Generalization of diffusion magnetic resonance imaging–based brain age prediction model through transfer learning

Chang-Le Chen, Yung-Chin Hsu, Li-Ying Yang, Yu-Hung Tung, Wen-Bin Luo, Chih-Min Liu, Tzung-Jeng Hwang, Hai-Gwo Hwu, Wen-Yih Isaac Tseng

Institute of Medical Device and Imaging, College of Medicine, National Taiwan University, Taipei, Taiwan
AcroViz Technology Inc., Taipei, Taiwan
Department of Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan
Graduate Institute of Brain and Mind Sciences, College of Medicine, National Taiwan University, Taipei, Taiwan
Molecular Imaging Center, National Taiwan University, Taipei, Taiwan

ARTICLE INFO
Keywords:
Brain age
Diffusion harmonization
Diffusion MRI
Neural network
Transfer learning
White matter

ABSTRACT
Brain age prediction models using diffusion magnetic resonance imaging (dMRI) and machine learning techniques enable individual assessment of brain aging status in healthy people and patients with brain disorders. However, dMRI data are notorious for high intersite variability, prohibiting direct application of a model to the datasets obtained from other sites. In this study, we generalized the dMRI-based brain age model to different dMRI datasets acquired under different imaging conditions. Specifically, we adopted a transfer learning approach to achieve domain adaptation. To evaluate the performance of transferred models, brain age prediction models were constructed using a large dMRI dataset as the source domain, and the models were transferred to three target domains with distinct acquisition scenarios. The experiments were performed to investigate (1) the tuning data size needed to achieve satisfactory performance for brain age prediction, (2) the feature types suitable for different dMRI acquisition scenarios, and (3) performance of the transfer learning approach compared with the statistical covariate approach. By tuning the models with relatively small data size and certain feature types, optimal transferred models were obtained with significantly improved prediction performance in all three target cohorts (p < 0.001). The mean absolute error of the predicted age was reduced from 13.89 to 4.78 years in Cohort 1, 8.34 to 5.35 years in Cohort 2, and 8.74 to 5.64 years in Cohort 3. The test–retest reliability of the transferred model was verified using dMRI data acquired at two timepoints (intraclass correlation coefficient = 0.950). Clinical sensitivity of the brain age prediction model was investigated by estimating the brain age in patients with schizophrenia. The prediction made by the transferred model was not significantly different from that made by the reference model. Both models predicted significant brain aging in patients with schizophrenia as compared with healthy controls (p < 0.001); the predicted age difference of the transferred model was 4.63 and 0.26 years for patients and controls, respectively, and that of the reference model was 4.39 and 0.09 years, respectively. In conclusion, transfer learning approach is an efficient way to generalize the dMRI-based brain age prediction model. Appropriate transfer learning approach and suitable tuning data size should be chosen according to different dMRI acquisition scenarios.

1. Introduction
Aging of the human brain is a complex process and is associated with major risks of cognitive decline and neurodegeneration (Lopez-Otin et al., 2013). Patients with neurodegenerative or mental disorders, such as Alzheimer’s disease and schizophrenia (SZ), often exhibit considerable changes in their brain structure and function, which may cause patients’ biological age to deviate from the normative aging trajectory (Cole, 2018). To transform the complicated aging trajectories of the human brain into a comprehensible index, an emerging neuroimaging-based
biomarker called “brain age,” which represents the aging status of the brain, has been developed. Brain age is derived from modern machine learning techniques, which depict the patterns of neuroimaging data from brain scans to predict the brain age (Cole and Franke, 2017). Accumulating evidence has demonstrated that the brain age has both scientific and clinical relevance; it can be used to characterize mortality risk and dementia risk, and can be used as a proxy of gerontological factors in older adults (Cole et al., 2015, 2017a; Liem et al., 2017; Wang et al., 2019).

T1-weighted image-based brain age prediction models can achieve satisfactory performance in healthy people and be sensitive to aberrant aging in patients with neurological diseases (Cole et al., 2015, 2017a; Franke et al., 2010; Gaser et al., 2013; Pardoe et al., 2017; Valizadeh et al., 2017; Kaufmann et al., 2019). However, the morphological features derived from T1-weighted images are relatively macroscopic and insensitive to the microstructural changes during neurodevelopment and neurodegeneration (Deijboley et al., 2005; Weston et al., 2015). Diffusion MRI techniques are considerably sensitive to tissue microstructure because they can detect diffusing water molecules and probe the geometry of surroundings at the cellular scale (Kincses and Vecsei, 2018). By exploring the diffusion process in the white matter, these techniques can highlight subtle brain changes, which are otherwise invisible when other imaging modalities are used. Because of its unique capabilities, dMRI is commonly used to characterize white matter changes during neurodevelopment and neurodegeneration (Norhoj Jespersen, 2018). Many studies have shown that during brain aging, white matter microstructure integrity degrades, leading to the disruption of communication within cerebral neural networks and consequently affecting relevant cognitive functions (Bennett and Madden, 2014; Catani and Ffytche, 2005; Filley, 2005).

Although dMRI is well suited for detecting age-associated changes in the brain, few studies have used dMRI techniques to model the brain age (Mwangi et al., 2013; Richard et al., 2018; Saha et al., 2018). Compared with data from regular imaging modalities, such as T1- and T2-weighted imaging, dMRI data demonstrate more technical challenges and higher intersite variabilities (Malyarenko et al., 2016; Mirzaalian et al., 2015). Intersite or interscanner variability arises from various factors including differences in head coil types, imaging gradient nonlinearity, magnetic field homogeneity, data reconstruction algorithms, and other scanner-related factors (Jovicich et al., 2014; Teipel et al., 2011; Zhu et al., 2011). These factors can cause nonlinear deviations in raw image data and consequently in estimated diffusion measures such as mean diffusivity (MD) and fractional anisotropy (FA) (Mirzaalian et al., 2015). The deviations can deteriorate the performance of a prediction model, which is trained using the data from one site and is intended to be applied to data from another site.

Some methods have been proposed to mitigate the influence of intersite variability such as meta-analysis, statistical covariates, and signal harmonization (Mirzaalian et al., 2018). Meta-analysis involves combining z-scores of a given scalar diffusion measure from all sites to determine group differences (Kochunov et al., 2014). However, this approach requires the subject population for each site to be sufficiently large to capture the variance of the population. The statistical covariate approach accounts for intersite differences through statistical variance analysis. It employs statistical models to regress out site-specific differences by using statistical covariates. This approach requires the source domain data to be accessible and the data size from different sites to be balanced. Signal harmonization involves harmonizing dMRI data by transforming raw signals, rather than derived diffusion indices, by using rotation invariant spherical harmonic features (Mirzaalian et al., 2016, 2018). This approach works under the premise that a group of traveling individuals must be scanned at different sites successively in a short period of time by using similar imaging acquisition parameters. It makes this approach costly and thus infeasible for common use.

In this study, we proposed to use the transfer learning approach to fine-tune the brain age model to fit the datasets acquired from different sites and under different imaging conditions. Transfer learning is frequently used in medical image analysis to overcome large data requirements typically needed for machine learning when linking trained models between different imaging modalities and domains (Banerjee et al., 2018; Christopher et al., 2018; Ghafoorian et al., 2017; Shan et al., 2018; Xiao et al., 2018). Traditional machine learning assumes that training and test data have the same data distribution. When the distributions of the training data and test data are different, the performance of a predictive learner declines (Weiss et al., 2016). Transfer learning allows transforming the model parameters fitted in data distribution from one domain into those from other domains based on the premise that those two domains are linked by a high-level common domain. In the case of the intersite variability of dMRI data, the processed diffusion measures, such as FA, obtained from various sites represent similar microstructural properties but may exhibit different observational distributions. Therefore, the transfer learning approach should be a plausible solution to improving the generalizability of the dMRI-based brain age prediction model.

To prove the feasibility of the transferred learning approach, this paper had two specific aims: validation and verification of the transferred models. The validation aim entailed a search of the minimal data size to tune the model parameters and hyperparameters for the satisfactory transfer of the models from the source domain to the target domain. We hypothesized that the errors in brain age prediction were significantly reduced after transfer learning. The verification aim involved the application of the transferred models to patients with schizophrenia. We hypothesized that the phenomenon of significant brain aging in schizophrenia could still be detected by the transferred models (Kochunov et al., 2013; Koutsouleris et al., 2014; Schnack et al., 2016).

2. Material and methods

Fig. 1 presents an overview of our transfer learning experiment. The source domain for brain age model training used the datasets obtained from the Cambridge Centre for Ageing and Neuroscience (CamCAN) cohort (Shafo et al., 2014; Taylor et al., 2017). The target domains for transferring the brain age model comprised three independent cohorts with different dMRI parameters: (1) the National Taiwan University Hospital (NTUH) cohort, (2) the Hammersmith Hospital (HH) subset in the Information eXtraction from Images (IXI) cohort (https://brain-deve
dopment.org/ixi-dataset/), and (3) the Guy’s Hospital (Guys) subset in the IXI cohort. The details of the datasets are described in Section 2.1. The transfer learning experiment included three parts: data preprocessing, model validation, and model verification. For data preprocessing, we used a reconstruction framework called mean apparent propagator (MAP) MRI to reconstruct diffusion-weighted images into several diffusion indices. These diffusion indices were separated into the tensor feature set and advanced feature set. These two sets represent the variability of diffusion acquisition schemes, which may capture distinct aspects of biological information in white matter aging (Coutu et al., 2014).

The two feature sets were used in the transfer learning experiment to assess the difference in performance in different diffusion acquisition schemes. For model validation, the brain age models were trained from the CamCAN cohort (source domain) by using the tensor and advanced features separately. After model training was completed, the experiment was conducted to transfer the pre-trained models to the target domains. The global searching of transfer learning entailed two parts: (1) determining how many tuning samples were needed from the target domain to fine-tune the brain age model and (2) comparing the transfer learning method with the statistical covariate approach and hybrid approach. The transfer learning process here was optimized using an agile optimization process because it facilitated rapid prototyping and broad searching. After the optimal tuning sample size and transfer strategy were determined, the refined optimization was employed to fine-tune the brain age model according to the best training configuration. For model verification, the transferred model in the NTUH cohort was applied to the
samples with two timepoint measurements to evaluate the repeatability of brain age prediction. Clinical sensitivity was evaluated by applying the transferred model to patients with SZ to detect advanced brain aging in this imaging cohort.

2.1. Datasets

For brain age modeling and the transfer learning experiment, the dMRI datasets of neurologically healthy and cognitively normal volunteers from four independent cohorts, namely the CamCAN, NTUH, IXI-HH, and IXI-Guys cohorts, were used. Table 1 summarizes the demographics of the four cohorts. The CamCAN cohort is a public access database funded by the Biotechnology and Biological Sciences Research Council, the UK Medical Research Council, and the University of Cambridge. The NTUH cohort was obtained from a private database from the corresponding author. All individuals in this cohort were anonymized and had no history of significant neurological or psychiatric problems. The HH and Guys cohorts were the two subsets of IXI datasets, a public database jointly developed by the Imperial College of Science Technology and Medicine and University College, London. In the experiment for repeatability assessment, 46 independent healthy adults aged 22–71 years were recruited from NTUH (mean age = 50.2 years, SD = 14.1 years, 30.4% were men) in whom two brain scan sessions were conducted 72 days apart on average. These individuals were enrolled using the same recruitment criteria as those for the NTUH cohort. For clinical sensitivity evaluation, we included 158 patients diagnosed with SZ by NTUH psychiatrists according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (age range = 16–62 years, mean age = 31.0 years, SD = 8.5 years, 45.6% were men), and 160 matched controls (age range = 18–62 years, mean = 31.7 years, SD = 9.9 years, 45.0% were men). None of the patients had a history of brain surgery or other neurological or psychiatric diseases. All participants recruited in NTUH provided written informed consent, as approved by the Institutional Review Board of the NTUH. All imaging data in NTUH were acquired using the same MRI scanner and imaging settings.

2.2. MRI acquisition

Table 1 summarizes the imaging parameters of T1-weighted images and diffusion-weighted images for each cohort. The participants in the CamCAN cohort were scanned using a 3T Siemens TIM Trio scanner with a 32-channel phased array head coil at Medical Research Council (UK) Cognition and Brain Sciences Unit in Cambridge, UK (Shafo et al., 2014). Two-shell DTI datasets were acquired using a twice-refocused diffusion pulsed-gradient spin-echo (SE) echo-planar imaging (EPI) sequence with the following imaging parameters: repetition time (TR) = 9100 ms, echo time (TE) = 104 ms; field of view (FOV) = 192 × 192 mm²; voxel size = 2 mm isotropic; 66 axial slices using 30 directions with b = 1000 s/mm², 30 directions with b = 2000 s/mm², and 3 images with b = 0 s/mm². High-resolution T1-weighted images were acquired using a three-dimensional (3D) magnetization-prepared rapid gradient echo (3D-MPRAGE) sequence: TR/TE = 2250/2.99 ms, inversion time (TI) = 900 ms, flip angle = 9°, FOV = 256 × 240 × 192 mm³, and voxel size = 1 mm isotropic. The participants in the NTUH cohort including normal individuals for the transfer learning experiment and repeatability assessment, clinical controls, and patients with SZ were scanned using a 3T Siemens TIM Trio scanner with a 32-channel phased array head coil, all by using the same imaging protocols. DSI datasets were acquired using the diffusion pulsed-gradient SE EPI sequence with a twice-refocused balanced echo (Reese et al., 2003; Wedeen et al., 2005): TR/TE = 9600/130 ms, slice thickness = 2.5 mm, acquisition matrix = 80 × 80, FOV = 200 × 200 mm², and in-plane spatial resolution = 2.5 × 2.5 mm². The diffusion-encoding acquisition scheme used in this dataset followed the DSI framework.
Table 1
Demographics and imaging parameters of four independent cohorts for brain age modeling.

|               | CamCAN | NTUH   | IXI-HH | IXI-Guys |
|---------------|--------|--------|--------|----------|
| Total N (% males) | 616 (49.5%) | 405 (42.2%) | 176 (48.9%) | 183 (43.7%) |
| Mean age (SD)   | 54.3 (18.4) | 44.4 (18.7) | 47.2 (16.8) | 56.0 (13.6) |
| Age range (years) | 18–88 | 18–92 | 20–82 | 24–86 |
| Manufacturer field | 3T | 3T | 3T | 1.5T |
| Head coil | Siemens TIM Trio | Siemens TIM Trio | Philips Intera | Philips Gyroscan Intera |
| Voxel size | 2.0 mm isotropic | 1.75 mm isotropic | 2.5 mm | 2.0 mm |
| Slice thickness | 2.0 mm | 2.5 mm | 2.0 mm | 2.35 mm |
| FOV (mm) | 192 × 192 | 224 × 224 | 224 × 224 | 224 × 224 |
| TE (ms) | 9100 | 9600 | 11894 | 9054 |
| TR (ms) | 104 | 130 | 51 | 80 |
| b-value (s/mm²) | 2000/30 directions | 102 directions with one b₀ image | b = 0; 1 image | b = 15; 15 directions |
| Acquisition scheme | Two-shell DTI | Single-shell DTI | Single-shell DTI | Single-shell DTI |

(Wedeen et al., 2005), which applied 102 diffusion-encoding gradients corresponding to the Cartesian grids in the half-sphere of the 3D diffusion-encoding space (q-space) within a radius of 3 units equivalent to \( b_{\text{max}} = 4000 \text{ s/mm}^2 \) (Kuo et al., 2008). Because the data in q-space were real and symmetrical around the origin, the acquired half-space data were projected to fill the other half of the sphere. High-resolution T1-weighted imaging was performed using a 3D-MPRAGE sequence: TR/TE/T1 = 2000/3 ms, flip angle = 9°, FOV = 256 × 192 × 208 mm³, and acquisition matrix = 256 × 192 × 208; this resulted in an isotropic spatial resolution of 1 mm³.

The images of the HH cohort in the IXI dataset were acquired using a 3T Philips Intera MRI scanner. Diffusion-weighted imaging used the twice-refocused diffusion pulsed-gradient SE EPI sequence: TR/TE/T1 = 2000/4/3 ms, slice thickness = 2 mm, voxel size = 1.75 × 1.75 × 2 mm, 15 unique directions with \( b = 1000 \text{ s/mm}^2 \) and 1 image with \( b = 0 \text{ s/mm}^2 \). The spatial correlation between T1- and diffusion-weighted images was evaluated by calculating the spatial correlation between the T1-weighted image-derived white matter tissue probability map and the diffusion-weighted image-derived FA map (alternatively GFA map for DSI dataset). Higher spatial correlation indicated better spatial alignment between T1- and diffusion-weighted images. The datasets with correlations higher than mean spatial correlation minus 2.5 standard deviations were included. The correction algorithm for motion and eddy currents with EDDY in FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy) was used to detect and replace slices affected by signal loss due to bulk motion during diffusion encoding (Andersson and Sotiropoulos, 2016). The eddy current induced nuisance factor of DSI datasets in NTUH was eliminated during the acquisition process by twice-refocused sequence (Reese et al., 2003). Motion-induced signal dropout caused by in-scanner head motion was inevitable in DSI datasets due to a long scan time, particularly in those with high b values. All acquired DSI datasets (5712 images per participant) were examined by comparing the signal in the central square of each image with the predicted signal attenuation. Signal deviation from the predicted distribution was considered signal loss. Data with more than 60 images of signal dropout per participant (1% of the total diffusion-weighted images) were discarded. After all image quality assurance procedures, the data of 96.9% (616/636), 92.7% (405/437), 97.2% (176/181), and 96.8% (183/189) of the patients in the CamCAN, NTUH, HH, and Guys cohorts were retained, respectively.

2.3.2. Diffusion data reconstruction
The diffusion indices derived from the dMRI dataset were computed using the regularization version of the framework of mean apparent propagator (ReMAP)-MRI (Hsu and Tseng, 2018; Ozarslan et al., 2013). The signal in q-space was fitted with a series expansion of Hermite basis functions, which describe diffusion in various microstructural geometries (Averbuch-Shahak, 2016). The zero-order term in the expansion series contained the diffusion tensor that characterized the Gaussian displacement distribution. Higher-order terms in the expansion series were the orthogonal corrections to the Gaussian approximation, and these were used for reconstructing the average propagator. Notably, because of the differences in acquisition schemes, the DTI and DSI datasets were fitted using the maximum of the sixth- and eighth-order terms of the expansion series, respectively. Axial diffusivity (AD), radial diffusivity (RD), MD, FA, and volume ratio (VR) in each voxel were determined by calculating SNR minus 2.5 standard deviations at their site were included. The degree of within-subject alignment between T1- and diffusion-weighted images was evaluated by calculating the spatial correlation between the T1-weighted image-derived white matter tissue probability map and the diffusion-weighted image-derived FA map (alternatively GFA map for DSI dataset). Higher spatial correlation indicated better spatial alignment between T1- and diffusion-weighted images. The datasets with correlations higher than mean spatial correlation minus 2.5 standard deviations were included. The correction algorithm for motion and eddy currents with EDDY in FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/eddy) was used to detect and replace slices affected by signal loss due to bulk motion during diffusion encoding (Andersson and Sotiropoulos, 2016). The eddy current induced nuisance factor of DSI datasets in NTUH was eliminated during the acquisition process by twice-refocused sequence (Reese et al., 2003). Motion-induced signal dropout caused by in-scanner head motion was inevitable in DSI datasets due to a long scan time, particularly in those with high b values. All acquired DSI datasets (5712 images per participant) were examined by comparing the signal in the central square of each image with the predicted signal attenuation. Signal deviation from the predicted distribution was considered signal loss. Data with more than 60 images of signal dropout per participant (1% of the total diffusion-weighted images) were discarded. After all image quality assurance procedures, the data of 96.9% (616/636), 92.7% (405/437), 97.2% (176/181), and 96.8% (183/189) of the patients in the CamCAN, NTUH, HH, and Guys cohorts were retained, respectively.
the first eigenvalue, mean of the second and third eigenvalues, mean of the three eigenvalues of the diffusion tensor, normalized standard deviation of the diffusion tensor, and product of eigenvalues divided by MD cubed, respectively (Alexander et al., 2007; Le Bihan et al., 2001). Generalized FA (GFA) was quantified as the standard deviation of the orientation distribution function (ODF) divided by the root-mean square of the ODF (Tuch, 2004). The non-Gaussianity (NG) series including NG, orthogonal NG components (NGO) and parallel NG components (NGP) were estimated by quantifying the dissimilarity between the propagator and its Gaussian counterpart, representing the generalization of diffusion kurtosis (Ozarslan et al., 2013). Herein, we used the aforementioned nine diffusion indices, namely FA, VR, AD, RD, MD, GFA, NG, NGO, and NGP, to represent various microstructural properties of the white matter, such as myelination degree and fiber caliber, density, and damage (Alexander et al., 2011; Falangola et al., 2013; Kumar et al., 2013). We divided the diffusion indices into tensor (FA, VR, AD, RD and MD) and advanced (GFA, NG, NGO and NGP) feature sets since these two sets represent different aspects of microstructural property in white matter (Goutu et al., 2014). During brain age analysis, MD is the common feature for both tensor and advanced features because it is the most sensitive to age, and it is considered advantageous for brain age prediction (Cox et al., 2016). Therefore, tensor features included FA, VR, AD and RD, whereas GFA, NG, NGO, NGP and MD were advanced features.

2.3.3. Tract-specific feature extraction

To extract effective features of white matter tract integrity for machine learning, tract-based automatic analysis was performed to sample the diffusion indices from 76 predefined major fiber tract bundles over the whole brain (Chen et al., 2015). For the DSI dataset, the 76 major fiber tract bundles were built in the DSI template NTU-DSI-122 (Hsu et al., 2015) through deterministic streamline-based tractography with multiple regions of interest defined in the automated anatomical labeling atlas (Tzourio-Mazoyer et al., 2002). The sampling coordinates of the 76 tracts were transformed from NTU-DSI-122 to individual DSI datasets with the corresponding deformation maps. The deformation maps were obtained through two-step registration, which included anatomical information provided by the T1-weighted images (Ashburner and Friston, 2011) and microstructural information provided by DSI datasets (Hsu et al., 2012). The sampling coordinates were aligned with the proceeding direction of each fiber tract bundle, and diffusion indices were sampled in the native space along the sampling coordinates normalized and divided into 100 steps. Finally, for each participant, we obtained the output of tract-specific analysis results, called 3D connectogram (x axis: 100 steps along sampling coordinates; y axis: 76 white matter tract bundles; z axis: 9 diffusion indices). For the DTI dataset, the analytic procedure was similar to that for the DSI dataset, except for the use of a different diffusion template and registration algorithm. The diffusion template for DTI used NTU-DSI-122-DTI, derived from NTU-DSI-122 by sampling the diffusion-weighted images with b-value 1000 s/mm² along 18 nonlinear directions via linear interpolation. The registration algorithm in the second step used the large deformation diffeomorphic metric mapping (LDDMM) framework of DTI (Cao et al., 2006), instead of the LDDMM of DSI. The 3D connectograms of all participants were used to represent white matter characteristics for brain age modeling.

2.4. Brain age modeling

To transfer the dMRI-based model from the source domain to the target domain, the brain age models were trained using tensor and advanced features separately in the CamCAN cohort. First, the CamCAN cohort was split into the training (N = 500) and test (N = 116) sets by using a conditionally random method. The distributions of age and sex in the two sets were statistically identical. The input data of the model were the connectograms derived from the tract-specific analysis and sex factor. Because of the high dimensionality of the connectogram data, the tract steps of connectogram were averaged into one feature, and the dimension of diffusion index was concatenated to the dimension of the tracts. In brief, the data matrix of the model input was a two-dimensional array consisting of the numbers of individuals and tract features (5 diffusion indices × 76 tracts = 380 tract features). Here, cascade neural network models, which outperformed the other common classical machine learning approaches in our internal test (see Supplementary Information 1), were used to predict age by using the tract features. The cascade neural network is a feedforward neural network involving connections from the input and every previous layer to the subsequent layer. It is similar to a simplified fully connected version of a dense block in densely connected convolutional networks, which alleviate the vanishing-gradient problem and strengthen feature propagation (Huang et al., 2017). The hyperparameters of the model, including numbers of hidden layers and neurons, penalty of regularization, and types of activation function, were determined through random search (Bergstra and Bengio, 2012). The details of hyperparameter settings were described in the Supplementary Information 2. The loss function of model optimization was specified as mean square error function optimized using gradient descent algorithm with an adaptive learning rate and constant momentum. A 10-fold cross-validation procedure was conducted within the training set to estimate the brain age model performance. The validation set performance was used to stop the model parameter updating and hyperparameter tuning (Supplementary Information 2). The training procedure was implemented using MATLAB R2018b (The MathWorks Inc., Natick, MA, USA) with a graphic processing unit NVIDIA GeForce GTX 1080Ti (NVIDIA Inc., Santa Clara, CA, USA) for accelerated computing. The performance of the trained brain age model was tested by predicting individuals’ brain age in the test set. To quantify model performance, metrics including Pearson’s correlation coefficient (rho) and mean absolute error (MAE) between the predicted and chronological ages were calculated.

2.5. Global searching of transfer learning approach and refined optimization

The right side of Fig. 1 provides an overview of the global searching experiment. The global searching was conducted to investigate two questions regarding the transfer learning approach: (1) How many tuning samples are needed to transfer the brain age model from the source domain to the target domain? and (2) How did the transfer learning approach perform in comparison with the statistical covariate approach and their hybrid version? The data cohorts in the target domain, namely the NTUH, HH, and Guys cohorts, were conditionally-randomly divided into the tuning-pool and test sets (NTUH: tuning pool N = 300, test set N = 105; HH: tuning pool N = 120, test set N = 56; Guys: tuning pool N = 120, test set N = 63). The tuning pool set served as the tuning sample database. During the global searching of the transfer learning approach, the subset of the tuning pool was conditional-randomly selected as the tuning set to re-train the pre-trained model. After the tuning process was completed, the transferred model was applied to the test set from the corresponding target domain for model performance evaluation. Moreover, the statistical covariate approach (also called “cotrain” in subsequent tables and figures) was used for combining both the tuning set from target domain and the training set from source domain to train a new brain age model from scratch with the site indicator as a predictor. The hybrid approach (also called “transfer learning with cotrain” [TLCO] in the subsequent tables and figures) was used for re-training the pre-trained brain age model by using a combination of the tuning and training sets with the site indicator. In addition, training a brain age model from scratch solely using tuning samples (called “single” approach) served as the baseline in each cohort of the target domain. Each point estimation of global searching was repeated 300 times. Both the pre-trained brain age models based on tensor features and advanced features were appraised according to the aforementioned global searching procedure. Because a large amount of simulated tuning experiments was needed, we optimized the tuning process by adopting an agile
optimization method that exploited a time-saving optimizer called scaled conjugate gradient (SCG) algorithm for fast optimization and the hyperparameter settings emulated as those of the training process in the target domain (Moller, 1993). Here, we defined two performance criteria to determine the adequate tuning sample size. The first was optimal performance; that is, the MAE of the test data predicted by the transfer learning approach, statistical covariate approach (cotrain), or hybrid approach (TLCO) was comparable to the standard MAE, which was calculated by the model trained using the maximal amount of the tuning pool set with the single approach. The second was the suboptimal performance; that is, the MAE of the test data was 1 year higher than the standard MAE. Setting up the suboptimal performance criterion allowed providing information about the relationship between data cost and performance improvement. All reported performance metrics pertain to the test set in each cohort, unless specified otherwise. After the tuning sample size, suitable transfer methods, and types of input features were determined in each target cohort, we conducted refined optimization, entailing the gradient descent algorithm with adaptive learning rate and momentum (GDX algorithm) for better optimization and fine-tuning the hyperparameter settings including parameter regularization, loss function, and layer-based parameter freezing (making the parameters of layer in the neural network model untrainable) with random search (Supplementary Information 2), with hold-out validation to transfer the pre-trained model. The performance of refined optimization was assessed by repeating 100 times to provide a robust estimate of transferred model accuracy. The code of brain age modeling and transfer learning are available for open-access in the author’s github repository (https://github.com/ChangleChen/BrainAge_TL).

2.6. Statistical analysis

Statistical analysis was performed using MATLAB R2018b. For transferred model evaluation, statistical metrics, including rho and MAE, between the predicted and chronological ages were calculated for the NTUH, HH and Guys cohorts under the circumstances of optimal transfer, suboptimal transfer, and the unadjusted pre-trained model (directly applying the model trained from the source domain to the data in the target domain without any tuning process). The statistical comparison of the performance between the agile and the refined models was conducted using Z test to compare the difference between two estimated distributions of average MAE under two model tuning conditions. To compare the performance between the CamCAN unadjusted model and the refined optimal transferred model, paired t-test was used to compare the absolute difference between predicted and chronological ages derived from the two models.

For test–retest reliability assessment, the transferred model in the NTUH cohort with optimal performance was applied to the NTUH longitudinal dataset to predict each individual’s brain age. The absolute error (AE) of within-subject predicted age (WSPA) derived from the transferred model was compared with those of the nonadjusted model and the reference model (the brain age model trained with all the tuning pool set of the NTUH cohort by using the training procedure identical to that used for the CamCAN pre-trained model) by using multiple paired t tests. The Shrout and Fleiss 2,1 type model of intraclass correlation coefficient (ICC) was used to measure the linear consistency between the ages predicted at two timepoints (Weir, 2005).

For clinical sensitivity analysis, the transferred model in the NTUH cohort was applied to patients with SZ and matched controls. The CamCAN pre-trained and NTUH reference models were also applied to these individuals to predict their brain age. The predicted age difference (PAD = predicted age – chronological age) was calculated in the patient and control groups to quantify the degree of aging. The PADs between the patient and control groups derived from the transferred model, the CamCAN pre-trained model, and the NTUH reference model were compared using multiple analyses of covariance while regressing out age and sex.

3. Results

3.1. Brain age modeling in the CamCAN dataset

The six-layer cascade neural network model using 10-fold cross-validation on the CamCAN cohort training set demonstrated strong linear correlation and low MAE between chronological and predicted age based on tensor features (rho = 0.947, 95% confidence interval [CI] = 0.942–0.951; MAE = 4.57, CI = 4.45–4.74) and advanced features (rho = 0.958, CI = 0.952–0.964; MAE = 4.16, CI = 3.97–4.36). These models also accurately predicted the brain age in the CamCAN independent test set based on tensor features (rho = 0.926, CI = 0.922–0.929; MAE = 5.71, CI = 5.57–5.85) and advanced features (rho = 0.943, CI = 0.941–0.945; MAE = 4.68, CI = 4.59–4.77). Fig. 2 depicts the scatter plots of the predicted brain age in both CamCAN training and test sets.

3.2. Global searching of transfer learning and refined optimization

3.2.1. Appraisal of tuning sample size and transfer learning approach against other approaches

Fig. 3 presents the experiment results of the agile optimization method. The metrics of Pearson correlation coefficient and MAE are displayed with the blue series and red series of lines, respectively. The procedure of comparing the performance of different features, transfer learning approaches, and sample sizes is described below. First, we compared which feature type (tensor features or advanced features) could have a lower MAE given the maximal tuning sample size to determine which one was favorable in each site. Next, we considered which transfer learning approach using as low samples as possible could achieve the optimal performance, which was the MAE derived from the model adopting the single approach with the maximal tuning sample size (indicated by the yellow horizontal lines in Fig. 3). Meanwhile, the amount of tuning sample size was determined in each site (indicated by the yellow vertical lines in Fig. 3). Finally, we loosened the appraised criteria to meet the suboptimal performance (indicated by the orange vertical lines in Fig. 3). Notably, when the horizontal lines in Fig. 3 are not illustrated, it means the performance derived with the minimal tuning sample size suffices to meet the selection criteria.

In the NTUH cohort, the model using advanced features outperformed that using tensor features. Compared with the other approaches, the transfer learning approach achieved low MAE under the same tuning sample size. Considering the MAE of the model adopting the single approach and the maximal tuning sample size, the transferred model using advanced features and the transfer learning approach achieved optimal performance with 192 tuning samples (36% data saved) and suboptimal performance with 75 tuning samples (75% data saved). The estimated tuning sample size here was calculated by interpolating between the global searching results.

In the HH cohort, the model using tensor features outperformed that using advanced features. The hybrid approach (TLCO) achieved low MAE under the same tuning sample size. The transferred model using tensor features and the hybrid approach achieved optimal performance with 19 tuning samples (84% data saved) and suboptimal performance with 10 tuning samples (92% data saved).

In the Guys cohort, the model using tensor features outperformed that using advanced features. The transfer learning approach achieved low MAE under the same tuning sample size. The transferred model using tensor features and the transfer learning approach achieved optimal performance with 66 tuning samples (45% data saved) and suboptimal performance with 10 tuning samples (92% data saved).

Table 2 lists the best combinations of tuning sample sizes and transfer approaches determined from the global searching experiment. The MAE of the test set made by the “optimal” transferred model was significantly lower than that made by the “suboptimal” transferred model only in NTUH cohort (NTUH: paired t(104) = 2.56, p = 0.012; Table 3). Although the MAE comparisons between those derived from the models
with the optimal and suboptimal criteria in HH and Guys cohorts did not reach significant difference (HH: paired \( t(55) = 0.63, p = 0.529 \); Guys: paired \( t(62) = 0.81, p = 0.419 \); Table 3), the optimal criteria still suggest that MAE can be decreased with more tuning samples.

### 3.2.2. Refined optimization

After determining the suitable feature type, transfer approach, and tuning sample size for each site, the refined optimization was conducted to improve model performance by using the GDX algorithm for optimization and random search for hyperparameter setting. In our advanced tuning, the elapsed time of GDX optimization (9202 sec elapsed for 100 times modeling) was approximately 7 times that of the SCG algorithm in agile optimization (1317 sec elapsed for 100 times modeling). The average MAE in the refined model was significantly decreased by 0.53 (\( z = 18.42, p < 0.001 \)), 0.86 (\( z = 5.56, p < 0.001 \)), and 0.77 (\( z = 6.21, p < 0.001 \)) years in the NTUH, HH, and Guys cohort, respectively compared to the agile version (Table 3). The correlation between chronological and predicted age increased in each cohort. After refined optimization, the transferred models based on the optimal and suboptimal criteria were applied to the test set in each site and compared with the performance of the CamCAN original model without adjustment. Given the suitable transfer learning approach and feature type with refined optimization, the performance of the CamCAN pre-trained model applied on the CamCAN cohort could be reproduced in each target domain with a smaller amount of tuning data. Fig. 4 demonstrates the predicted age derived from the CamCAN model (i.e., apply the pre-trained model without any fine-tuning) and the transferred models with suboptimal and optimal transfer performance. The MAE of the test set made by the refined optimal transferred models in each site was significantly lower than that made by the CamCAN pre-trained model. The statistical results were as follow: NTUH: paired \( t(104) = 10.58, p < 0.001 \); HH: paired \( t(55) = 3.85, p < 0.001 \); Guys: paired \( t(62) = 4.37, p < 0.001 \). In particular, compared with the other sites, MAE in the NTUH cohort decreased most considerably.

### 3.3. Test–retest reliability

Based on the previous results, the transferred model “NTUH-AF-TL-192” from the NTUH cohort was applied to the NTUH longitudinal data. The ICC of the predicted age between the two timepoint measurements of the CamCAN model, transferred model, and reference model (trained using 300 tuning samples from the NTUH cohort) were 0.795, 0.950, and 0.961, respectively, indicating that the repeatability of the predicted age of the transferred model was similar to that of the reference model and was apparently better than that of the original CamCAN model. The AE-WSPA of the transferred model (3.08 years; \( t(45) = 3.18, p = 0.003 \)) and that of the reference model (3.24 years; \( t(45) = 2.23, p = 0.031 \)) were significantly lower than that of the original CamCAN model (4.59 years), highlighting a lower deviation between two timepoint measurements when the transferred model was adopted (Fig. 5). No significant difference was noted in AE-WSPA between the transferred and reference models (\( t(45) = 0.39, p = 0.6998 \)).
In the reference model from the NTUH cohort, the PADs were significantly higher in patients with SZ than in controls (mean PAD-SZ = 4.39 years, SD = 8.07 years; mean PAD-control = 0.09 years, SD = 4.85 years; F(1,314) = 36.15, p < 0.001; Fig. 6). Similarly, for the transferred model, the PADs of patients were significantly higher than those of controls (mean PAD-SZ = 4.63 years, SD = 8.02 years; mean PAD-control = 0.26 years, SD = 5.00 years; F(1,314) = 33.76, p < 0.001), indicating a discerning power comparable to that of the reference model. However, the PADs generated by the CamCAN model substantial increased in both groups, resulting in no significant difference between the two groups (mean PAD-SZ = 18.30 years, SD = 9.06 years; mean PAD-control = 16.87 years, SD = 8.57 years; F(1,314) = 1.608, p = 0.21). In addition, the PADs of the control group did not differ significantly between the transferred and reference models (t(159) = 1.40, p = 0.164).

### 4. Discussion

The dMRI-based brain age prediction model provides a valid indicator to track an individual’s aging status of the cerebral white matter, which may assess the risk of age-related neurological diseases. However, the interscanner variability of dMRI signals prohibits model...
Table 3
Results of test data in each site predicted by corresponding transferred model by using the agile and refined optimization methods.

| Optimization | Agile optimization | Rho | MAE      | Rho | MAE      |
|--------------|--------------------|-----|----------|-----|----------|
| NTUH-AF-TL-192 | 0.945 (0.944,0.946) | 0.949 (0.949,0.950) | 5.141 (5.110,5.172) | 4.784 (4.766,4.801) |
| NTUH-AF-TL-75  | 0.921 (0.919,0.924) | 0.937 (0.936,0.937) | 6.107 (6.003,6.212) | 5.413 (5.386,5.440) |
| HH-TF-TLCO-19  | 0.900 (0.890,0.910) | 0.916 (0.916,0.917) | 6.058 (5.765,6.351) | 5.354 (5.323,5.386) |
| HH-TF-TLCO-10  | 0.882 (0.863,0.900) | 0.922 (0.921,0.923) | 6.619 (6.077,7.162) | 5.605 (5.569,5.641) |
| Guys-TF-TL-66   | 0.805 (0.799,0.811) | 0.863 (0.862,0.865) | 6.472 (6.360,6.584) | 5.644 (5.612,5.677) |
| Guys-TF-TL-10   | 0.782 (0.757,0.807) | 0.831 (0.830,0.833) | 6.786 (6.308,7.264) | 6.080 (6.049,6.112) |

The transferred model in each site was fine-tuned according to the optimal and suboptimal combination from the global searching. Code in the left column = site, suitable feature type, transfer method, and required tuning sample size. AF: advanced features; TF: tensor features; TL: transfer learning; TLCO: transfer learning with cotrain. Performance measures include rho and MAE.

---

Fig. 4. Performance of the brain age models applied to the test set from each target domain. The table shows the performance of the test set (i.e., rho and MAE) between chronological and predicted age. In each target domain, the brain age was predicted by three prediction models including the original CamCAN model (no adjustment), transferred model with suboptimal performance (suboptimal transfer), and transferred model with optimal performance (optimal transfer). Here, the transferred model was optimized using the refined optimization method during the transfer process. The scatter plots show the distribution of chronological versus predicted age. The dots are coded with grayscale shades to indicate the AE between the chronological and predicted age.
For the first time, our study demonstrates that this problem can be overcome by appropriate transfer learning. The pre-trained model was constructed using the CamCAN cohort and was transferred to three target datasets acquired with different imaging protocols. By tuning the models with relatively small data size and certain feature types, optimal transferred models were obtained with satisfactory prediction performance in all three target cohorts. The test–retest reliability of the transferred model was verified in the two timepoint measurements. When applying the transferred model to patients with SZ, significant brain aging was found in patients as compared with healthy controls. Therefore, the transfer learning approach is feasible to generalize the dMRI-based brain age prediction model.

The global searching results indicated that the pre-trained model constructed using the source domain (CamCAN: 3T two-shell scheme) can be transferred to three independent target domains (NTUH: 3T half-sphere grid-sampling scheme, HH: 3T single-shell scheme, Guys: 1.5T single-shell scheme) with satisfactory prediction performance. Using suitable sample sizes, feature types, and transfer approaches, the transferred model could achieve prediction performance comparable to that of the pre-trained model. The transferred model worked equally well for different target domains acquired using various imaging protocols (NTUH: rho = 0.949, MAE = 4.78 years with 192 tuning samples; HH: rho = 0.916, MAE = 5.35 years with 19 tuning samples; Guys: rho = 0.863, MAE = 5.64 years with 66 tuning samples). If the same amount of tuning data was used, in most cases, the transfer learning approach facilitated better prediction than did the models trained from scratch (single or statistical covariate approach).

The pre-trained model using tensor and advanced features predicted an individual’s age with MAEs of 5.71 and 4.68 years, respectively, consistent with the values reported by Mwangi et al. (2013) and Richard et al. (2018). Moreover, the model using advanced features provided better prediction performance than did that using tensor features under the image acquisition conditions of the CamCAN cohort. The imaging parameters of the CamCAN cohort were designed to estimate the diffusion kurtosis tensor on top of the diffusion tensor (Shafto et al., 2014). By using this imaging protocol, the non-Gaussian diffusion indices derived from higher-order terms of the average propagator provide detailed information on the microstructure beyond the diffusion tensor, enabling better brain age prediction (Coutu et al., 2014; Gong et al., 2014). Although the advanced features allow to provide better prediction than the tensor features, it is suggested that combining both tensor and advanced features would develop a prediction model with better performance.

Different imaging scenarios including acquisition schemes, main magnetic fields, and imaging parameters may affect SNR and diffusion measures, and these factors may contribute to the intersite variability of dMRI datasets, prohibiting data pooling for model training or generalization of an established brain age prediction model to new datasets obtained from other sites. Although a satisfactory pre-trained model in the source domain data is used, the prediction outcome can be severely biased if the model is directly applied to the datasets from other institutions. Our results indicated that if no adjustment was made to the pre-trained model, the target domain data acquired using 3T DSI scheme (NTUH cohort) yielded poor prediction performance (MAE = 13.9 years)—worse than the data acquired using 3T single-shell scheme (HH cohort, MAE = 8.34 years). The difference in performance may be related to the distance of feature distributions between target and source domains. Given that the acquisition scheme of the HH cohort (3T single-
shell scheme) is similar to the CamCAN scheme (3T two-shell scheme), the feature distribution of the HH cohort is possibly closer to that of the CamCAN cohort than that of the NTUH cohort (3T half-sphere grid-sampling scheme). This speculation can be clarified using more datasets with varied acquisition schemes in shell numbers and b values.

The transfer learning approach not only achieved satisfactory performance in the target domain but also improved data usage efficiency. In our global searching results, the models tuned using the transfer learning approach achieved better performance (i.e., lower MAE) than the newly trained models using the same amount of tuning data. In addition, if we used the maximal amount of tuning data to train a model from scratch, the model produced a comparable accuracy compared with that of the transferred model which merely required fewer tuning data, demonstrating the efficiency of the transfer learning approach. Considering optimal performance as the goal of transfer learning, the transfer learning approach economized tuning data by 36%, 84%, and 45% data-saving rates for the NTUH, HH, and Guys cohorts, respectively. If the goal was to achieve suboptimal performance, the approach reduced the amount of tuning data by 75%, 92%, and 92% data-saving rates for the NTUH, HH, and Guys cohorts, respectively.

Our global searching of the tuning sample size for the suboptimal performance provides two insights: (1) It provides a reference to seek a compromised solution under limited data conditions and (2) It clarifies the relationship between data cost and performance improvement. Specifically, improving performance from suboptimal to optimal performance requires approximately 120 additional individuals for the NTUH cohort, but less than 10 additional individuals for the HH cohort. These results suggest that data acquired with imaging settings similar to that of the HH cohort (3T single-shell) are most suitable for transferring the brain age prediction model originally trained using the CamCAN cohort.

Compared with the statistical covariate approach (cortrain), the tuning methods with the transfer learning approach provided lower MAE in most of the global searching results (Fig. 3). When the tuning sample size increased, MAE became similar among the different approaches, except
for the TLCO approach, which used tensor features in the HH cohort (Fig. 3B). The TLCO approach attained apparently lower MAE than did other methods when tuning data increased. Because feature distributions between the HH and CamCAN cohorts are similar, tuning data combined with all source domain data enables more efficient optimization of the pre-trained parameter setting to attain better prediction.

Transfer learning approaches provide a useful and practical solution for domain adaptation in the field of machine learning. In the real world, dataset shift usually occurs because the joint distribution of inputs and outputs differs between the training and inference stages (Quionero-Candela et al., 2009). Dataset shift basically comprises three types of shift: covariate shift, target shift, and concept shift (Kouw et al., 2018). Covariate shift refers to the changes in the distribution of covariates (i.e. independent variables, inputs) while the conditional distribution of target variable is conserved across domains (Bouvier et al., 2019). In this case, both transfer learning and statistical covariate approaches and their combination are able to deal with this dataset shift problem. In our results, the combination approach (i.e. TLCO) appears to improve the predictions better when the target distributions are similar between source and target domains. Target shift refers to the changes in the distribution of the target variable (i.e. dependent variable, output), occurring when the distributions of target variables are extremely imbalanced between source and target domains. Concept shift denotes the shift in the relationship between covariates and target variable. Transfer learning instead of statistical covariate method allows to ameliorate the impact of the latter two shifts by re-weighting the mapping between feature spaces and the target variable across domains (Kouw et al., 2018). Given the minimal assumptions of dataset shift between domains, transfer learning provides a better general solution to fine-tune the pre-trained model for domain adaptation as we have observed in our global searching experiment.

Notably, our global searching was conducted using the agile optimization method to save the elapsed computation time. The resulting transferred model was trained using the refined optimization method, which adopted the dedicated optimizer and customized hyperparameters tuning to customize the suitable tuning settings for different datasets. In our experience, setting mean square error can prevent the model from overfitting by the outliers. Also, in the cascade neural network architecture for brain age modeling, the approach of re-training the entire pre-trained model usually achieved better performance compared to that of tuning the parameters in some layers and freezing the rest. Briefly, we demonstrated that refined optimization can improve model performance with the same amount of tuning data and thus achieve higher data efficiency, being recommended to the readers who would use our script to tune the model.

Feature types may affect model performance depending on the dMRI acquisition imaging protocols. In our global searching results, the transferred model of the NTUH cohort using advanced features exhibited better prediction than did that using tensor features. Theoretically, GFA and NG indices are estimated using higher-order terms of Hermite functions, thus being able to capture subtle and additional information of microstructural property and more sensitive to white matter aging as compared to the conventional tensor metrics (Gong et al., 2014; Teipel et al., 2014). However, in the HH and Guys cohorts, the models using tensor features yielded lower MAE (i.e., better prediction) than those using advanced features. The single-shell acquisition scheme with standard sampling schemes and typical b values (1000 s/mm²) can hardly provide useful information beyond the diffusion tensor (Jensen et al., 2005; Yan et al., 2013). When mean apparent propagator MRI reconstruction is applied to data acquired using a simple diffusion schemes and low b value setting such as those used in the HH and Guys cohorts, the advanced indices, such as NG series, can become overwhelmed by the noise, yielding a poor brain age prediction outcome. On the contrary, the advanced schemes, such as DSI, provide stable estimation of advanced features, enabling model performance improvement. If the advanced acquisition scheme is available, it is recommended that combining all metrics including the advanced features and tensor features might be helpful in improving the accuracy of a brain age prediction model.

SNR in the dMRI data linearly varies with the main magnetic field changes. Higher main magnetic field provides higher SNR; this can reduce uncertainty in diffusion index estimation (Alexander et al., 2006; Polders et al., 2011). Given the similarities in the imaging acquisition scheme used here, our results demonstrated that the transferred model tuned using 3T scanner data provided lower MAE and higher linear consistency of the predicted age than that tuned using 1.5T scanner data. Our results indicate that the SNR of the target domain is a critical factor when applying the transfer learning approach.

The test–retest reliability of the transferred model was satisfactory, indicating that the transferred learning approach is useful in longitudinal studies. Our results demonstrated that the two timepoint measurements of the 3T DSI data achieved high test–retest reliability (ICC = 0.950) for brain age predicted by the refined transferred model NTUH-AF-TL-192, which was comparable to that predicted by the NTUH reference model (ICC = 0.961). AE-WSPAs were also satisfactory, suggesting that the brain age predicted by the transferred model is similar in reproducibility to that predicted by the reference model. The test–retest reliability of diffusion measures derived from dMRI has been demonstrated (Boekel et al., 2017; Zhou et al., 2018). Our results further provide consistent reliability in brain age prediction for transferred models. High test–retest reliability supports the use of the transferred model in longitudinal follow-up, which can aid in detecting the aberrant aging trajectory. Further research on between-scanner or between-site reliability on traveling individuals across sites is warranted.

Our results also confirmed the sensitivity of the transferred model in detecting advanced brain aging in patients with SZ. The transferred model showed prediction performance comparable to that of the reference model. In patients, the mean PAD calculated by the transferred model was 4.63 years, which is close to PAD of 4.39 years calculated by the reference model. In controls, the mean PAD calculated by the transferred and reference models were also comparable (0.26 and –0.09 years, respectively). By contrast, when the model trained using the CamCAN cohort was directly applied to the NTUH cohort, PAD considerably increased in both patients and controls—compromising the power of discerning advanced brain aging in patients with SZ. Our results indicate that the transfer learning approach not only improves brain age prediction accuracy but also maintains the sensitivity of advanced brain aging in patients with SZ.

Transfer learning is an approach that generalizes the established dMRI-based brain age prediction model to other dMRI datasets from different sites or scanners. However, if the goal is to find statistical inference among sites such as that in meta-analysis, transfer learning is useless because the transfer learning framework does not assume any statistical premises. This is the main difference between machine learning and classical statistics, and one should choose the appropriate approach to meet the research purpose.

In the topic of brain age estimation through machine learning, the actual metric used to present biological aging of brain is called predicted age difference (PAD, also known as brain age delta, or brain age gap), which is calculated by the brain’s estimated age minus the individual’s chronological age. In most brain age studies, the derived PAD has a significant negative correlation with the chronological age in the training and test sets, resulting in an age-related bias. This bias can be minimized by statistical adjustment (Cole et al., 2015; Chen et al., 2019) or corrected by establishing a correction model. For transferring a brain age model estimation (Beheshti et al., 2019; Smith et al., 2019). In the scope of the present study, the transfer learning framework focuses on the generalization of brain age models per se. Therefore, we suggest that after readers implement a model with the transfer learning approach, the abovementioned correction approaches should be considered as a post hoc adjustment.

There are some limitations in this study. First, this approach depends on the tuning sample from the target domain for the adjustment of pre-trained model parameters. The small amount of tuning data can affect
the distribution of the model parameters in the target domain. If the tuning sample cannot represent the distribution of the data population in the target domain, the adjusted parameters of the transferred model may become biased, potentially leading to erroneous brain age prediction. In addition, the age of the tuning sample should be evenly distributed across lifespan to ensure balanced parameter adjustment. Second, our brain age framework involved dMRI preprocessing (dMRI reconstruction and tract-based feature extraction) and model building. Because the transfer learning approach intends to adjust the prediction model parameters alone, the format of model inputs for the source and target domains should be identical. The dMRI preprocessing in the source and target domains should adopt the same analytical pipeline to extract common feature types from dMRI data. Third, in the two phases of the experiment (global searching and refined optimization), we did not perform nested cross-validation because no extra data were available in the target site. Optimistically biased results might occur and affect the determination of hyperparameters in the second phase. Further tests in public are needed to address this issue.

In summary, we constructed brain age prediction models by using the dMRI dataset from the CamCAN repository and used the transfer learning approach to transfer the established models from the source domain to the target domains, with satisfactory model performance. By using the two-shell dMRI data (\(b = 1000, 2000 \text{ s/mm}^2\)) acquired using 3T MRI scanner as the pre-trained model materials, the transfer learning approach aimed in economizing the amount of tuning data needed to achieve optimal performance at the data-saving rates of 36%, 84%, and 45% for the data acquired using 3T DSI (\(b_{\text{max}} = 4000 \text{ s/mm}^2\)), 3T single-shell (\(b = 1000 \text{ s/mm}^2\)), and 1.5T single-shell (\(b = 1000 \text{ s/mm}^2\)) schemes, respectively. Under the same amount of tuning data were used, the transferred model achieved the best predictive performance when the 3T single-shell data with tensor features using TLCO approach were used, indicating that the data acquired using a clinically feasible imaging protocol, such as the current protocol, are the most suitable for model transferring. The use of advanced and tensor features was preferable if the data were acquired using advanced and single-shell acquisition schemes, respectively. In addition, the transfer learning approach demonstrated satisfactory test–retest reliability and clinical sensitivity in revealing advanced brain aging in patients with SZ. The scripts, code used to establish brain age prediction model and perform transfer learning approach, and the pre-trained model are available online. The readers who are interested in our method can utilize our framework (scripts available on our github repository) to create their own model and transfer the model to new datasets, even the model is proposed to conduct other tasks such as classification. In conclusion, we demonstrated an effective approach which allows generalization of the dMRI-based brain age prediction model efficiently. We also provided a general guideline for applying the transfer learning approach in different data acquisition scenarios.

Author's contribution

C.C. and W.T. conceived the study and were in charge of overall planning. L.Y., Y.T., W.L. helped in data collection of diffusion MRI datasets and verified the numerical results of the simulation. C.L., T.H. and H.H. enrolled the clinical participants. C.C. and Y.H. performed the diffusion image analysis. C.C. designed the prediction model, and conducted the experiments, statistical analyses and results visualization. C.C., L.Y., Y.T., W.L. and W.T. contributed to the interpretation of the results. C.C. and W.T. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Research data for this article

The data from Cambridge Centre for Ageing and Neuroscience (http://www.mrc-cbu.cam.ac.uk/datasets/camcan) and Information eXtraction from Images (http://brain-development.org/ixi-dataset/) organizations are available online in accordance with their data access declaration. The data acquired from National Taiwan University Hospital (NTUH) are not available due to confidentiality agreement of NTUH Research Ethics Committee. Scripts of brain age modeling and transfer learning method are available in an online open-access repository. Code of imaging process conditionally available upon request from the corresponding author.

Ethical approval

All procedures performed in this study involving human participants from the National Taiwan University Hospital (NTUH) were in accordance with the ethical standards of the NTUH Research Ethics Committee (REC) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent in the study was obtained from all individual participants who were recruited in the NTUH.

Declaration of competing interest

The authors declare that they have no financial/non-financial and direct/potential conflict of interest.

CRediT authorship contribution statement

Chang-Le Chen: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Visualization. Yung-Chin Hsu: Conceptualization, Software, Validation, Data curation. Li-Ying Yang: Validation, Data curation. Yu-Hung Tung: Validation, Data curation. Wen-Bin Luo: Validation, Data curation. Chih-Min Liu: Resources. Tzung-Jeng Hwang: Resources. Hai-Gwo Hwu: Resources. Wen-Yih Isaac Tseng: Conceptualization, Supervision, Writing - review & editing, Project administration, Funding acquisition.

Acknowledgement

This research was supported by National Health Research Institute (NHRI) Taiwan (grant: NHRI-EX109-10928NI) and partially supported by Ministry of Science and Technology (MOST) Taiwan (grant: 106-2314-B-002-242-MY3). We are grateful to the data source of the CamCAN project (http://www.mrc-cbu.cam.ac.uk/datasets/camcan/) and the IXi project (http://brain-development.org/ixi-dataset/). The CamCAN research project was supported by the Biotechnology and Biological Sciences Research Council (Grant #BB/H008217/1), and the European Union’s Horizon2020 research and innovation programme (Grant agreement #732592). We thank Wallace Academic Editing for assistance with English editing that greatly improved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116831.

References

Alexander, A.L., Hurley, S.A., Sammon, A.V., Adluru, N., Honeinbor, A.P., Mossahebi, P., et al., 2011. Characterization of cerebral white matter properties using quantitative magnetic resonance imaging stains. Brain Connect. 1, 423–446.
Alexander, A.L., Lee, J.E., Lazar, M., Field, A.S., 2007. Diffusion tensor imaging of the brain. Neurotherapeutics 4, 316–329.
Alexander, A.L., Lee, J.E., Wu, Y.C., Field, A.S., 2006. Comparison of diffusion tensor imaging measurements at 3.0 T versus 1.5 T with and without parallel imaging. Neuroimaging Clin. 16, 299–309 xi.
Andersson, J.L.R., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063–1078.
Ashburner, J., Friston, K.J., 2011. Diffusomorphic registration using geodesic shooting and Gauss-Newton optimisation. Neuroimage 55, 954–967.
Avram, A.V., Sarlis, J.E., Barnett, A.S., Ozsencan, E., Thomas, C., Irfanoglu, M.O., et al., 2016. Clinical feasibility of using mean apparent propagator (MAP) MRI to characterize brain tissue microstructure. Magn. Reson. Imaging 35, 431–442.

Banerjee, I., Crawley, A., Bhethanabothula, M., Daldrop-Link, H.E., Rubin, D.L., 2018. Transfer learning on fused multi-parametric MR images for classifying histopathological subtypes of rhabdomyosarcoma. Comput. Med. Imag. Graph. 65, 62–71.

Benshti, I., Nugent, S., Potvin, O., Duchesne, S., 2019. Bias-adjustment in neuroimaging-based brain age frameworks: a robust scheme. Neuroimage: Clinical. 24, 102063.

Benett, L.J., Madden, D.J., 2014. Disconnected aging: cerebral white matter integrity and age-related differences in cognition. Neuroimage 276, 187–205.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. J. Mach. Learn. Res. 13, 281–305.

Bockting, G., Forstmann, B.U., Pfeuffer, M.C., 2017. A test-retest reliability analysis of diffusion measures of white matter tracts relevant for cognitive control. Psychophysiology 54, 43–53.

Bouvier, V., You, P., Hudelot, C., Chastagnol, C., 2019. Hidden Covariate Shift: A Minimal Assumption for Domain Adaptation kXPreprint arXiv:1907.12299.

Cao, Y., Miller, M.I., Mori, S., Winslow, R.L., Younes, L., 2006. Diffeomorphic matching of diffusion tensor images. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2006. p. 67.

Cattani, M., Frithche, D.H., 2005. The rises and falls of disconnection syndromes. Brain 128, 2224–2239.

Chen, Y.J., Lo, Y.C., Fan, C.C., Hwang, T.J., Liu, C.M., et al., 2015. Automatic lesion segmentation. Med. Image Comput. Comput. Assist. Interv. 10435, 516–524.

Chen, C.-L., Shih, Y.C., Liou, H.H., Hsu, Y.C., Lin, F.H., Tseng, W.Y.I., 2019. Premature white matter aging in patients with right mesial temporal lobe epilepsy: a machine learning approach. Neuroimage: Clinical. 24, e20205.

Christopher, M., Belghith, A., Bovd, C., Proudfoot, J.A., Goldbaum, M.H., Wehren, R.N., et al., 2018. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci. Rep. 8, 16685.

Cole, J.H., 2017. Neuroimaging 2018. NeuroImage. 173, 1–17.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing. Neuroimage: Clinical. 24, 102033.

C.-L. Chen et al. NeuroImage 217 (2020) 116831

Chen, Y.J., Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., et al., 2001. Diffusion and decay in diffusion-weighted magnetic resonance imaging, and reconstruction of signal-to-noise ratios in MR images: in diffusion imaging. Magn. Reson. Med. 46, 473–481.

Chen, C.-L., Shih, Y.C., Liou, H.H., Hsu, Y.C., Lin, F.H., Tseng, W.-Y.I., 2019. Premature white matter aging in patients with right mesial temporal lobe epilepsy: a machine learning approach. Neuroimage: Clinical. 24, e20205.

Christopher, M., Belghith, A., Bovd, C., Proudfoot, J.A., Goldbaum, M.H., Wehren, R.N., et al., 2018. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci. Rep. 8, 16685.

Cole, J.H., 2017. Neuroimaging 2018. NeuroImage. 173, 1–17.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing. Neuroimage: Clinical. 24, 102033.
prediction based on diffusion tensor imaging and brain morphometry. PeerJ 6, e5908.
Saha, S., Fagnozzi, A., George, J., Colditz, P.B., Boyd, R., Rose, S., et al., 2018. Investigating brain age deviation in preterm infants: a deep learning approach. Data driven treatment response assessment and preterm, perinatal, and paediatric image analysis. In: First International Workshop, DATRA 2018 and Third International Workshop, PIPPI 2018, Held in Conjunction with MICCAI 2018. Springer, Granada, Spain, pp. 87–96.
Schonack, H.G., van Haren, N.E., Nieuwenhuis, M., Hulshoff Pol, H.E., Cahn, W., Kahn, R.S., 2016. Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study. Am. J. Psychiatr. 173, 607–616.
Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., et al., 2014. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurol. 14, 204.
Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M.K., Sun, L., et al., 2018. 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. IEEE Trans. Med. Imag. 37, 1522–1534.
Smith, S.M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T.E., Miller, K.L., 2019. Estimation of brain age delta from brain imaging. Neuroimage 200, 528–539.
Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., et al., 2017. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. Neuroimage 144, 262–269.
Tuch, D.S., 2004. Q-ball imaging. Magn. Reson. Med. 52, 1358–1372.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.
Valizadeh, S.A., Hanggi, J., Merillat, S., Jancke, L., 2017. Age prediction on the basis of brain anatomical measures. Hum. Brain Mapp. 38, 997–1008.
Wang, J., Krol, M.J., Tiulpin, A., Dubost, F., de Bruijne, M., Vernooij, M.W., Adams, H.H.H., Ikrar, M.A., Niessen, W.J., Roshchupkin, G.V., 2019. Gray matter age prediction as a biomarker for risk of dementia. Proc. Natl. Acad. Sci. U.S.A. 116, 21213–21218.
Wedeen, V.J., Hagmann, P., Tseng, W.Y., Reese, T.G., Weisskoff, R.M., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. 54, 1377–1386.
Weir, J.P., 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J. Strength Condit Res. 19, 231–240.
Weston, P.S., Simpson, L.J., Ryan, N.S., Ourselin, S., Fox, N.C., 2015. Diffusion imaging changes in grey matter in Alzheimer’s disease: a potential marker of early neurodegeneration. Alzheimer’s Res. Ther. 7, 47.
Xiao, T., Liu, L., Li, K., Qin, W., Yu, S., Li, Z., 2018. Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. BioMed Res. Int. 2018, 4605191.
Yan, X., Zhou, M., Ying, L., Yin, D., Fan, M., Yang, G., et al., 2013. Evaluation of optimized b-value sampling schemas for diffusion kurtosis imaging with an application to stroke patient data. Comput. Med. Imag. Graph. 37, 272–280.
Zhou, X., Sakaie, K.F., Debblins, J.P., Narayanan, S., Fox, R.J., Lowe, M.M., 2018. Scan-rescan repeatability and cross-scanner comparability of DTI metrics in healthy subjects in the SPRINT-MS multicenter trial. Magn. Reson. Imaging 53, 105–111.
Zhu, T., Hu, R., Qiu, X., Taylor, M., Tsu, Y., Yiannoutsos, C., et al., 2011. Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study. Neuroimage 56, 1398–1411.