 0.022	0.732 ± 0.021	0.786 ± 0.012	0.840 ± 0.012
NPDL (w/o non-neg)	0.743 ± 0.019	0.781 ± 0.007	0.830 ± 0.011	0.860 ± 0.002
NPDL (ours)	0.752 ± 0.007	0.782 ± 0.008	0.828 ± 0.004	0.863 ± 0.005

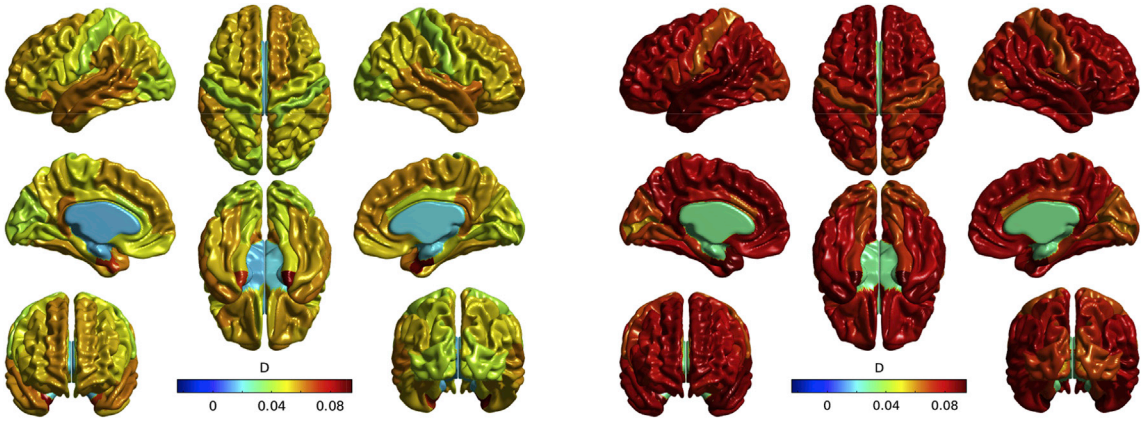
Tables 2–4 give the prediction accuracy in terms of mean RMSE, MAE and correlation coefficient (R). To evaluate the impact of feature variability on prediction, we report results obtained for training and testing sets containing subjects within increasing age ranges. Specifically, four age ranges are considered: 3–10 years, 3–14 years, 3–18 years and 3–21 years, the number of subjects in each age range are indicated in the tables. The last age range (i.e., 3–21 years) contains all 841 subjects of the PING dataset. As can be seen, NPDL achieves a higher accuracy than DPL or the proposed method without non-negativity constraints. In comparison with DPL, using NDPL without non-negativity constraints leads to a reduction of 16.82% in RMSE and 12.40% in MAE, for the age range including all subjects. Likewise, also imposing non-negativity yields a reduction of 17.77% in RMSE and 12.43% in MAE, compared to DPL. As expected, a lower prediction accuracy is observed for broader age ranges due to the more challenging regression problem. In contrast, all methods give a higher correlation for larger age ranges, reflecting the fact that the overall trend is easier to determine over the complete process of brain development.

Plots of predicted age and prediction residuals (predicted age minus real age) obtained by our model on the entire cohort of 841 subjects are shown in Fig. 3. We see that the model learns the general relationship



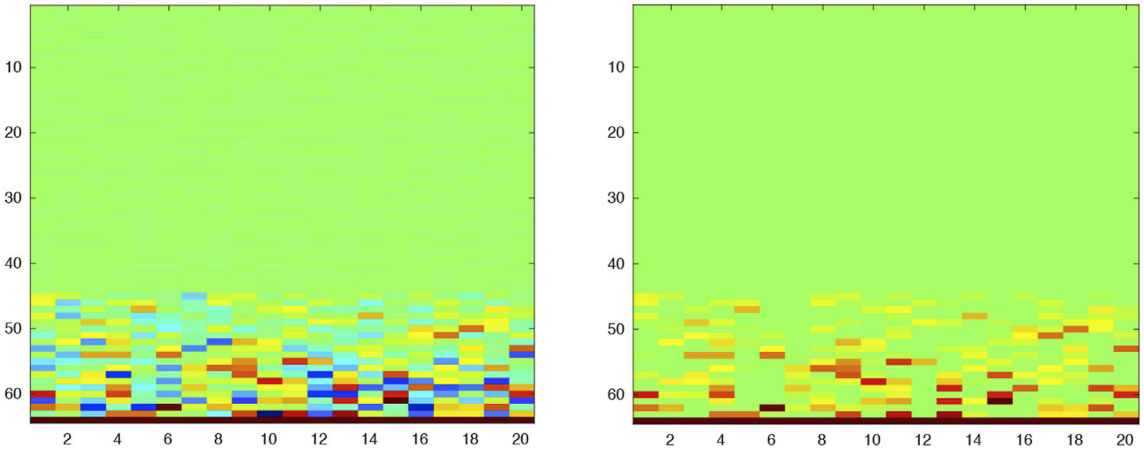
(a) Predicted age (b) Predicted age residuals

Fig. 3. Predicted age and prediction residuals versus age.



a) Same age class (RMSE=0.0541) b) Different age class (RMSE=0.0791)

Fig. 4. Example of cortical thickness reconstruction residuals for a 5 year old subject obtained with a) the dictionary of the same age class, and b) the dictionary corresponding to age 6.



(a) Without non-negativity (b) With non-negativity

Fig. 5. Example of coefficients matrix **A** for age class 5, with and without non-negativity constraints.

between brain features and age, resulting in a high correlation between real and predicted values. However, when considering prediction residuals, we observe a negative slope where the age of younger subjects is overestimated and the age of older ones is underestimated. This can be attributed to the well-known phenomenon of regression toward the mean, reported in several studies on brain development (Erus et al.,

Table 5
Age prediction accuracy, in terms of mean RMSE and MAE, obtained by NDPL on all subjects, using different feature types.

	Surface area	Thickness	Volume	DWI	All
RMSE	5.149 ± 0.087	3.806 ± 0.029	4.637 ± 0.113	5.698 ± 0.047	3.286 ± 0.056
MAE	3.946 ± 0.077	2.799 ± 0.026	3.507 ± 0.078	4.464 ± 0.038	2.614 ± 0.036

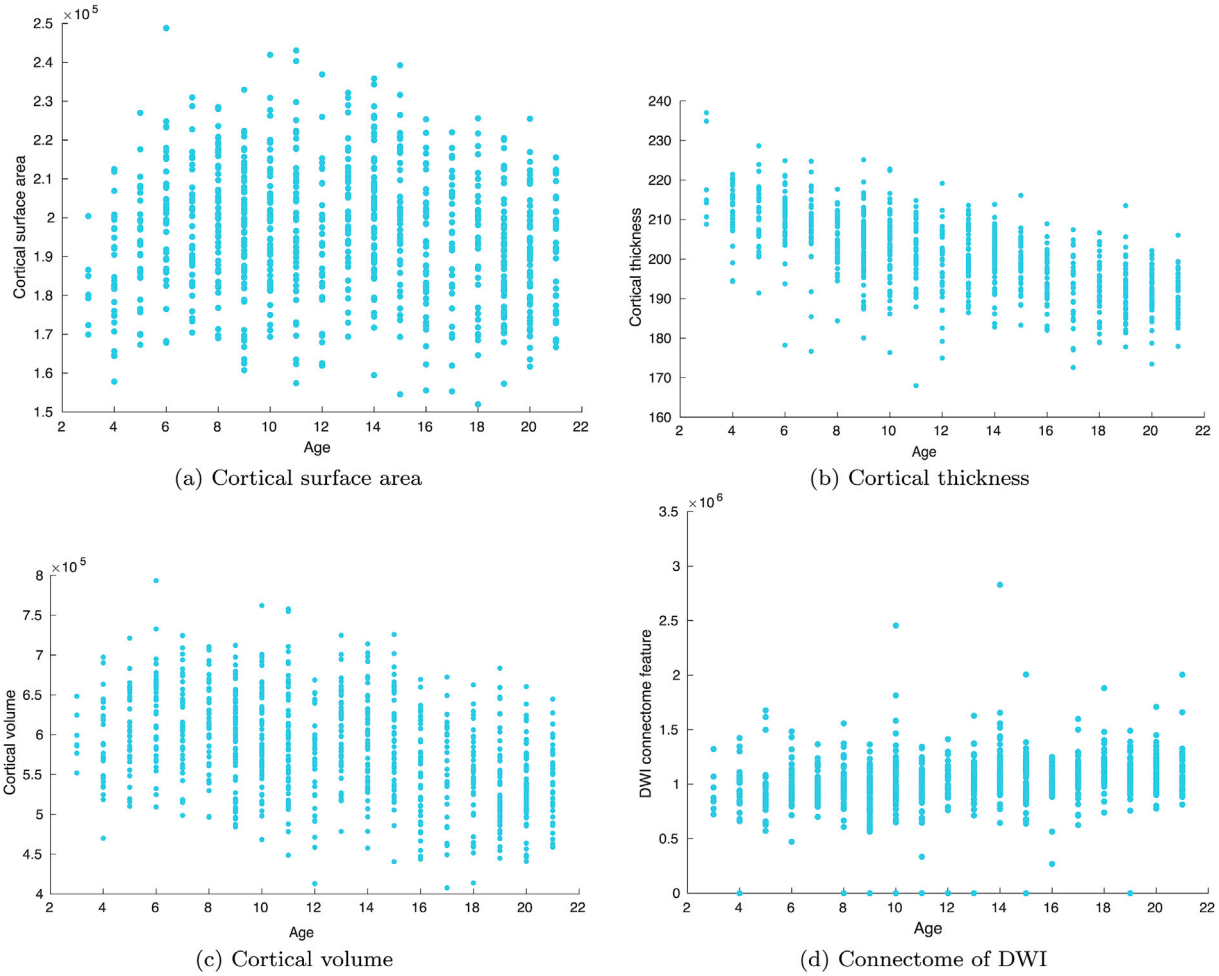
2014; Varikuti et al., 2018).

To illustrate how classification works, we show in Fig. 4 an example of cortical thickness reconstruction residuals (absolute value) for a subject of age 5, obtained with the same age class dictionary or with a dictionary corresponding to age 6. We can see that the dictionary corresponding to the subject's real age leads to a smaller residual over the whole cortex, with an RMSE of 0.0541 compared to 0.0791 for the dictionary of age 6. The impact of imposing non-negativity on projected features is illustrated in Fig. 5, showing an example of feature matrix **A** obtained by our NPDL model and the model without non-negativity constraints. As can be seen, enforcing features to be non-negative yields a sparser representation, which illustrates the ability of our model to act as a feature selection method.

4.2. Impact of features

In order to evaluate the usefulness of each feature type (i.e., cortical surface area, thickness, volume and singular vectors of DWI connectivity matrix) for predicting brain age, we measured the relative performance of NPDL trained with these different sets of features. Table 5 reports the mean RMSE and MAE for predicting the brain age of all 841 PING subjects, using individual feature types or the combination of all features. Comparing features types, we see that cortical thickness features yield the lowest error (mean RMSE of 3.806), followed by cortical volume (mean RMSE of 4.637). The worst prediction is obtained for structural connectivity based on DWI (mean RMSE of 5.698). Moreover, we observe that combining all feature types, as described in Section 3.6, yields a significantly better prediction, with a relative reduction of 13.66% in RMSE and 6.61% in MAE, compared to employing only cortical thickness features. To better understand the predictive value of each type of features, Fig. 6 plots the distribution of features for subjects of increasing age. In the figure, each point corresponds to the sum of feature values for a given subject and feature type (e.g., for cortical features, the values are summed across the 62 cortical regions). The relationship between different brain features and age is further depicted in Fig. 7. Each plot of this figure shows the value of individual features (i.e., 62 region measurements for cortical thickness, area and volume, 62 SVD components for DWI connectome) for all 841 subjects sorted by increasing age. Separate dot colors are employed for each of the 62 features of every feature type. We can see that the thickness in all parcels decreases linearly with age. Following the results of Table 5, cortical thickness features show the highest correlation with chronological age, while DWI-based connectome features display almost no correlation.

To further assess the relative importance of individual features for predicting brain age, we show in Table 6 the regression coefficients α_i obtained using the model of Eq. (8) for each of the 10 cross-validation



(a) Cortical surface area (b) Cortical thickness

(c) Cortical volume (d) Connectome of DWI

Fig. 6. Feature values versus age. Each dot corresponds to the sum of features values (e.g., over cortical regions) for a single subject.

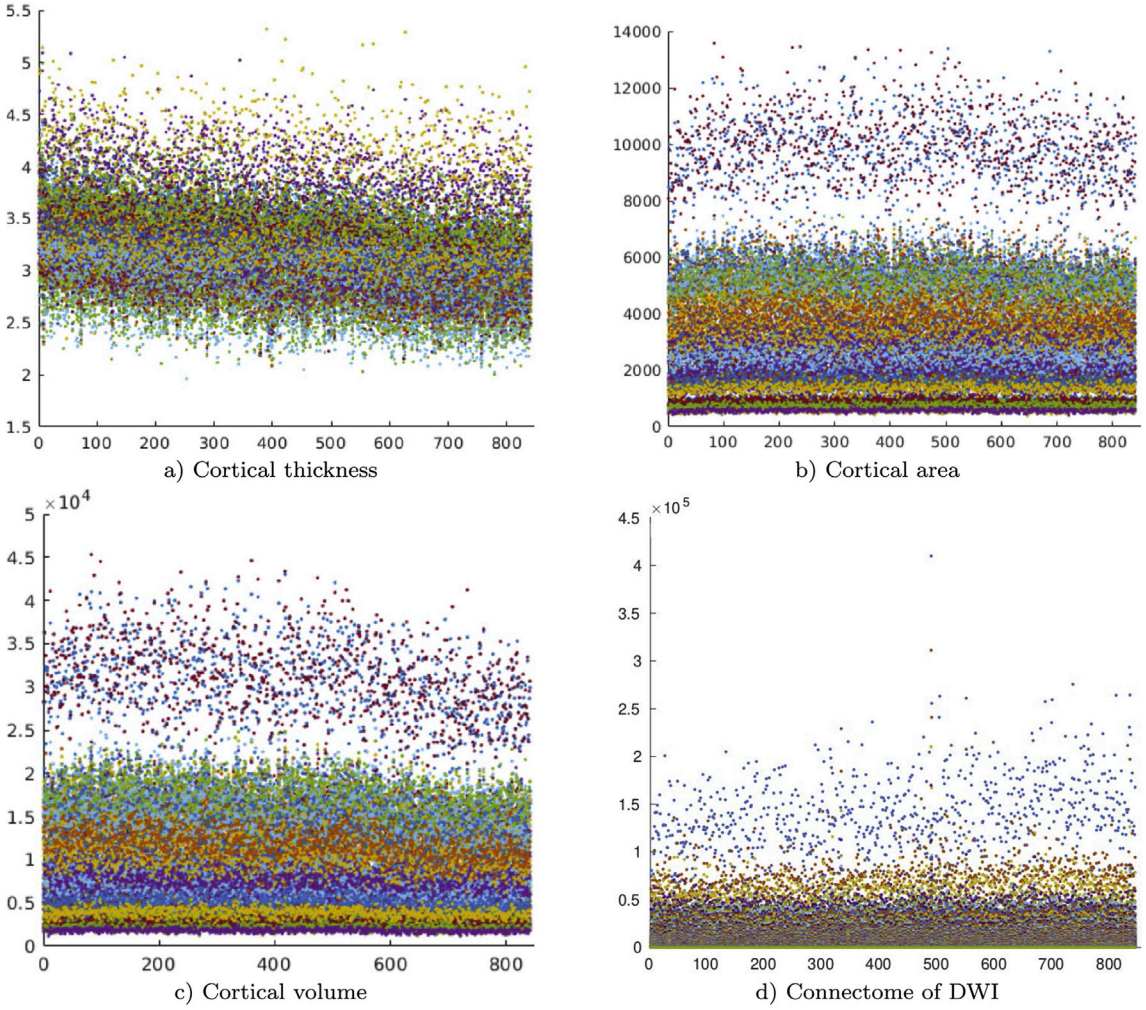


Fig. 7. Distribution of individual features of each feature type for all 841 subjects sorted by increasing age (each dot color corresponds to a different feature).

Table 6
Regression coefficients α_i for each cross-validation fold and mean values over all 10 folds.

	1	2	3	4	5	6	7	8	9	10	Mean
Surf. area	0.074	0.074	0.072	0.117	0.135	0.082	0.099	0.099	0.096	0.095	0.094
Thickness	0.581	0.581	0.597	0.626	0.598	0.564	0.547	0.547	0.600	0.583	0.583
Volume	0.310	0.310	0.282	0.236	0.246	0.310	0.315	0.315	0.285	0.289	0.290
DWI	0.065	0.065	0.067	0.065	0.071	0.086	0.079	0.079	0.067	0.071	0.071

folds. Although estimated coefficients vary across folds, their respective contribution to the final prediction remains similar, with cortical thickness being the most important feature, followed by volume, surface area and finally DWI connectome. This follows the results obtained with individual feature sets (see Table 5), where cortical thickness measurements led to the highest prediction accuracy.

4.3. Predicted age and cognition

We first study the correlation between age and four NTCB cognitive measures, i.e., Flanker Inhibitory Control (Flanker), Attention Test (Attention), Dimensional Change Card Sort (DCCS), and language measures of the Oral Reading Recognition tests (Reading). The correlation coefficient (R) obtained on the cohort of 636 subjects with provided cognitive scores is given in Table 7 (first row). Scatter plots of cognitive scores versus age for the same cohort are shown in Fig. 8. A mild positive correlation can be observed for all cognitive measures, with the highest correlation of R = 0.346 obtained for the Reading score. Next, we evaluate the correlation between the same cognitive measures and the brain age predicted by our model. For this experiment, we randomly split the 636 subjects into a training set of 509 subjects and a testing set of 127 subjects. Results are provided in Table 7 and Fig. 9. A similar correlation with cognitive scores is found for the actual age and the predicted age of

Table 7
Correlation coefficient (R) between cognitive scores and real age, predicted age or prediction residuals (predicted age minus real age). First row: cohort of 636 subjects with cognitive scores. Other rows: 127 subjects in the testing set.

		Reading	Flanker	Attention	DCCS
Real age	**(all)**	0.346	0.281	0.245	0.279
Real age	**(test)**	0.363	0.285	0.270	0.285
Predicted age	**(test)**	0.366	0.301	0.282	0.282
Prediction residual	**(test)**	− 0.190	− 0.135	− 0.132	− 0.155

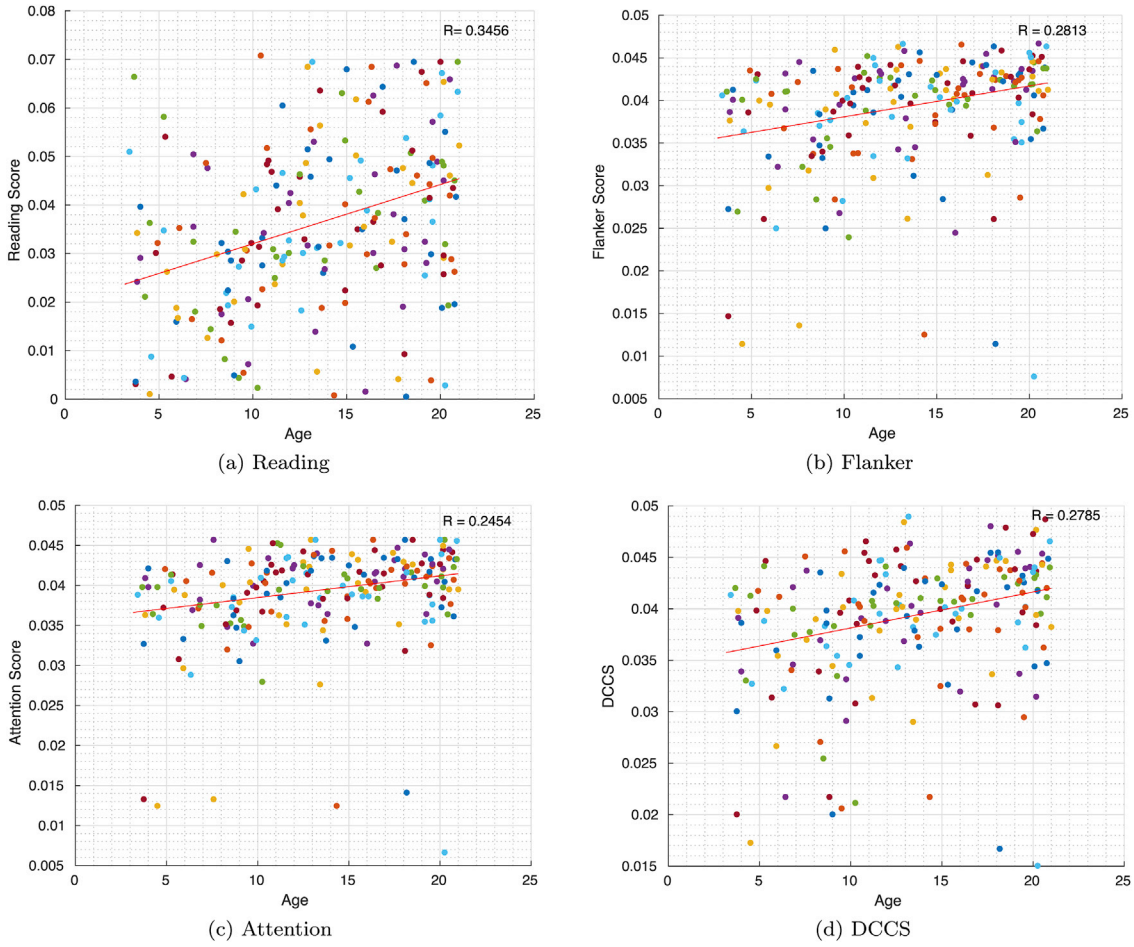


Fig. 8. Cognitive scores with respect to subject age.

test subjects, as presented in the second and third rows of Table 7. From the last row of this table and Fig. 9, we also observe that prediction residuals are negatively correlated with cognitive scores.

In the next experiment, we determine whether differences between chronological age and brain age predicted by our model correlates with cognitive measures from the NTCB. We split the 127 test subjects in two groups: subjects whose predicted age is greater or equal to their real age (*Above* group), and those with a predicted age less than the real age (*Below* group). We then compared the distribution of cognitive scores in the two groups. The top row of Fig. 10 shows the histogram of scores for the Above (blue line) and Below (red line) groups, with group averages given between parentheses. We observed that subjects with a predicted age greater or equal to their real age have lower cognitive scores, on average, than those in the other group. We performed a one-tailed *t*-test with null hypothesis of no difference between the groups and report the p-values in Fig. 10. Results indicate that observed differences may not be due to random sampling, for all cognitive scores ($p < 0.05$). A possible confound in this analysis is that cognition is highly correlated with age in developing subjects. To eliminate this confound, we normalized cognitive scores of subjects by subtracting the mean score of subjects with the same age, and dividing the result by the standard deviation of same-age subjects. Hence, normalized values indicate how individual subjects compare to those of the same age. The bottom row of Fig. 10 gives the distribution of normalized scores for the subjects groups used in the previous analysis (i.e., predicted age Above or Below the real age). Once again, we observe higher mean normalized scores in the Below group compared to the Above group. However, we cannot establish significance in this case.

4.4. Impact of gender

We evaluate the sensitivity of the proposed model to gender differences by using female and male data separately for training. Results in Fig. 11 reveal that the model may be less accurate on the female cohort, with a larger mean and variance of prediction residuals for this cohort (differences are significant with a *t*-test p-value of 0.0352). To gain additional insight, we plot in Fig. 12 the distribution of cortical thickness in relation to age for male and female subjects, since this feature type has the greatest influence on prediction performance. We compare the principal component of distributions computed for male and female (dashed lines in the figure), which estimates the mean cortical thickness as function of age. We find a smaller mean thickness for female compared to male of similar age, which could partly explain why the female group has higher prediction error compared to the male cohort.

5. Discussion

Several studies have investigated the relevance of neuroimaging features such as cortical surface measures for analyzing brain development and brain maturity (Cole and Franke, 2017; Ball et al., 2017; Brown et al., 2012; Becker et al., 2018; Varikuti et al., 2018; Franke et al., 2018; Pardoe and Kuzniecky, 2018; Cole, 2017; Yamaguchi and Honma, 2012). Multi-modal and multivariate neuroimaging methods have been revealed as a powerful tool for deriving spatial changes during aging (Cole et al., 2017b; Fjell et al., 2013; Sowell et al., 2003). As reported in (Rosenberg et al., 2018), large-scale datasets and computational tools based on machine learning have brought forward new ways to study developmental changes in brain structure and behavior, that could offer translational

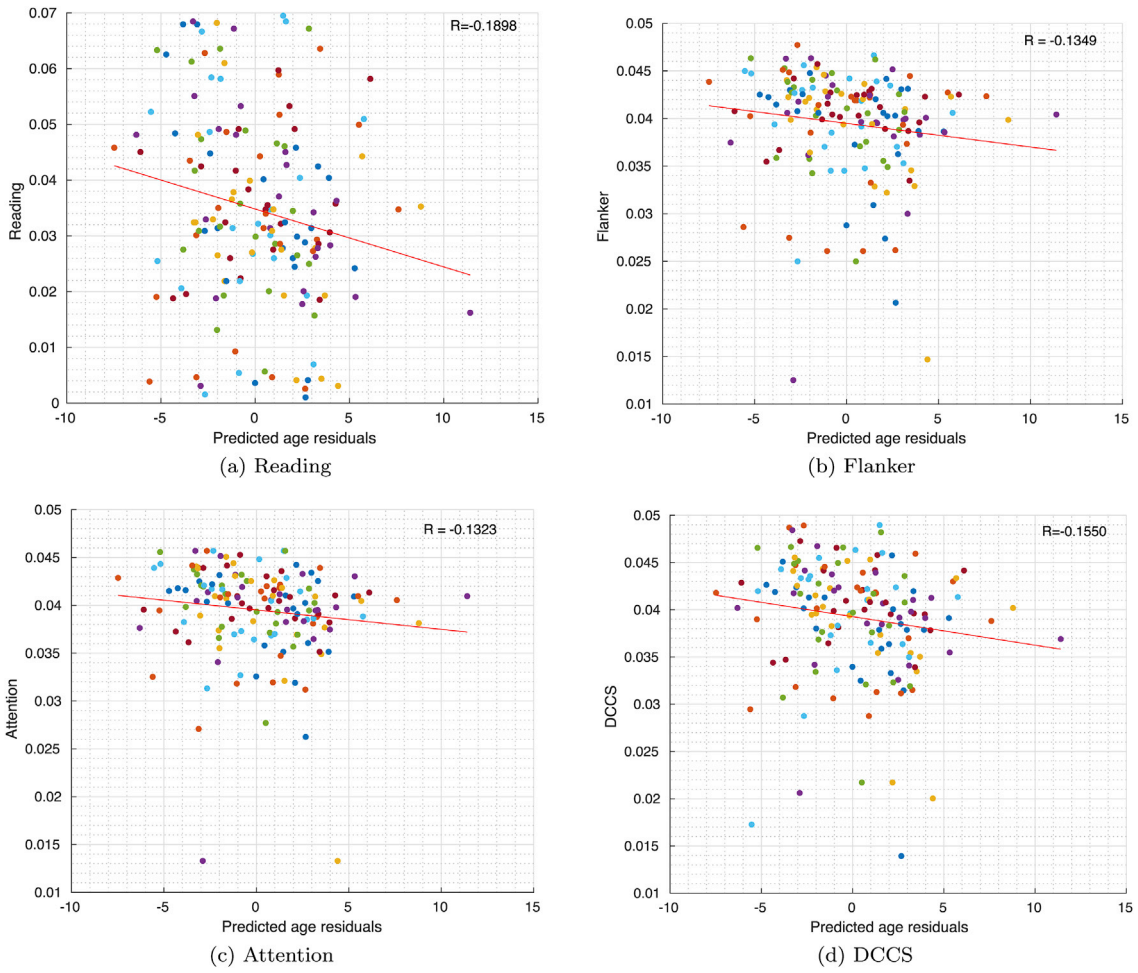


Fig. 9. Cognitive scores with respect to prediction residuals (predicted age minus real age).

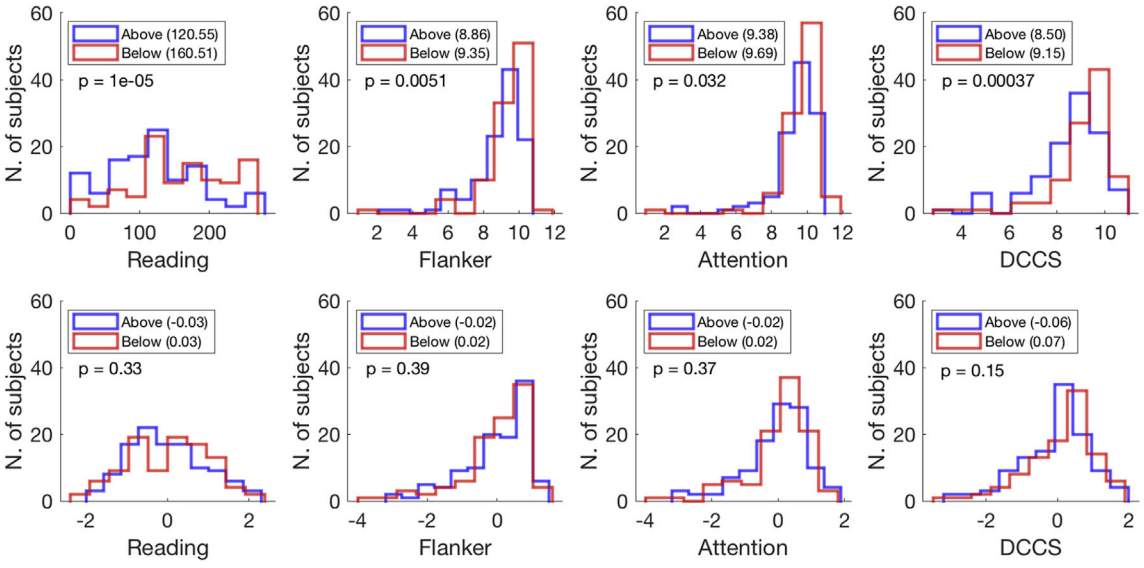
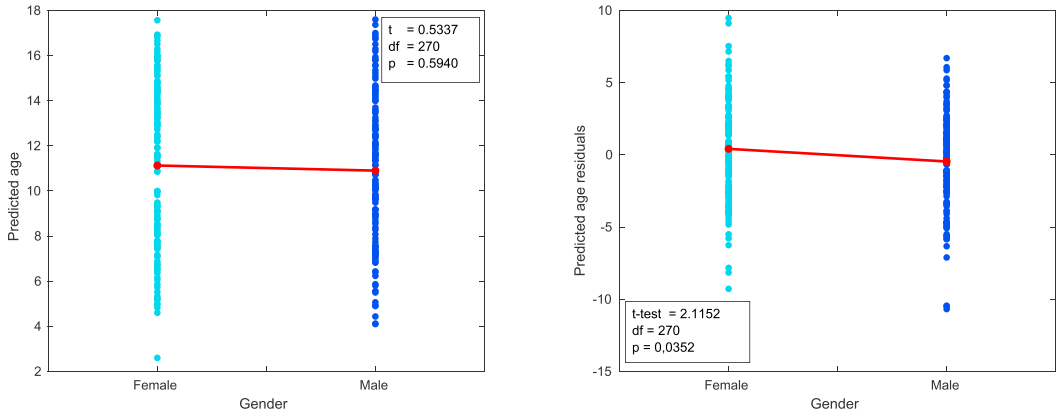


Fig. 10. Distribution of cognitive scores for subjects whose predicted age is greater or equal to (Above) or less than (Below) their real age. Top: unnormalized scores. Bottom: cognitive scores normalized within each age group. Values between parentheses are the distribution mean.

benefits for personalized medicine and education. In (Cole et al., 2017b), Cole et al. found that combining multivariate measures of biological ageing helped in building accurate biomarkers for predicting cognitive deterioration and mortality. Ball et al. (2017) suggested that variability in brain development could be modeled using a low dimensional data representation. Our work follows these studies by showing the potential of machine learning to derive neuroimaging biomarkers for predicting brain maturity (Tables 2 and 3).



(a) Predicted age (b) Predicted age residuals

Fig. 11. Predicted age and residuals versus gender.

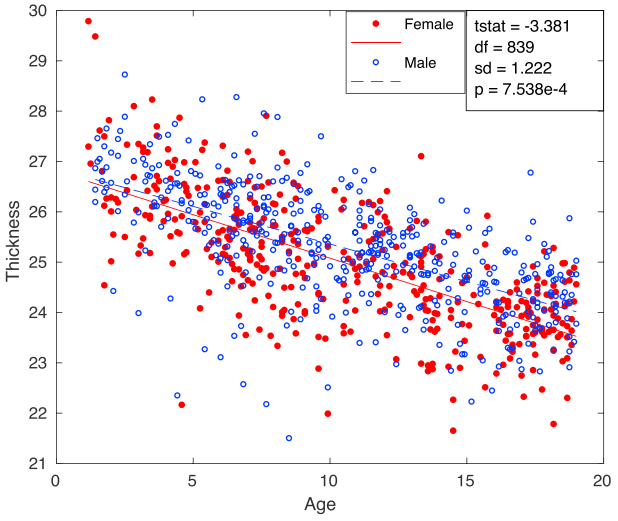


Fig. 12. Cortical thickness versus age for female and male subjects. Dashed lines correspond to the principal component of the distributions.

We propose an extension of standard dictionary learning to predict brain maturation progression in the PING dataset. By adding orthogonality and non-negativity constraints to the paired dictionary approach, we introduce potential modeling benefits, including better control of overfitting, implicit feature selection and improved interpretability. Experiments on predicting the age of 841 PING subjects between 3 and 21 years demonstrated the advantage of our NPDL model, compared to Dictionary Pair Learning (DPL) (Gu et al. Feng), as well as the benefit of imposing non-negativity constraints while learning the class-specific dictionaries. Compared to recent studies based on the same data, however, the proposed method leads to a higher prediction error. For instance, an MAE of 1.79 years is reported in (Ball et al., 2017), using cortical measures corrected for global differences in intracranial volume. Whereas this previous study used cortical measures at the voxel level, for up to 300,000 measurements per surface, we here only considered aggregated measures in atlas parcels. Hence, employing finer-grained information could possibly improve the accuracy of our method.

The relation between gray matter density and brain age is well-established in the literature (Khundrakpam et al., 2015; Giedd and Rapoport, 2010; Sowell et al., 2002). Moreover, recent studies have proposed predictive models combining several cortical measures (Ball et al., 2017; Erus et al., 2014; Brown et al., 2012). However, the link between structural connectivity and brain maturity had not been fully explored. Our results showed cortical thickness to be the most discriminative feature among cortical measures, followed by cortical volume (Table 5). This differs from the study in (Ball et al., 2017), also using the PING cohort, where cortical thickness was found to be the least predictive measure. This may be due to different measures of the cortical thickness and should be further investigated. Although measures of DWI-based connectome show a slight increase with age (Fig. 6), results of our study indicate this information to have a low predictive power for brain maturity.

We analyzed the relationship between brain age predictions and cognition. Specifically, we considered the signed difference between the predicted brain age and chronological age of subjects as a biomarker to determine four NTCB cognitive measures, i.e., Flanker Inhibitory Control (Flanker), Attention Test (Attention), Dimensional Change Card Sort (DCCS), and language measures of the Oral Reading Recognition tests (Reading). Results showed that subjects with a predicted age above their real age had higher mean cognitive scores than those in the opposite group. Although this difference is significant when using unnormalized scores, significance disappears when normalizing the scores of a subject with respect to same age subjects in the cohort. This could confirm the result in (Ball et al., 2017), where no significant association was found between model accuracy and cognitive performance. As pointed out in previous studies such as (Le et al., 2018) and observed in our experiments (Fig. 3), using prediction residuals (i.e., predicted age minus actual age) as biomarker for cognitive scores can induce systematic bias due to a regression toward the mean effect. Additional steps could be used to avoid this bias, for instance including chronological age as regression covariate (see (Le et al., 2018)).

Our experiments on the impact of gender revealed a reduced accuracy of our model for predicting the brain age of female subjects. When comparing the distribution of cortical thickness values across gender, we found a slightly higher mean for male subjects of the same age (Fig. 12). This observation is in line with previous studies indicating gender effects in brain anatomical features, see for instance (Luders et al., 2006, 2012).

6. Conclusion

We proposed a novel non-negative projective dictionary learning (NDPL) approach for predicting brain maturity. This approach extends conventional methods for learning discriminative dictionaries by imposing both orthogonality and non-negativity constraints, which reduces redundancy in the learned representation and acts as an implicit feature selection technique. Experiments on the task of modeling brain age showed the advantages of our approach compared to Dictionary Pair Learning (DPL) and the model without non-negativity constraints. Our

analysis indicates that cortical thickness is the most discriminative feature for predicting brain age. Moreover, results highlight differences in the prediction accuracy obtained for male and female subjects and suggest a possible link between prediction residuals and cognition.

Acknowledgements

This work was supported, in part, by the FRQS Quebec (CCC-Axis 246110, 271636), Brain Canada/HBHL/Exp (247858), National Nature Science Foundation of China (NSFC: 61902220, 61602277, U1609218), Canadian Open Neuroscience Platform(CONP RSA), NIH-NIBIB P41 EB019936 (ReproNim), NIH-NIMH R01 MH083320 and NIH-NIMH R01 MH083320 (CANDIShare), as well as the Healthy Brains for Healthy Lives (HBHL) initiative.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116226.

Appendix

1. The flowchart of data pre-processing

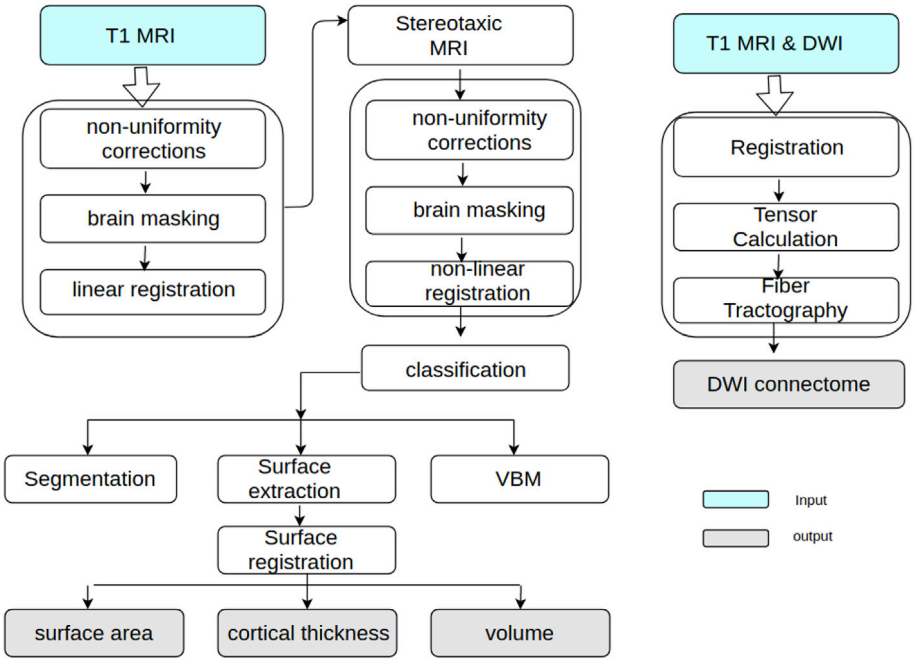


Fig. 13. Flowchart of data pre-processing.

References

Adeli, E., Meng, Y., Li, G., Lin, W., Shen, D., 2018. Multi-task Prediction of Infant Cognitive Scores from Longitudinal Incomplete Neuroimaging Data. NeuroImage.

Akshoomoff, N., Newman, E., Thompson, W.K., McCabe, C., Bloss, C.S., Chang, L., Amaral, D.G., Casey, B., Ernst, T.M., Frazier, J.A., et al., 2014. The NIH Toolbox Cognition Battery: results from a large normative developmental sample (PING). Neuropsychology 28, 1.

Ball, G., Adamson, C., Beare, R., Seal, M.L., 2017. Modelling neuroanatomical variation during childhood and adolescence with neighbourhood-preserving embedding. Sci. Rep. 7, 17796.

Barnes, K.A., Cohen, A.L., Power, J.D., Nelson, S.M., Dosenbach, Y.B., Miezin, F.M., Petersen, S.E., Schlaggar, B.L., 2010. Identifying basal ganglia divisions in individuals using resting-state functional connectivity MRI. Front. Syst. Neurosci. 4, 18.

Becker, B.G., Klein, T., Wachinger, C., Initiative, A.D.N., et al., 2018. Gaussian process uncertainty in age estimation as a measure of brain abnormality. Neuroimage 175, 246–258.

Brown, T.T., Kuperman, J.M., Chung, Y., Erhart, M., McCabe, C., Hagler Jr., D.J., Venkatraman, V.K., Akshoomoff, N., Amaral, D.G., Bloss, C.S., et al., 2012. Neuroanatomical assessment of biological maturity. Curr. Biol. 22, 1693–1698.

Cole, J.H., 2017. Neuroimaging-derived brain-age: an ageing biomarker? Aging (Albany NY) 9, 1861.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: Innovative brain ageing biomarkers. Trends Neurosci. 40 (12), 681–690.

Cole, J.H., Poudel, R.P., Tsagkrasoulis, D., Caan, M.W., Steves, C., Spector, T.D., Montana, G., 2017. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. Neuroimage 163, 115–124.

Cole, J.H., Ritchie, S.J., Bastin, M.E., Hernández, M.V., Maniega, S.M., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., et al., 2017. Brain age predicts mortality. Mol. Psychiatry.

Collins, D., Neelin, P., Peters, T., Evans, A., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage 31, 968–980.

Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., et al., 2010. Prediction of individual brain maturity using fmri. Science 329, 1358–1361.

Erus, G., Battapady, H., Satterthwaite, T.D., Hakonarson, H., Gur, R.E., Davatzikos, C., Gur, R.C., 2014. Imaging patterns of brain development and their relationship to cognition. Cerebr. Cortex 25, 1676–1684.

Eskildsen, S.F., Coupé, P., Fonov, V.S., Pruessner, J.C., Collins, D.L., Initiative, A.D.N., et al., 2015. Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. Neurobiol. Aging 36, S23–S31.

Evans, A.C., Group, B.D.C., et al., 2006. The NIH MRI study of normal brain development. Neuroimage 30, 184–202.

Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U., Church, J.A., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2009. Functional brain networks develop from a "local to distributed" organization. PLoS Comput. Biol. 5, e1000381.

Fjell, A.M., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A.M., Walhovd, K.B., et al., 2013. Critical ages in the life course of the adult brain: nonlinear subcortical aging. Neurobiol. Aging 34, 2239–2247.

Franke, K., Luders, E., May, A., Wilke, M., Gaser, C., 2012. Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. Neuroimage 63, 1305–1312.

Franke, K., Gaser, C., Roseboom, T.J., Schwab, M., de Rooij, S.R., 2018. Premature brain aging in humans exposed to maternal nutrient restriction during early gestation. Neuroimage 173, 460–471.

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I., 2014. DIPY, a library for the analysis of diffusion MRI data. Front. Neuroinf. 8, 8.

Giedd, J.N., Rapoport, J.L., 2010. Structural MRI of pediatric brain development: what have we learned and where are we going? Neuron 67, 728–734.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nat. Neurosci. 2, 861.

Gorski, J., Pfeuffer, F., Klamroth, K., 2007. Biconvex sets and optimization with biconvex functions: a survey and extensions. Math. Methods Oper. Res. 66, 373–407.

S. Gu, L. Zhang, W. Zuo, X. Feng, Projective dictionary pair learning for pattern classification, in: Advances in Neural Information Processing Systems,2014 pp. 793–801..

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

Jernigan, T.L., Brown, T.T., Hagler Jr., D.J., Akshoomoff, N., Bartsch, H., Newman, E., Thompson, W.K., Bloss, C.S., Murray, S.S., Schork, N., et al., 2016. The pediatric imaging, neurocognition, and genetics (PING) data repository. Neuroimage 124, 1149–1154.

Khundrakpam, B.S., Tohka, J., Evans, A.C., Group, B.D.C., et al., 2015. Prediction of brain maturity based on cortical thickness at different spatial resolutions. Neuroimage 111, 350–359.

Kiar, G., Bridgeford, E., Chandrashekhar, V., Mhembere, D., Burns, R., Gray Roncal, W., Vogelstein, J., 2017. A Comprehensive Cloud Framework for Accurate and Reliable Human Connectome Estimation and Meganalysis bioRxiv.

Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I., Evans, A.C., 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. Neuroimage 27, 210–221.

Le, T.T., Kuplicki, R.T., McKinney, B.A., Yeh, H.-w., Thompson, W.K., Paulus, M.P., 2018. A nonlinear simulation framework supports adjusting for age when analyzing brainage. Front. Aging Neurosci. 10, 317.

Li, J., Wu, Y., Zhao, J., Lu, K., 2017. Low-rank discriminant embedding for multiview learning. IEEE transactions on cybernetics 47, 3516–3529.

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S.K., Huntenburg, J.M., Lampe, L., Rahim, M., Abraham, A., Craddock, R.C., et al., 2017. Predicting brain-age from multimodal imaging data captures cognitive impairment. Neuroimage 148, 179–188.

Luders, E., Narr, K.L., Thompson, P.M., Rex, D.E., Woods, R.P., DeLuca, H., Jancke, L., Toga, A.W., 2006. Gender effects on cortical thickness and the influence of scaling. Hum. Brain Mapp. 27, 314–324.

Luders, E., Sánchez, F.J., Tosun, D., Shattuck, D.W., Gaser, C., Vilain, E., Toga, A.W., 2012. Increased cortical thickness in male-to-female transsexualism. J. Behav. Brain Sci. 2, 357.

Pardoe, H.R., Kuzniecky, R., 2018. NAPR: a cloud-based framework for neuroanatomical age prediction. Neuroinformatics 16, 43–49.

Pfefferbaum, A., Sullivan, E.V., Hedehus, M., Lim, K.O., Adalsteinsson, E., Moseley, M., 2000. Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. Magn. Reson. Med.: An Official Journal of the International Society for Magnetic Resonance in Medicine 44, 259–268.

Rosenberg, M.D., Casey, B., Holmes, A.J., 2018. Prediction complements explanation in understanding the developing brain. Nat. Commun. 9, 589.

Salat, D., Tuch, D., Hevelone, N., Fischl, B., Corkin, S., Rosas, H., Dale, A., 2005. Age-related changes in prefrontal white matter measured by diffusion tensor imaging. Ann. N. Y. Acad. Sci. 1064, 37–49.

Shaw, P., Lerch, J., Greenstein, D., Sharp, W., Clasen, L., Evans, A., Giedd, J., Castellanos, F.X., Rapoport, J., 2006. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. Arch. Gen. Psychiatr. 63, 540–549.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.

Sowell, E.R., Trauner, D.A., Gamst, A., Jernigan, T.L., 2002. Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. Dev. Med. Child Neurol. 44, 4–16.

Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A.W., 2003. Mapping cortical change across the human life span. Nat. Neurosci. 6, 309.

Tohka, J., Zijdenbos, A., Evans, A., 2004. Fast and robust parameter estimation for statistical partial volume models in brain mri. Neuroimage 23, 84–97.

Varikuti, D.P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K.R., Jockwitz, C., Caspers, S., Moebus, S., Amunts, K., et al., 2018. Evaluation of non-negative matrix factorization of grey matter in age prediction. Neuroimage 173, 394–410.

Walhovd, K.B., Fjell, A.M., Brown, T.T., Kuperman, J.M., Chung, Y., Hagler, D.J., Roddey, J.C., Erhart, M., McCabe, C., Akshoomoff, N., et al., 2012. Long-term influence of normal variation in neonatal characteristics on human brain development. Proc. Natl. Acad. Sci. 109, 20089–20094.

Yamaguchi, K., Honma, K., 2012. Development of the human abducens nucleus: a morphometric study. Brain Dev. 34, 712–718.

Zhang, M., Desrosiers, C., 2018. Structure Preserving Image Denoising Based on Low-Rank Reconstruction and Gradient Histograms. Computer Vision and Image Understanding.

Zhang, M., Desrosiers, C., 2018. High-quality image restoration using low-rank patch regularization and global structure sparsity. IEEE Trans. Image Process. 28, 868–879.

Zhang, M., Desrosiers, C., Zhang, C., 2018a. Atlas-based reconstruction of high performance brain MR data. Pattern Recognit. 76, 549–559.

M. Zhang, C. Desrosiers, Y. Guo, C. Zhang, B. Khundrakpam, A. Evans, Brain status prediction with non-negative projective dictionary learning, in: International Workshop on Machine Learning in Medical Imaging, Springer,2018b pp. 152–160..

W. Zhou, Y. Yang, Z. Yu, Discriminative dictionary learning for EEG signal classification in brain-computer interface, in: Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on, IEEE, pp. 1582–1585..

Zhu, X., Suk, H.I., Lee, S.W., Shen, D., 2016. Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 63, 607–618.

Zijdenbos, A.P., Forghani, R., Evans, A.C., 2002. Automatic pipeline analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. IEEE Trans. Med. Imaging 21, 1280–1291.