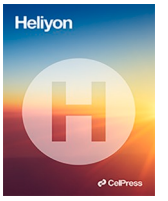
Heliyon 10 (2024) e32375

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Predicting cerebrovascular age and its clinical relevance: Modeling using 3D morphological features of brain vessels

Hwan-ho Cho [a], Jonghoon Kim [b], Inye Na [b], Ha-Na Song [c], Jong-Un Choi [c,d], In-Young Baek [c], Ji-Eun Lee [c], Jong-Won Chung [c], Chi-Kyung Kim [e], Kyungmi Oh [e], Oh-Young Bang [c], Gyeong-Moon Kim [c], Woo-Keun Seo [c,d,**], Hyunjin Park [b,f,*]

[a] *Department of Electronics Engineering, Incheon National University, Incheon, South Korea*
[b] *Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea*
[c] *Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea*
[d] *Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, South Korea*
[e] *Department of Neurology, Korea University Guro Hospital, Korea University College of Medicine, Seoul, South Korea*
[f] *Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea*

ARTICLE INFO

Keywords:
Cerebrovascular morphology
Age prediction
Machine learning
Cardiovascular disease
Risk factors
Personalized marker

ABSTRACT

Aging manifests as many phenotypes, among which age-related changes in brain vessels are important, but underexplored. Thus, in the present study, we constructed a model to predict age using cerebrovascular morphological features, further assessing their clinical relevance using a novel pipeline.

Age prediction models were first developed using data from a normal cohort (n = 1181), after which their relevance was tested in two stroke cohorts (n = 564 and n = 455). Our novel pipeline adapted an existing framework to compute generic vessel features for brain vessels, resulting in 126 morphological features. We further built various machine learning models to predict age using only clinical factors, only brain vessel features, and a combination of both. We further assessed deviation from healthy aging using the age gap and explored its clinical relevance by correlating the predicted age and age gap with various risk factors.

The models constructed using only brain vessel features and those combining clinical factors with vessel features were better predictors of age than the clinical factor-only model (r = 0.37, 0.48, and 0.26, respectively). Predicted age was associated with many known clinical factors, and the associations were stronger for the age gap in the normal cohort. The age gap was also associated with important factors in the pooled cohort atherosclerotic cardiovascular disease risk score and white matter hyperintensity measurements.

Cerebrovascular age, computed using the morphological features of brain vessels, could serve as a potential individualized marker for the early detection of various cerebrovascular diseases.

* Corresponding author. Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea.
** Corresponding author. Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea.
E-mail addresses: mcastenosis@gmail.com (W.-K. Seo), hyunjinp@skku.edu (H. Park).

https://doi.org/10.1016/j.heliyon.2024.e32375
Received 6 February 2024; Received in revised form 31 May 2024; Accepted 3 June 2024
Available online 4 June 2024

1. Introduction

Aging of the brain can present as multifaceted phenotypes, including shrinkage of brain volume, destructive changes in the subcortical network, decreased cognitive function, and an increased risk of stroke [1]. Owing to the overarching clinical and economic impacts of brain aging, researchers have striven to find ways to assess this phenomenon. Growing evidence suggests that predicted age (i.e., brain age) is clinically relevant; thus, increased brain age is generally considered a risk factor for various neurological diseases [2–5]. In particular, this feature is widely adopted not only to predict chronological age, but also to evaluate the association of biological or clinical factors with the age gap (i.e., the difference between predicted and chronological age) [2,6]. This is because the age gap can reflect the degree of abnormality, and thus might serve as a surrogate marker of brain health [6–8].

Age-related changes in brain vessels, known as cerebrovascular aging, occur as part of the normal brain aging process. Cerebrovascular aging is characterized by increased inflammation, endothelial dysfunction, stiffening of cerebral arteries, and narrowing of the vascular lumen [9]. These changes in the brain vessels are associated with stroke, vascular cognitive dysfunction, and

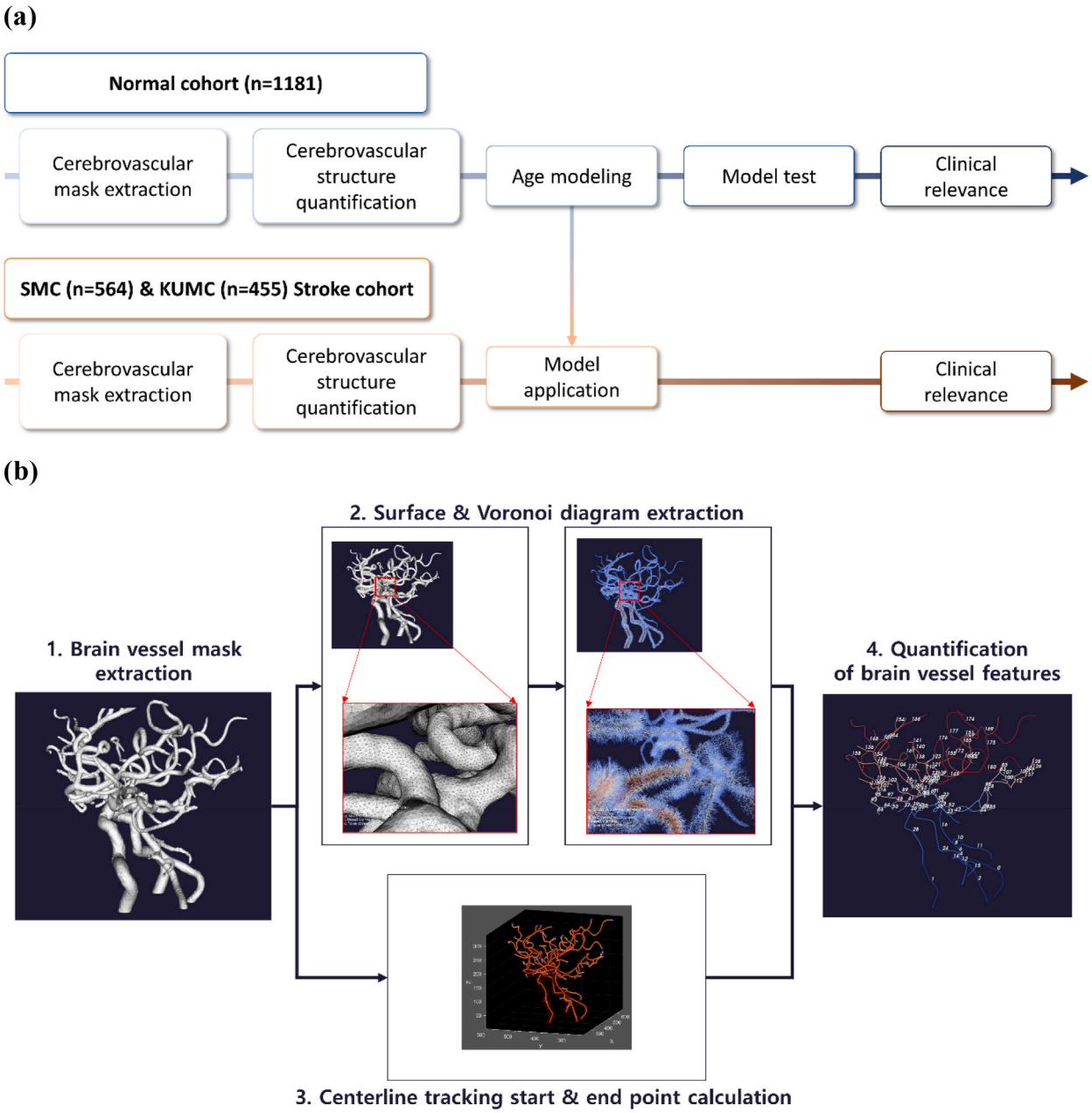


Fig. 1. (a) The overall workflow of the study. (b) The procedures followed to compute the cerebrovascular geometrical features.

neurodegenerative diseases [10]. Therefore, age prediction based on brain imaging data, particularly cerebrovascular age, is an important topic in neuroimaging research [7,8]. In previous studies, regression-based machine learning using structural neuroimaging, such as T1-weighted magnetic resonance imaging, has been used to estimate brain age in healthy controls, and the model has been applied to diseased cohorts to assess abnormalities [8]. However, quantitative approaches using vascular structural images, such as magnetic resonance angiography (MRA), are relatively rare, because standardized pipelines for cerebrovascular imaging analysis are limited.

The Vascular Modeling Toolkit (VMTK), which was originally developed to model aortic hemodynamics and assess various geometrical vessel features, is a representative method for computing generic vessel features from routine medical imaging data [11–13]. The VMTK is an open-source Python library that provides various low-level image-processing functions tailored for the quantification of tubular structures using medical images. As such, it is useful for building pipelines for cerebrovascular analysis. The framework cannot be directly applied to brain vessels because of the difference in vessel size, but serves as a starting point for further development.

In this study, we aimed to predict age using cerebrovascular morphological features and assess their clinical relevance in both normal and stroke cohorts (Fig. 1a). To this end, we constructed a pipeline to extract the morphological features of brain vessels based on VMTK [11], and subsequently used it to predict age. To explore the clinical relevance of this model, the association between predicted age and various risk factors, including the pooled atherosclerotic cardiovascular disease (ASCVD) risk score [14] and white matter hyperintensity measurements [15], was evaluated.

2. Methods

2.1. Data

This retrospective multicenter study included health screening data from a healthy control cohort, as well as the medical records of two cohorts from stroke registries. Data from the healthy control cohort were obtained from a previously constructed digitalized cerebrovascular atlas database. For this database, we selected patients who underwent health screening tests at the Health Promotion Center of the Samsung Medical Center between 2014 and 2016. Among those who visited the center, we selected subjects who fulfilled

Table 1
Demographic data of the normal cohort.

Clinical variables	Level	Training (n = 827)	Test (n = 354)	p-value
Age		53.378 (9.038)	54.223 (9.939)	0.154
Sex				0.398
	Female	389 (0.470)	176 (0.497)	
	Male	438 (0.530)	178 (0.503)	
Weight		65.077 (11.653)	65.531 (12.217)	0.545
Height		165.478 (7.857)	165.175 (8.295)	0.550
ASCVD risk		5.330 (5.945)	5.974 (7.077)	0.148
WMHV		2.303 (4.530)	1.948 (3.662)	0.266
WMHC		11.977 (6.978)	11.778 (7.512)	0.707
SBP		118.456 (15.725)	119.836 (16.974)	0.178
DBP		73.177 (10.412)	72.065 (10.861)	0.097
HTN				0.686
	No	624 (0.755)	271 (0.766)	
	Yes	203 (0.245)	83 (0.234)	
Cholesterol		192.940 (36.533)	194.514 (34.492)	0.490
HDLC		59.814 (16.037)	60.647 (16.455)	0.417
LDLC		122.530 (32.959)	125.215 (32.591)	0.198
HL				0.068
	No	598 (0.723)	274 (0.774)	
	Yes	229 (0.277)	80 (0.226)	
DM				0.132
	No	746 (0.902)	329 (0.929)	
	Yes	81 (0.098)	25 (0.071)	
Smoker				0.493
	No experience	519 (0.628)	214 (0.605)	
	Ex-smoker	184 (0.223)	90 (0.254)	
	Current smoker	124 (0.150)	50 (0.141)	
Alcohol				0.120
	No	329 (0.398)	133 (0.376)	
	1~2 cups per week	107 (0.129)	61 (0.172)	
	Half bottle	137 (0.166)	53 (0.150)	
	One bottle	194 (0.235)	91 (0.257)	
	Over on bottle	60 (0.073)	16 (0.045)	

Note. *ASCVD* atherosclerotic cardiovascular disease; *WMHV* white matter hyperintensity volume; *WMHC* white matter hyperintensity lesion count; *SBP* systolic blood pressure; *DBP* diastolic blood pressure; *HTN* hypertension; *HDLC* high-density lipoprotein cholesterol; *LDLC* low-density lipoprotein cholesterol; *HL* hyperlipidemia; *DM* diabetes mellitus. Numbers in parentheses represent standard deviations for continuous variables.

the following criteria: (1) available cerebrovascular time-of-flight (TOF) and fluid attenuation recovery inverted (FLAIR) images; (2) available clinical information, including sex, age, ASCVD risk score, blood pressure, lipid profile, smoking status, hypertension, diabetes mellitus, height, and weight; and (3) no evidence of a previous stroke, aneurysm of the intracranial vessel, or neoplasm on brain imaging. We selected 1181 cases for whom brain MRA intracranial TOF images were available, spanning an age range of 16–88 years. The ASCVD risk score is a quantitative estimation of the ten-year calculated risk of atherosclerotic cardiovascular disease [14]. FLAIR images were used to calculate white matter hyperintensity measurements using the lesion segmentation tool (LST) in the SPM 12 pipeline [16].

To assess the clinical relevance of the predicted cerebrovascular age, two datasets for acute stroke registries from the Samsung Medical Center (SMC) (n = 564) and Korea University Guro Hospital (KUMC) (n = 455) were considered. Each stroke registry enrolled patients with acute stroke or transient ischemic attack within seven days of symptom onset. A diagnosis of stroke was made in patients who presented with focal neurological symptoms or signs, in whom relevant brain lesions were found on brain images, especially diffusion-weighted images. The age ranges of the two disease cohorts were 26–92 years for the KUMC and 20–95 for the SMC.

This study was approved by the Institutional Review Boards (IRBs) of Samsung Medical Center (SMC-2021-04-072) and Korea University Guro Hospital (KUMC- 2019GR0300). The requirement for informed consent for healthy controls was waived by the IRB owing to the retrospective nature of the study and the anonymization of personal information. Written informed consent was obtained from all the patients who participated in the stroke registry. The demographic information of the normal cohort is given in Table 1, while that of the diseased cohort is given in Supplementary Table S1.

2.2. Imaging acquisition

The intracranial arteries were imaged using a 3.0 T Philips Achieva magnetic resonance imaging scanner (Philips Medical Systems), equipped with a 32-element phased-array receiver head coil for the SMC normal and stroke registries, and a 1.5 T S Magnetom Sonata MRI scanner (Siemens Medical Solutions) equipped with a 16-channel phased-array head coil for external validation images of patients in the KUMC stroke registry. This study used whole-brain three-dimensional (3D) MRA images collected from each participant using the TOF protocol. With an isotropic voxel size configured to $0.284 \times 0.284 \times 0.284\ mm^3$, the parameters were as follows: echo time, 4.59 ms; repetition time, 22 ms; flip angle, 23°; RBW, 130 Hz/pixel; GRAPPA factor, 3; 32 reference lines for SMC registry images and echo time, 4.97 ms; repetition time, 37 ms; flip angle, 25°; RBW, 65 Hz/pixel; 16 reference lines for KUMC stroke registry images. The acquisition parameters for the FLAIR images were as follows: The SMC registry images had an echo time of 125 ms, a repetition time of 11000 ms, an inversion time of 2800 ms, a pixel spacing of 0.469×0.469 mm, a slice thickness of 5 mm, and a Magnetic Field Strength of 3 T. The KUMC stroke registry images had an echo time of 95 ms, repetition time of 10000 ms, inversion time of 2605 ms, pixel spacing of 0.599×0.599 mm, slice thickness of 5 mm, and Magnetic Field Strength of 3 T. An expert neurologist (SWK) compared the image quality between healthy and diseased cohorts using the SMC data, finding no differences. However, the image quality of the

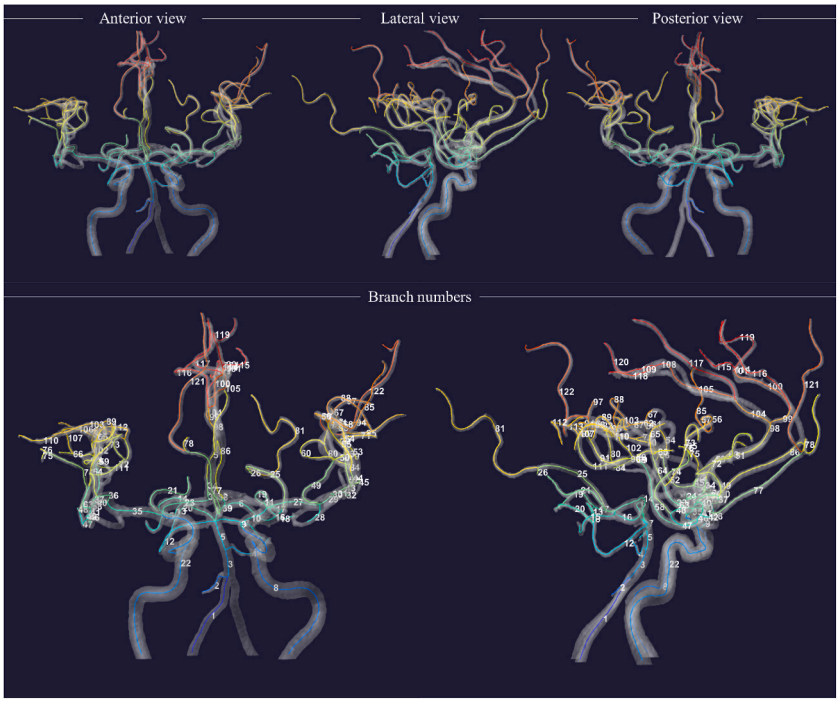


Fig. 2. The brain vessel mask, centerlines extracted from the vessel mask, and branch splitting results.

KUMC data was inferior to that of the SMC data because of the difference in the magnet strength.

2.3. Calculation of the cerebrovascular features

Fig. 1b shows the overall procedure for calculating cerebrovascular geometric features. We built our pipeline based on VMTK and made the following adaptations for brain vessel quantification: Non-contrast-enhanced MRA was used to extract vessel features for all cohorts. First, we extract a candidate brain vessel mask using a region-growing algorithm [17]. Three seed points were set on the internal carotid artery (ICA) and basilar artery (BA) in the TOF image, and the threshold was manually adjusted to improve the vessel mask. Brain vessel segmentation covers the segmental level of major cerebrovascular structures (e.g., M1-M4 segments of the ICA). A full list of extracted vessel names is provided in Supplementary Table S2. The segmented binary vessel masks were reviewed by a single investigator (SHN), and confirmed by a board-certified vascular neurologist (SWK). The reliability of the vessel segmentation was evaluated by comparing the features from the binary vessel mask with manual segmentation labels of a large vessel (i.e., the intracranial artery), created by one of the investigators (J-UC). Our investigations confirmed that the feature values, which are described below, from the manual and semiautomatic segmentation results were significantly correlated, with a p-value of less than 0.001.

The surface nodes and edges of the binary vessel mask were calculated using a marching-cube algorithm, and adjusted to isotropic triangles by applying spatial interpolation to the coordinates of each surface node [18]. The edges of each plane obtained by calculating a 3D Voronoi diagram that evenly divided the volume space between each surface node were used as centerline candidates [19]. Subsequently, centerline tracking was performed using the edges of the Voronoi diagram from the bottom part of the ICA or BA as the starting point and the termination of each brain vessel as the endpoint. To obtain the start and end points, 3D skeletonization was applied to a binary vessel mask [20]. The skeleton points were joined using a tree structure. All the leaf nodes of the tree were defined as endpoints, and the first node of the ICA or BA was defined as the starting point. The obtained centerlines were further merged and resampled at intervals of 0.2841 mm to match the MRA resolution, and divided into various branches based on bifurcation (Fig. 2).

Various morphological characteristics of the brain vessels were calculated using the obtained centerline. From all the centerline points, the sectional area, sectional perimeter, minimum diameter, maximum diameter, and circularity in the section perpendicular to the centerline direction were calculated. In addition, the curvature and torsion, which are structural characteristics of the centerline, were calculated. Subsequently, the diameter of the maximum inscribed sphere was obtained using the centerline and vessel mask surfaces. Finally, after assigning all the center points to the various vessel branches, the bifurcation angle of the bifurcation point was calculated. A full list of the computed geometrical features is provided in Supplementary Table S3.

All of the above processes were carried out using the open-source Python library VMTK (http://www.vmtk.org) [11–13], simple insight toolkit (SITK, https://sitk.org) [21,22], and sci-kit image (skimage, https://scikit-image.org) [23]. We built a wrapper software encompassing the entire procedure with a graphical user interface using pyQt5 for enhanced usability (https://github.com/Hwan-ho/BrainVesselFeatureExtractor).

2.4. Prediction of age using brain vessel features

To build age prediction models using vessel features, we computed 126 ($=14 \times 9$) features that were first-order statistics ($n = 14$) of each feature ($n = 9$) from all the centerline points. The control cohort data were divided into training (examinees who visited the SMC between 2014 and 2015, $n = 827$) and test (examinees who visited the SMC in 2016, $n = 354$) sets according to the time of visit for the health checkup. The features were not normalized to preserve the physical units of the structural measurements of the brain vessels. Feature selection was performed in two steps using only training data. In the first step, L1-normal regularization (i.e., the least absolute shrinkage and selection operator [LASSO]) was adopted [24]. To address the multicollinearity problem among the selected features in the first step, we applied recursive feature elimination based on the variance inflation factor (VIF). Among the features with a VIF greater than five, the feature with the largest VIF was removed until all features in the selected feature set had a VIF of less than five [25].

Our age prediction models were built using a random forest (RF) regressor [26] comprising 256 decision trees trained using a tree-bagging algorithm, an RF regressor comprising 256 trees trained using the least-squares boosting (LSBoost) algorithm for 1024 epochs with a learning rate of 0.01, a linear regressor, and an ensemble of the three previous models using the selected vessel features. The random forest model inherently possesses ensemble characteristics by utilizing multiple decision trees to introduce randomness. It is known for its ability to reduce overfitting and stabilize predictions. The linear regression model predicts the target variable as a linear combination of features and is the simplest and most commonly used model for regression tasks. In our ensemble approach, we set the predicted age as the average output of the three models. The prediction model was further refined using a linear transform. Additional linear transformation does not affect the performance metrics based on informational measures, such as correlation, but can improve the absolute error of the prediction. To evaluate the model performance, the correlation, adjusted R-squared, mean squared error (MSE), and mean absolute error (MAE) between the predicted and chronological ages were calculated.

We denoted the predicted age from the model as "cerebrovascular age," because it is based on vessel measures. Our designation was partly motivated by the "brain age" terminology proposed in other studies, where brain age is computed using brain imaging features. Several age-prediction models have previously been developed. Our baseline model (model 1) included sex, weight, and height. However, these factors are often referred to as potential sources of bias in age prediction [8]. We subsequently built four models (models 2–5); an RF regression model trained by a typical tree-bagging algorithm), an RF regression model trained by the LSBoost algorithm, a linear regression model, and an ensemble model averaging the prediction results of the three models. We built additional models (models 6–9) by combining the first four models (models 2–5) with sex, weight, and height information to probe the

incremental effects of the three factors. Nine models were constructed and compared.

To demonstrate the stability of the proposed models, we conducted a nested 5-fold cross-validation on the training set (n = 827), in which the training set was divided into two subsets, inner training (n = 662) and inner validation (n = 165), each time leaving a different validation set five different times. All of our models were constructed using the inner training set, and their performance was evaluated using the inner validation sets. The mean and standard deviation of the cross-validation were further reported.

2.5. Exploring the clinical relevance of predicted age

We employed two strategies to explore the clinical relevance of these models. First, the correlations between predicted age (i.e., cerebrovascular age) and the ten-year primary risk of ASCVD and conventional vascular risk factors were evaluated [14]. Second, the volume and number of cerebral white matter hyperintensity lesions measured from FLAIR images for the control group were correlated with chronological age and cerebrovascular age. White matter hyperintensity lesions were segmented by the lesion growth algorithm as implemented in the LST toolbox version 2.0.15 (www.statisticalmodelling.de/lst.html) for SPM [27]. Measurements of white matter hyperintensity can reflect an increased risk of cerebrovascular events such as stroke, dementia, and death [15]. Thus, they were considered for correlation analysis.

We also evaluated the association between the clinical parameters and the age gap, defined as the difference between predicted and chronological age (i.e., predicted age – chronological age), to explore the degree of abnormality. Existing age prediction studies have reported results that are biased toward the mean age of the training set, leading to underestimation in older adults and overestimation in younger adults [6]. To partially mitigate this effect, we applied a bias correction based on regressing chronological age before associating the age gap with clinical parameters [7].

We further applied our age prediction models to diseased cohorts from two medical centers to explore the diagnostic value of our model. Existing studies on predicting brain age have shown decreased performance when applied to diseased cohorts, owing to the shift in cohort characteristics [6,8]. Even with degraded performance, the application in the diseased cohort provides important insights, particularly with respect to the age gap. To apply our models to stroke cohorts, the brain vessel masks of stroke patients from two independent centers were used as inputs. The same brain vessel features were also calculated. All the parameters of each model were fixed and applied to the vessel features computed from the diseased cohorts. The clinical relevance was explored in the same manner as in the normal cohort.

2.6. Statistical analysis

We used two-sample Student's t-tests to compare continuous-valued information, and chi-square tests to compare categorized information in the demographic table. To compare model performance, bootstrapping was performed 100 times using a 0.2 hold-out ratio to construct the distribution of performance metrics, while a paired *t*-test was applied after normality was confirmed using the Shapiro-Wilk test. Statistical significance was set at $p < 0.05$. Bonferroni correction was further applied to associate model predictions with risk scores. RF modeling and all statistical analyses were conducted using the Statistics and Machine Learning Toolbox in

Table 2
Age prediction results in the normal cohort.

Training							
Model No.	Features	Model	R	Adj R2	p-value	MSE	MAE
1	Baseline		0.696	0.483	<0.001	42.212	5.082
2	Vessel features only	RF	0.901	0.812	<0.001	15.360	2.957
3		LSBoost	0.903	0.815	<0.001	15.079	3.022
4		Linear	0.405	0.163	<0.001	68.398	6.293
5		**Ensemble**	**0.864**	**0.747**	**< 0.001**	**20.667**	**3.669**
6	Vessel features combined with sex, height, and weight	RF	0.912	0.832	<0.001	13.711	2.835
7		LSBoost	0.908	0.823	<0.001	14.422	2.975
8		Linear	0.501	0.250	<0.001	61.241	6.053
9		**Ensemble**	**0.871**	**0.758**	**< 0.001**	**19.773**	**3.616**
Test							
Model No.	**Features**	**Model**	**R**	**Adj R2**	**p-value**	**MSE**	**MAE**
1	Baseline		0.285	0.079	<0.001	90.998	7.712
2	Vessel features only	RF	0.329	0.106	<0.001	88.353	8.019
3		LSBoost	0.339	0.112	<0.001	87.684	7.957
4		Linear	0.368	0.133	<0.001	85.657	7.179
5		**Ensemble**	**0.356**	**0.124**	**< 0.001**	**86.496**	**7.543**
6	Vessel features combined with sex, height, and weight	RF	0.436	0.188	<0.001	80.204	7.271
7		LSBoost	0.442	0.193	<0.001	79.678	7.257
8		Linear	0.507	0.255	<0.001	73.584	6.664
9		**Ensenble**	**0.481**	**0.229**	**< 0.001**	**76.154**	**6.818**

Note. *RF* random forest regressor trained by tree bagging; *LSBoost* random forest regressor trained by least-squares boosting; *R* correlation; *Adj R2* adjusted R-squared; *MSE* mean squared error; *MAE* mean absolute error.

MATLAB (MathWorks, Natick, MA, USA).

3. Results

3.1. Modeling of cerebrovascular age in the normal cohort

Table 2 shows the training and test results for age prediction in the normal cohort. Fig. 3a-d shows a plot of the predicted and chronological ages of the representative model (model 5, ensemble model). Plots from the other models are provided in the Supplementary Materials (Figs. S1–S8). The baseline model (model 1: age prediction using sex, weight, and height) showed a significant association with chronological age in both the training (correlation 0.696, p-value 1.3430×10^{-120}) and test sets (correlation 0.285, p-value 2.854×10^{-8}), owing to the natural bias of height and weight according to sex and age, as reported in other studies [8].

Age prediction improved for the models using vessel features (models 2–5) for the training and test sets compared with the baseline. In the test set, the correlation coefficient of the ensemble model (Model 5) using vessel feature information was 0.356, compared to 0.285 for the baseline model ($p < 0.001$). The MSE of the ensemble model was 86.496, compared with 90.998 for the baseline model in the test set ($p < 0.001$).

When the vessel features and other clinical factors such as sex, height, and weight were used together (models 6–9), the prediction was further improved in the test set. The correlation coefficient of the combined ensemble model (Model 9) was 0.481, compared with

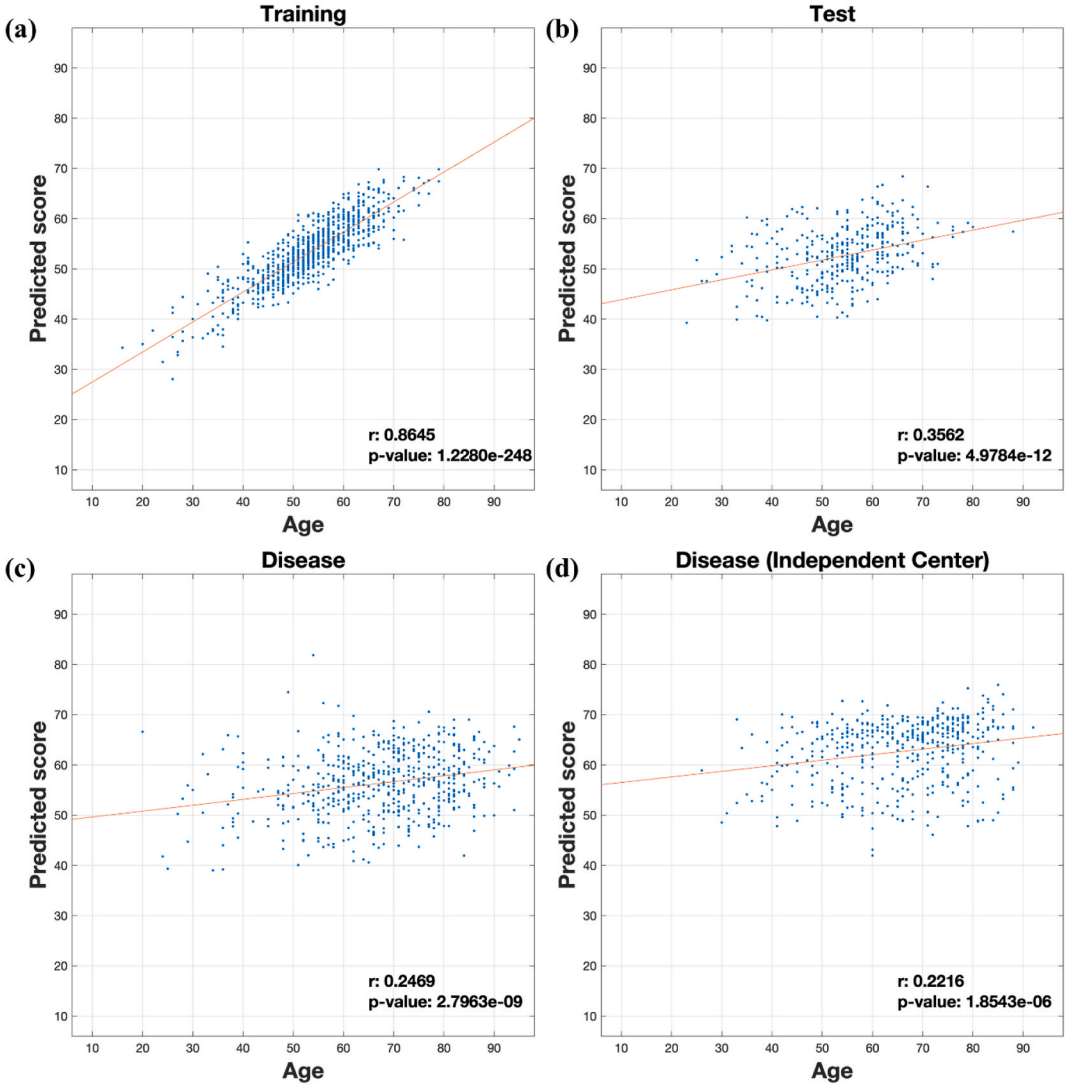


Fig. 3. Plots of the predicted vs. chronological age of the ensemble model that considered brain vessel features (model 5) for the (a) training set, (b) test set, (c) diseased SMC cohort, and (d) diseased KUMC cohort. Red lines are the regression lines for prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Correlation coefficient between the predicted age and clinical parameters in the normal cohort.

Training														
Model No.	Features	Model	ASCVD risk	WMHV	WMHC	SBP	DBP	HTN	Cholesterol	HDLC	LDLC	HL	DM	Smoke	Alcohol
1	Baseline		0.358	0.126	0.143	0.067	0.017	0.108	−0.055	−0.133	0.005	0.140	0.087	−0.058	−0.099
2	Vessel features	RF	0.606	0.291	0.368	0.091	0.058	0.177	−0.072	−0.141	−0.014	0.154	0.164	−0.025	−0.119
3	only	LSBoost	0.611	0.285	0.342	0.071	0.027	0.164	−0.080	−0.118	−0.025	0.131	0.157	−0.029	−0.138
4		Linear	0.352	0.243	0.319	0.088	0.178	0.168	−0.091	−0.149	−0.051	0.096	0.094	0.159	0.060
5		**Ensemble**	**0.595**	**0.298**	**0.372**	**0.087**	**0.072**	**0.181**	**−0.084**	**−0.142**	**−0.027**	**0.143**	**0.158**	**0.008**	**−0.100**
6	Vessel features combined with sex,	RF	0.553	0.267	0.323	0.066	0.013	0.154	−0.058	−0.098	−0.008	0.151	0.142	−0.098	−0.177
7	height, and weight	LSBoost	0.558	0.254	0.306	0.060	0.000	0.152	−0.066	−0.092	−0.013	0.130	0.134	−0.092	−0.166
8		Linear	0.289	0.201	0.213	0.064	0.086	0.123	−0.069	−0.072	−0.044	0.100	0.080	0.004	−0.064
9		**Ensenble**	**0.524**	**0.263**	**0.310**	**0.067**	**0.025**	**0.155**	**−0.068**	**−0.095**	**−0.019**	**0.139**	**0.133**	**−0.078**	**−0.157**
Test Model No.	**Features**	**Model**	**ASCVD risk**	**WMHV**	**WMHC**	**SBP**	**DBP**	**HTN**	**Cholesterol**	**HDLC**	**LDLC**	**HL**	**DM**	**Smoke**	**Alcohol**
1	Baseline		0.132	0.051	0.012	0.081	0.114	0.156	0.034	−0.079	0.052	0.101	0.115	0.057	0.043
2	Vessel features	RF	0.381	0.198	0.246	0.137	0.164	0.181	−0.073	−0.285	0.013	0.114	0.127	0.270	0.151
3	only	LSBoost	0.368	0.188	0.225	0.116	0.127	0.176	−0.100	−0.260	−0.021	0.135	0.149	0.239	0.119
4		Linear	0.415	0.279	0.311	0.117	0.179	0.129	−0.033	−0.273	0.052	0.123	0.185	0.264	0.097
5		**Ensemble**	**0.400**	**0.221**	**0.263**	**0.130**	**0.159**	**0.175**	**−0.079**	**−0.284**	**0.008**	**0.130**	**0.155**	**0.267**	**0.133**
6	Vessel features combined with sex,	RF	0.381	0.222	0.193	0.126	0.101	0.158	−0.028	−0.217	0.039	0.102	0.155	0.106	−0.014
7	height, and weight	LSBoost	0.295	0.201	0.196	0.112	0.105	0.193	−0.028	−0.211	0.040	0.126	0.147	0.098	0.005
8		Linear	0.344	0.300	0.234	0.123	0.084	0.153	−0.019	−0.156	0.039	0.160	0.209	0.054	−0.122
9		**Ensemble**	**0.334**	**0.246**	**0.216**	**0.126**	**0.104**	**0.180**	**−0.027**	**−0.210**	**0.042**	**0.133**	**0.174**	**0.095**	**−0.036**

Note. *RF* random forest regressor trained by tree bagging; *LSBoost* random forest regressor trained by least-squares boosting; *ASCVD* atherosclerotic cardiovascular disease; *WMHV* white matter hyperintensity volume; *WMHC* white matter hyperintensity lesion count; *SBP* systolic blood pressure; *DBP* diastolic blood pressure; *HTN* hypertension; *HDLC* high-density lipoprotein cholesterol; *LDLC* low-density lipoprotein cholesterol; *HL* hyperlipidemia; *DM* diabetes mellitus; *Underline* represent a significant association.

Table 4
Correlation coefficient between the corrected age gap and clinical parameters in the normal cohort.

Training														
Model No.	Features	Model	ASCVD risk	WMHV	WMHC	SBP	DBP	HTN	Cholesterol	HDLC	LDLC	HL	DM	Smoke	Alcohol
1	Baseline		−0.138	−0.054	−0.095	0.032	0.047	0.005	−0.003	−0.089	0.026	0.068	−0.034	0.024	0.047
2	Vessel features	RF	0.126	0.173	0.208	0.077	0.185	0.097	−0.006	−0.118	0.008	0.083	0.046	0.167	0.132
3	only	LSBoost	0.036	0.133	0.120	0.030	0.113	0.065	−0.025	−0.066	−0.016	0.028	0.028	0.159	0.082
4		Linear	0.149	0.173	0.228	0.068	0.206	0.117	−0.065	−0.119	−0.047	0.047	0.032	0.222	0.156
5		**Ensemble**	**0.115**	**0.175**	**0.202**	**0.064**	**0.184**	**0.102**	**−0.035**	**−0.111**	**−0.020**	**0.058**	**0.039**	**0.201**	**0.135**
6	Vessel features combined with sex,	RF	−0.001	0.127	0.115	0.020	0.086	0.042	0.029	−0.018	0.023	0.075	−0.011	0.002	0.005
7	height, and weight	LSBoost	−0.089	0.074	0.047	0.005	0.052	0.038	0.009	−0.005	0.011	0.025	−0.028	0.014	0.032
8		Linear	0.024	0.108	0.086	0.037	0.113	0.055	−0.035	−0.026	−0.039	0.040	0.000	0.067	0.042
9		**Ensemble**	**−0.020**	**0.113**	**0.091**	**0.024**	**0.094**	**0.050**	**−0.001**	**−0.018**	**−0.004**	**0.051**	**−0.014**	**0.032**	**0.030**
Test															
Model No.	**Features**	**Model**	**ASCVD risk**	**WMHV**	**WMHC**	**SBP**	**DBP**	**HTN**	**Cholesterol**	**HDLC**	**LDLC**	**HL**	**DM**	**Smoke**	**Alcohol**
1	Baseline		−0.009	−0.060	−0.092	0.052	0.120	0.084	0.063	−0.056	0.073	0.058	0.060	0.096	0.096
2	Vessel features	RF	0.234	0.084	0.146	0.107	0.175	0.100	−0.045	−0.271	0.036	0.065	0.065	0.328	0.219
3	only	LSBoost	0.210	0.069	0.121	0.085	0.137	0.092	−0.073	−0.244	0.001	0.085	0.086	0.297	0.190
4		Linear	0.255	0.161	0.207	0.083	0.194	0.034	0.001	−0.258	0.081	0.069	0.120	0.332	0.173
5		**Ensemble**	**0.240**	**0.100**	**0.157**	**0.097**	**0.172**	**0.086**	**−0.049**	**−0.270**	**0.033**	**0.077**	**0.089**	**0.332**	**0.207**
6	Vessel features combined with sex,	RF	0.111	0.066	0.048	0.088	0.115	0.048	0.014	−0.198	0.074	0.035	0.076	0.177	0.069
7	height, and weight	LSBoost	0.079	0.042	0.050	0.071	0.120	0.085	0.015	−0.191	0.075	0.061	0.066	0.169	0.093
8		Linear	0.108	0.129	0.069	0.080	0.100	0.022	0.032	−0.127	0.082	0.091	0.125	0.134	−0.036
9		**Ensemble**	**0.105**	**0.077**	**0.058**	**0.085**	**0.121**	**0.060**	**0.020**	**−0.190**	**0.082**	**0.063**	**0.090**	**0.175**	**0.056**

Note. *RF* random forest regressor trained by tree bagging; *LSBoost* random forest regressor trained by least-squares boosting; *ASCVD* atherosclerotic cardiovascular disease; *WMHV* white matter hyperintensity volume; *WMHC* white matter hyperintensity lesion count; *SBP* systolic blood pressure; *DBP* diastolic blood pressure; *HTN* hypertension; *HDLC* high-density lipoprotein cholesterol; *LDLC* low-density lipoprotein cholesterol; *HL* hyperlipidemia; *DM* diabetes mellitus; *Underline* represent a significant association.

0.285 for the baseline model ($p < 0.001$). The MSE of the combined ensemble model was 76.154, compared with 90.998 for the baseline model ($p < 0.001$). Taken together, these results confirmed the additive value of clinical factors for predicting age, in addition to vessel features.

We further observed some decreases in the performance metrics between the training and test sets. Such occurrences are common in machine learning models. The results of 5-fold cross-validations are given in Supplementary Table S4 and are largely consistent with those shown in Table 2.

3.2. Exploring the clinical relevance of cerebrovascular age

Table 3 shows the correlation coefficients between cerebrovascular age and various clinical parameters in our models. Significant associations between cerebrovascular age and risk scores were observed in both the training and testing sets. These were stronger for the models that reflected vessel features than for the baseline model. Overall, a positive correlation was found between cerebrovascular age and ASCVD risk score, indicating that higher cerebrovascular age is associated with a higher risk of stroke within ten years. Both the volume of white matter hyperintensity (WMHV) and the count of white matter hyperintensity lesions (WMHC) also showed significant positive associations with cerebrovascular age, indicating that higher cerebrovascular age is associated with a worse status of the hyperintense region. The results of both ASCVD and white matter hyperintensity measurements indicated that cerebrovascular age could reflect known risk factors for brain aging.

Cerebrovascular age was significantly correlated with blood pressure status. In the test set, the cerebrovascular age using vessel features showed a significant positive correlation with hypertension (HTN), implying that higher cerebrovascular age is associated with a worse blood pressure status. In contrast, the correlation coefficient between chronological age and HTN was 0.064, which was not significant ($p > 0.05$). This indicates that signs of hypertension that cannot be captured by a simple chronological age can be detected by measuring the cerebrovascular age. Overall, the baseline model showed a lower correlation with HTN than models that reflected vessel features.

Among the measures of cholesterol-related status, high-density lipoprotein (HDL) cholesterol levels showed a significant negative association with cerebrovascular age, indicating that HDL cholesterol levels tend to decrease with increasing cerebrovascular age. However, this trend was not observed to exist with chronological age (correlation coefficient −0.09). Hyperlipidemia also showed a significant positive correlation with cerebrovascular age, indicating that individuals with hyperlipidemia tended to have a higher cerebrovascular age. The probability of diabetes was also significantly and positively correlated with cerebrovascular age.

Incorporating clinical factors, such as sex, height, and weight, with vessel features enhances age prediction performance, but degrades the correlation between cerebrovascular age and certain clinical parameters. Specifically, the association between ASCVD risk, WMHC, and HDL cholesterol weakened in the test set when brain vessel features were combined with the aforementioned clinical factors. Therefore, using brain vessel features alone, without clinical factors, offers the advantage of capturing the clinical status of cerebrovascular health.

Table 4 shows the correlation coefficients between the age gap of each model and various clinical parameters. The age gap was defined as the predicted age minus the chronological age following bias correction (see Methods). Thus, a positive age gap indicates

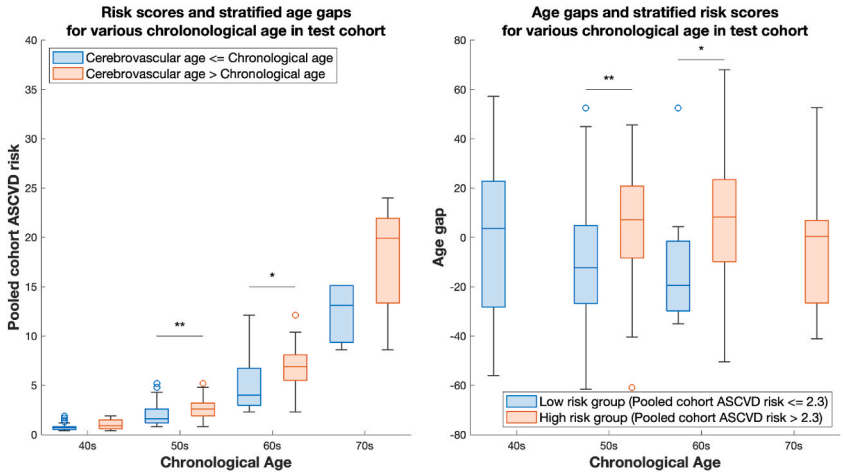


Fig. 4. Relevance of age gaps in different chronological age groups. (a) Box plot of the ASCVD risk score according to chronological age. Blue boxes represent a negative age gap (i.e., younger cerebrovascular age than the chronological age), while red boxes represent a positive age gap (i.e., older cerebrovascular age than the chronological age). Asterisks indicate significant differences in ASCVD risk between negative and positive age gap groups in the same chronological age (*p-value <0.05, **p-value <0.001). (b) Box plot of the age gap according to chronological age. Blue boxes represent the low ASCVD risk group, while red boxes represent the high ASCVD risk group. Asterisks indicate significant differences in the age gap between the low and high ASCVD risk groups in the same chronological age (*p-value <0.05, **p-value <0.001). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

that an individual's cerebrovascular age is older than their chronological age. The correlation coefficient between chronological age and age gap was close to zero after bias correction, indicating that the age gap and chronological age were largely uncorrelated and that the age gap could provide complementary information. The clinical relevance of the age gap was primarily observed when sex, weight, and height were not considered. Age gaps from the models using vessel features (models 2–5) showed significant positive associations with the ASCVD risk score, DBP, smoking, and alcohol intake. These results indicate that a cerebrovascular age older than the chronological age results in higher risk scores and higher blood pressure. HDL cholesterol also showed a significant negative association with age gaps when using vessel features (models 2–5), implying that cerebrovascular age is younger than chronological age if the HDL cholesterol level is high. A peculiar aspect of our results is that when sex, weight, and height information were included in age prediction, the statistical significance of the clinical relevance of the age gap decreased, particularly that for ASCVD risk. This suggests that performing age prediction without potential bias factors (i.e., sex, weight, and height) is advantageous for evaluating the clinical relevance of the age gap.

Our results showed that the age gap was significantly associated with the pooled ASCVD risk score. To explore the clinical relevance of the age gap in more detail, we investigated whether there was a difference in ASCVD risk scores according to age gap trends (i.e., positive age gap vs. negative age gap) within the same chronological age in the test cohort (Fig. 4a). In each chronological age group (40s, 50s, and 60s), individuals with a positive age gap had a significantly higher ASCVD risk score than those with a negative age gap in the same age group ($p < 0.001$, 0.004, and 0.005, respectively, for 40s, 50s, and 60s). In other words, individuals with a cerebrovascular age higher than their chronological age are more likely to develop atherosclerotic cardiovascular diseases, including stroke, within ten years compared with individuals of the same chronological age. In addition, we investigated whether there was a difference in the age gap according to ASCVD risk within the same age group (Fig. 4b). ASCVD risk groups were stratified through the application of the median value of the training cohort as a threshold. For each age group (40s, 50s, and 60s), the age gaps of individuals in the low ASCVD risk group were significantly higher than those in the high ASCVD risk group in the same age group (p-values 0.010, <0.001, <0.001 respectively for 40s, 50s, and 60s). Furthermore, the age gap in the low ASCVD group was negative, meaning that individuals with negative age gaps were predicted to have a lower ASCVD risk than their chronological age. In summary, the difference between the cerebrovascular age predicted by our model and the chronological age can be used as a robust measure to reflect the relative risk of cerebrovascular disease.

3.3. Application of models to diseased cohorts

Table 5 shows the age prediction results for the two cohorts. As in the normal cohort, age predictions in the diseased cohort showed significant correlations with the chronological age; however, the prediction errors were higher than those of the normal cohort. When vessel features were involved, the MSE values of the ensemble model were 171.400 and 156.173 in the SMC and KUMC stroke cohorts, respectively, compared with the MSE of the ensemble model for the normal test set of 86.496 ($p < 0.001$).

The cerebrovascular ages of individuals in their 40s–70s in the disease cohort are presented in Fig. 5. The cerebrovascular age was significantly greater than that of the normal cohort across all age ranges. Cerebrovascular age can indicate the structural health of brain vessels, while patients with stroke are more likely to have an abnormal vessel structure affected by the disease. As the disease

Table 5
Age prediction results in the diseased cohorts.

Diseased cohort 1 (SMC)							
Model No.	Features	Model	R	Adj R2	p-value	MSE	MAE
1	Baseline		0.358	0.126	<0.001	159.184	14.551
2	Vessel features only	RF	0.256	0.064	<0.001	170.575	14.228
3		LSBoost	0.242	0.057	<0.001	171.825	14.435
4		Linear	0.182	0.031	<0.001	176.465	14.527
5		**Ensemble**	**0.247**	**0.059**	**< 0.001**	**171.400**	**14.279**
6	Vessel features combined with sex, height, and weight	RF	0.407	0.164	<0.001	152.361	13.304
7		LSBoost	0.427	0.181	<0.001	149.172	13.048
8		Linear	0.409	0.165	<0.001	152.065	12.941
9		**Ensemble**	**0.443**	**0.195**	**< 0.001**	**146.677**	**13.003**
Diseased cohort 2 (KUMC)							
Model No.	**Features**	**Model**	**R**	**Adj R2**	**p-value**	**MSE**	**MAE**
1	Baseline		0.222	0.047	<0.001	156.164	13.350
2	Vessel features only	RF	0.181	0.030	<0.001	158.888	11.901
3		LSBoost	0.156	0.022	<0.001	160.244	11.469
4		Linear	0.274	0.073	<0.001	151.925	10.375
5		**Ensemble**	**0.222**	**0.047**	**< 0.001**	**156.173**	**10.943**
6	Vessel features combined with sex, height, and weight	RF	0.318	0.099	<0.001	147.619	10.913
7		LSBoost	0.276	0.074	<0.001	151.773	10.889
8		Linear	0.331	0.107	<0.001	146.293	10.144
9		**Ensemble**	**0.350**	**0.121**	**< 0.001**	**144.087**	**10.173**

Note. *RF* random forest regressor trained by tree bagging; *LSBoost* random forest regressor trained by least-squares boosting; *R* correlation; *Adj R2* adjusted R-squared; *MSE* mean squared error; *MAE* mean absolute error.

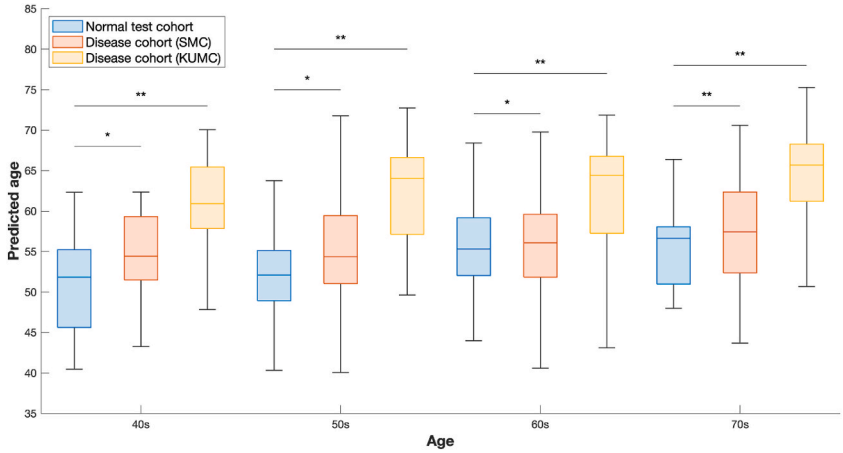


Fig. 5. Predicted age according to chronological age. Asterisks indicate significant differences in the predicted age between the normal and disease cohorts in the same chronological age range (*p-value <0.05, **p-value <0.001).

cohort had a high incidence of ASCVD (Supplementary Tables S5 and S6), the cerebrovascular age was expected to be significantly higher than that of the normal cohort. Consequently, the association between chronological age and cerebrovascular age was weaker because of the increased prediction error in the diseased cohorts compared to that in the normal cohort, as shown in Table 5.

Across all age ranges, the cerebrovascular age of the KUMC was higher than that of the SMC (Fig. 5). This could be due to the demographic differences between the two centers (Supplementary Table S1). Patients at KUMC are more likely to have hypertension and to exhibit higher cholesterol and lower HDL cholesterol levels, which can negatively affect their cerebrovascular health. These differences in clinical conditions may affect the structure of the brain vessels, leading to a higher predicted cerebrovascular age in patients at KUMC.

Associations between cerebrovascular age, corrected age gap, and various clinical parameters in the stroke cohorts have been reported (Tables S7 and S8). In contrast to the results obtained in the normal cohort, there was little to no association between cerebrovascular age, age gap, and clinical parameters in the stroke cohort. Because our models were constructed based on the structural characteristics of the normal cohort, it may be challenging to associate clinical characteristics with cerebrovascular age in the disease cohort. This difficulty may arise because individuals with a disease may undergo treatment, resulting in the influence of external factors, such as medications, that can affect various clinical conditions.

To evaluate the contribution of each vessel feature to age prediction, we dissected our linear model (Model 4), which was included in the ensemble model. Table 6 presents the selected vessel features and their multivariate linear coefficients against cerebrovascular age in the linear model (Model 4). Fig. 6 provides a visual representation of the contributions of different features to the various categories. All features had a VIF of less than 5, indicating that they were relatively independent. Of the 12 selected features, six were thickness features (50 %), one was a sectional shape feature (8 %), four were tortuosity features (33 %), and one was a sectional thickness feature (8 %) (Fig. 6a). The standard deviations of the minimum-maximum axis length ratio ($p = 0.003$), 25th percentile of curvature ($p = 0.049$), and maximum perimeter ($p = 0.013$) were significant. Fig. 6b presents a visualization of this data considering the p-value and sign of the coefficient of each feature (i.e., the length and direction of the bar denote the sign of the coefficient

Table 6
Contribution of various vessel features to the age prediction model (model 4; Linear). Bold text denotes significant features with a p-value less than 0.05.

Feature name	Coefficients	p-value	VIF	Category	Interpretation
(Intercept)	−17.182	0.176	–	–	–
MIS_Percentile 5	29.754	0.126	2.497	Thickness	Least Thickness
MIS_Percentile 75	3.973	0.342	3.167	Thickness	Greatest Thickness
MinD_Min	2.235	0.498	1.983	Thickness	Least minor axis
MinD_Max	2.000	0.218	4.473	Thickness	Greatest minor axis
MaxD_Percentile 25	11.251	0.070	3.228	Thickness	Least major axis
MaxD_Percentile 95	0.939	0.357	3.154	Thickness	Greatest major axis
MMR_STD	**−207.895**	**0.003**	**3.211**	**Sectional Shape**	**Sectional shape variation**
Curvature_Min	30.587	0.259	2.011	Tortuosity	Least Tortuosity
Curvature_Percentile 25	**80.303**	**0.049**	**2.243**	**Tortuosity**	**Least Tortuosity**
Torsion_Min	417.514	0.569	1.024	Tortuosity	Least Tortuosity
Torsion_IQR	−6.671	0.062	1.26	Tortuosity	Tortuosity variation
Perimeter_Max	**0.840**	**0.013**	**3.133**	**Surface**	**Greatest Perimeter**

Note. *MIS* maximum inscribed sphere; *MinD* minimum diameter; *MaxD* maximum diameter; *MMR* minimum-maximum axis length ratio; *STD* standard deviation; *IQR* interquartile range.

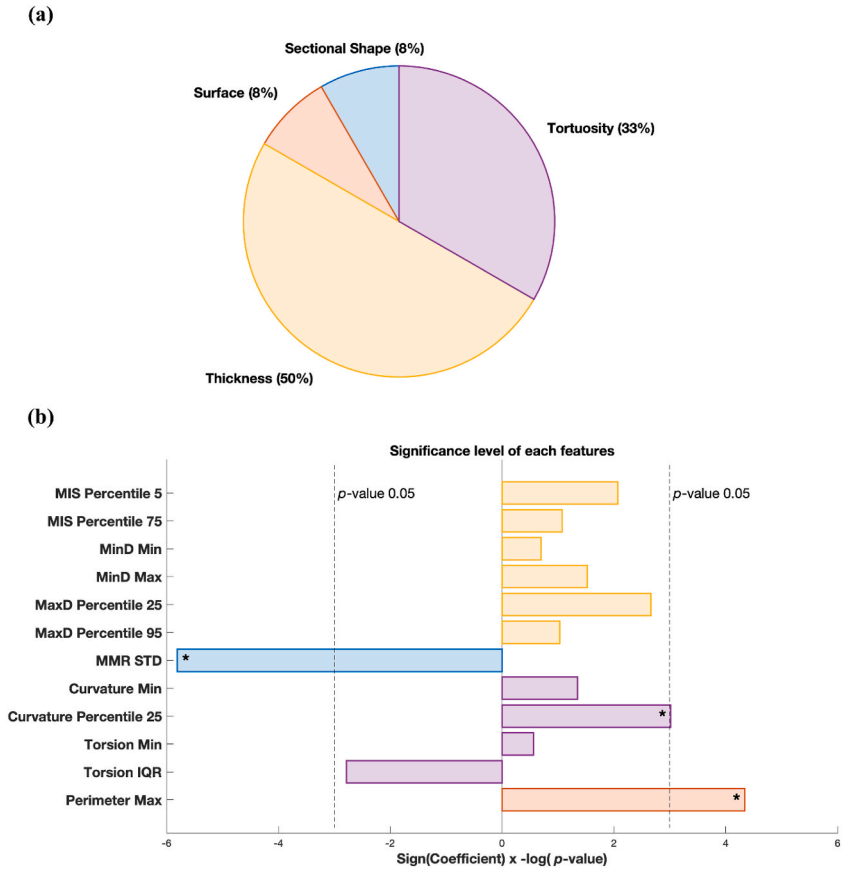


Fig. 6. Contribution of various vessel features to the age prediction model (a) ratio of the different feature components colored by categories, and (b) contribution of various features whose weights were computed as the negative log of p-value multiplied by the sign of the coefficient. Asterisks indicate significant features.

multiplied by-log (p-value)). Although there were many selected thickness features, none were significant. The standard deviation of the minimum-maximum axis length ratio represented the sectional shape variation, and had a negative coefficient, indicating that the diversity of the cerebrovascular sectional shape decreased with age. The 25th percentile of the curvature represents the lower 25th percentile value of the curvature distribution derived from the entire set of cerebrovascular nodes, and showed a positive coefficient, indicating that as people age, there is less chance of having a straight segment of the brain vessel, and the brain vessel becomes tortuous. The maximum perimeter had a positive coefficient, and was related to the greatest cross-sectional perimeter of the brain vessels, suggesting that an increase of 1 mm in the greatest cross-sectional perimeter would result in a 0.840 increase in age. Despite our exploration of multivariate coefficients in our linear model for clarity, our ensemble model incorporates nonlinear RF regressors. The contribution of each feature to the RF regressors can be evaluated based on the reduction in the error attributed to each feature. However, this method does not provide insight into the directionality of the contribution, which is crucial for interpreting machine learning models in medicine. Our assessment offers only a limited interpretation of the linear model; however, we believe that this partial information can provide a starting point for understanding the models.

4. Discussion

In the present study, we proposed a method to predict age using brain vessel features, and assessed its clinical relevance using various clinical parameters. Our pipeline for computing vessel features started with the brain vessel mask from TOF images obtained using the region-growing algorithm. The surface mesh of the vessel was subsequently obtained using the marching cube algorithm and refined to be close to the isosurface. The starting points of the BA or ICA and the endpoints of various vessels were automatically extracted from the 3D skeleton of the vessel mask. Based on the 3D Voronoi diagram, the centerlines of the blood vessels were tracked from every starting point to various endpoints. Various morphological features of the vessel were further computed along the centerline. The entire pipeline was constructed from in-house Python code and open-source Python software.

To predict age using brain vessel features, feature selection was performed using LASSO feature selection and recursive feature elimination using VIF. Age prediction models were constructed using RF regression with two training strategies, linear regression, and an ensemble of the three models. The association between cerebrovascular age and chronological age or various clinical parameters

was evaluated. The clinical relevance of this age gap was also explored. Before evaluating the clinical relevance of the age gap, bias correction was applied by removing the linear trend of the age effect. The models were then validated using data from normal and stroke cohorts extracted from two centers.

Vascular aging is a complex process involving multiple mechanisms, including oxidative stress, mitochondrial damage, genomic instability, and epigenetic alterations [28]. When developing treatments for age-related vascular diseases, it is essential to diagnose and monitor vascular age using clinically available tests. Arterial stiffness is the most well-known surrogate for monitoring vascular age [29]. Past research has shown that aortic stiffness markedly increases with advancing age, and is associated with an increased risk of cardiovascular disease [30]. Several markers may be used to assess vascular aging, including intima-medial thickness, pulse wave velocity, augmentation index, and echo-tracking methods [31]. Among these, carotid-femoral pulse wave velocity (cfPWV) is considered the gold standard for testing arterial stiffness. Indeed, in one population-based cohort study, increased cfPWV was identified as an independent predictor of coronary heart disease and stroke [32]. However, arterial stiffness measured by aortic pulse wave velocity lacks representativeness of cerebrovascular age, because the cerebral vasculature has distinctive histological and hemodynamic properties of the systemic vasculature. To date, no specific test has been introduced in clinical practice to directly assess cerebral arterial aging or stiffness.

We presented an ensemble model combining three models using vessel features (Model 5) as a representative model, considering both prediction performance and clinical relevance. When sex, height, and weight information were combined, age prediction improved, but the clinical relevance of cerebrovascular age and age gap weakened. One reason for this could be overfitting owing to known bias factors (i.e., sex, height, and weight). As such, among the models using vessel features, we chose Model 5 as the representative model, as it had the best age prediction performance and the strongest clinical relevance. The age prediction of Model 5 showed a correlation coefficient of 0.356 ($p < 0.001$) for the chronological age of the normal test set, and significant correlations with correlation coefficients of 0.247 ($p < 0.001$) and 0.222 ($p < 0.001$) in the SMC and KUMC cohorts, respectively. The model showed significant associations not only with the risk score, but also with blood pressure-related measures, such as SBP, DBP, and HTN, as well as with cholesterol-related risk factors, such as HDL cholesterol and hyperlipidemia. The age gap calculated from this model was also correlated with a variety of risk factors, including ASCVD, WHMC, DBP, and HDL cholesterol.

We further observed a drop in performance for the test set compared to the training set for our healthy cohort. Although this phenomenon is common in many cases, we believe that it could be attributed to the nature of the two RF-based models. By design, RF tends to fit closely to the training data; thus, the drop in performance in the test set could be aggravated compared to the linear regression model.

In our results for the disease cohort, discrepancies in MRI scanners and imaging protocols between the SMC and KUMC should be considered a limitation. Specifically, the differing magnetic field strengths at these centers (3 T at SMC and 1.5 T at KUMC) are likely to have influenced our vascular segmentation outcomes, due to an impact on the visualization of smaller arterial structures. As such, we conducted an additional analysis to assess the minimum arterial diameter in the two cohorts. Our findings revealed that the average minimum diameter was 0.1275 mm (SD = 0.0372 mm) at SMC and 0.1517 mm (SD = 0.0509 mm) at KUMC, with a statistically significant difference ($p < 0.001$). This significant variation suggests that a higher magnetic field strength can enhance the detection of finer vascular features.

The gap between the chronological and cerebrovascular ages can be used as an indicator of the extent to which a specific person is outside the normal range of health. Therefore, it can be used to measure the risk of aging or disease compared with a healthy aging group. In one study, Smith et al. showed that a strong correlation between predicted brain age and chronological age was not an optimal estimator of the age gap [6,7]. This is because the cerebrovascular age is biased toward the mean of the training cohort, owing to the nature of the optimization algorithm. We further corrected the age gap by removing the chronological age component from the cerebrovascular age. This analysis further showed significant associations between the corrected age gap and several non-imaging or imaging features using larger-scale (>10,000 participants) UK Biobank data. Similar to our study, the authors reported significant associations between age gap and blood pressure-related factors such as SBP and DBP, smoking, alcohol intake, diabetes, and cognitive status. In our results, the correlation between the age gap and cerebrovascular age was very high before chronological age was regressed out. Therefore, we removed the age effect to avoid biased results when exploring the clinical association with age gap.

In one prior cross-sectional study, Williamson et al. investigated the correlation between MRI-based cerebrovascular structure measures and modifiable cardiovascular risk factors in young adults aged 18–40 years with no clinical evidence of brain vascular disease [33]. The authors utilized TOF MRI to calculate cerebrovascular density, caliber, and tortuosity and assessed white matter hyperintensity and volume using FLAIR images. Through their analysis, various clinical parameters were found to be significantly associated with cerebrovascular caliber, tortuosity, and white matter hyperintensity. Similarly, our ensemble model (Model 5) demonstrated significant associations with all risk factors except cholesterol and LDL cholesterol levels. Specifically, the identified age gap was significantly associated with ASCVD, WMHC, DBP, HDL cholesterol, smoking, and alcohol intake. The association between cerebrovascular age and the clinical parameters identified in our study replicated these findings. We further observed that the vessel-only models had stronger correlations with cardiovascular risk factors than the combined models. The stronger correlation in the vessel-only model could be attributed to the intrinsic correlation between demographics and risk factors. However, it is generally observed that height and weight tend to decrease with age for both sexes in the Korean population [34,35]. As such, the effects of intrinsic correlation could be lower than expected.

This study has several limitations. First, our age prediction models were built using regular health checkup data from one hospital. It was difficult to obtain an independent normal cohort owing to privacy concerns. Therefore, the models were not validated in a normal external cohort. Second, our cohorts, both control and diseased, were concentrated in the age range of 40–70 years, leading to the underrepresentation of those outside the age range. However, the majority of the participants in the normal cohort also fell within

this age range (88.823 %). Third, the accuracy of the brain vessel feature extraction pipeline relies entirely on the quality of brain vessel mask segmentation. Consequently, if the brain vessel mask is not segmented appropriately, the subsequent vessel features may be inadequate. Fourth, there was a mismatch in age distributions between the healthy and diseased cohorts; thus, prediction in the out-of-range older patients in the diseased cohort could be biased. Therefore, the results in older patients should be interpreted with caution. Future studies should match the age distribution between the training and test cohorts to allow for an unbiased prediction. Fifth, TOF MRI is susceptible to various artifacts, including bias field inhomogeneity and ringing artifacts. The extended acquisition time required for TOF imaging makes it particularly vulnerable to motion artifacts. Signal dropouts can also occur in regions with turbulent blood flow, potentially affecting the results. These factors can lead to either over- or under-estimation of the vessel mask, ultimately affecting the extracted vessel features. Finally, the structural information of brain vessels can only reflect certain aspects of brain aging, and the physiological properties related to vascular aging still need to be validated. Unfortunately, physiological information on the brain vessels was not available in the present study. Despite these limitations, our method has the potential to be applied in clinics to aid diagnosis and risk assessment. However, careful validation of the proposed methods needs to be performed before they can be applied in the clinic.

In this study, we predicted age using the morphological features of major vessels at the segmental level and further demonstrated their clinical relevance. Our prediction models showed significant associations with known risk factors for cerebrovascular disease. Cerebrovascular age based on vessel features has clinical relevance as it reflects the degree of abnormality from healthy aging. This makes it a potential individualized marker for early detection of various cerebrovascular diseases. Future studies should assess improved cerebrovascular aging by combining structural and functional brain imaging. Combining structural insights from cerebrovascular morphology with perfusion MRI analysis would provide a comprehensive physiological understanding of brain vessel imaging. Use of contrast-enhanced MRA (CE-MRA) instead of regular non-contrast-enhanced MRA, may further improve vessel segmentation, and hence, the prediction results. However, in future work, we need to develop additional image preprocessing for CE-MRA to use the proposed pipeline. The plaque information can also be used to provide health information. However, it is difficult to extract plaque information because TOF images do not accurately reflect vessel wall information. We plan to incorporate plaque information into future studies.

Data availability statement

The pipeline for extracting brain vessel features will be made available on GitHub (https://github.com/Hwan-ho/BrainVesselFeatureExtractor) upon publication, and the experimental code will be shared upon request. Brain vessel data were accessed with permission from the IRB of the SMC.

CRediT authorship contribution statement

Hwan-ho Cho: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Jonghoon Kim:** Writing – review & editing, Data curation. **Inye Na:** Writing – review & editing. **Ha-Na Song:** Writing – review & editing, Resources, Data curation. **Jong-Un Choi:** Writing – review & editing, Investigation, Data curation. **In-Young Baek:** Data curation. **Ji-Eun Lee:** Data curation. **Jong-Won Chung:** Supervision, Resources, Data curation. **Chi-Kyung Kim:** Supervision, Data curation, Conceptualization. **Kyungmi Oh:** Supervision, Data curation. **Oh-Young Bang:** Writing – review & editing, Supervision, Project administration. **Gyeong-Moon Kim:** Writing – review & editing, Supervision, Project administration. **Woo-Keun Seo:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Hyunjin Park:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Research Foundation (NRF-2020M3E5D2A01084892) of South Korea. This research was further supported by the K-Brain Project of the National Research Foundation (NRF), funded by the Korean government (MSIT) (RS-2023-00265393).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e32375.

References

[1] R. Peters, Ageing and the brain, Postgrad. Med. 82 (2006) 84.
[2] J.H. Cole, K. Franke, Predicting age using neuroimaging: Innovative brain ageing Biomarkers, Trends Neurosci. 40 (2017) 681–690.
[3] J.H. Cole, S.J. Ritchie, M.E. Bastin, et al., Brain age predicts mortality, Mol. Psychiatr. 23 (23) (2018) 1385–1392. *5* 2017.
[4] N. Raz, K.M. Rodrigue, Differential aging of the brain: patterns, cognitive correlates and modifiers, Neurosci. Biobehav. Rev. 30 (2006) 730–748.
[5] T. Kaufmann, D. van der Meer, N.T. Doan, et al., Common brain disorders are associated with heritable patterns of apparent aging of the brain, Nat. Neurosci. 22 (22) (2019) 1617–1623, 10 2019.
[6] S.M. Smith, D. Vidaurre, F. Alfaro-Almagro, et al., Estimation of brain age delta from brain imaging, Neuroimage 200 (2019) 528–539.
[7] H. Peng, W. Gong, C.F. Beckmann, et al., Accurate brain age prediction with lightweight deep neural networks, Med. Image Anal. 68 (2021) 101871.
[8] L. Baecker, R. Garcia-Dias, S. Vieira, et al., Machine learning for brain age prediction: Introduction to methods and clinical applications, EBioMedicine 72 (2021) 103600.
[9] T. Yang, Y. Sun, Z. Lu, et al., The impact of cerebrovascular aging on vascular cognitive impairment and dementia, Ageing Res. Rev. 34 (2017) 15–29.
[10] R.E. Mrak, W.S.T. Griffin, D.I. Graham, Aging-associated changes in Human brain, J. Neuropathol. Exp. Neurol. 56 (1997) 1269–1275.
[11] R. Izzo, D. Steinman, S. Manini, et al., The vascular modeling toolkit: a Python library for the analysis of tubular structures in medical images, J. Open Source Softw. 3 (2018) 745.
[12] L. Antiga, M. Piccinelli, L. Botti, et al., An image-based modeling framework for patient-specific computational hemodynamics, Med. Biol. Eng. Comput. 46 (2008) 1097–1112.
[13] M. Piccinelli, A. Veneziani, D.A. Steinman, et al., A framework for geometric analysis of vascular structures: application to cerebral aneurysms, IEEE Trans. Med. Imag. 28 (2009) 1141–1155.
[14] G. Dc, L.-J.G.B. Dm, et al., ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines, Circulation 129 (2013), https://doi.org/10.1161/01.CIR.0000437741.48606.98. Epub ahead of print 24 June 2014.
[15] S. Debette, H.S. Markus, The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis, BMJ 341 (2010) 288.
[16] Schmidt P, Wink L. LST: A Lesion Segmentation Tool for SPM.
[17] R. Adams, L. Bischof, Seeded region growing, IEEE Trans. Pattern Anal. Mach. Intell. 16 (1994) 641–647.
[18] W. Heiden, T. Goetze, J. Brickmann, Fast generation of molecular surfaces from 3D data fields with an enhanced "marching cube" algorithm, J. Comput. Chem. 14 (1993) 246–250.
[19] F. Grélard, F. Baldacci, A. Vialard, et al., New methods for the geometrical analysis of tubular organs, Med. Image Anal. 42 (2017) 89–101.
[20] T.C. Lee, R.L. Kashyap, C.N. Chu, Building skeleton models via 3-D medial surface Axis thinning algorithms, CVGIP Graph. Models Image Process. 56 (1994) 462–478.
[21] B.C. Lowekamp, D.T. Chen, L. Ibáñez, et al., The design of simpleITK, Front. Neuroinf. 7 (2013) 45.
[22] Z. Yaniv, B.C. Lowekamp, H.J. Johnson, et al., SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research, J. Digit. Imag. 31 (2018) 290–303.
[23] S. Van Der Walt, J.L. Schönberger, J. Nunez-Iglesias, et al., Scikit-image: image processing in python, PeerJ 2014 (2014) e453.
[24] R. Tibshirani, Regression selection and shrinkage via the lasso, J. Roy. Stat. Soc. B (1996), https://doi.org/10.2307/2346178. Epub ahead of print.
[25] James Gareth, Witten Daniela, Hastie Trevor, et al., An Introduction to Statistical Learning, Springer, New York, 2013.
[26] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32. *1* 2001; 45.
[27] P. Schmidt, C. Gaser, M. Arsic, et al., An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis, Neuroimage 59 (2012), https://doi.org/10.1016/j.neuroimage.2011.11.032. Epub ahead of print.
[28] Z. Ungvari, S. Tarantini, A.J. Donato, et al., Mechanisms of vascular aging, Circ. Res. 123 (2018) 849–867.
[29] G.F. Mitchell, Arterial stiffness in aging: does it have a place in clinical practice? Hypertension 77 (2021) 768–780.
[30] Jae S. Young, K.S. Heffernan, S. Kurl, et al., Association between estimated pulse wave velocity and the risk of stroke in middle-aged men, Int. J. Stroke 16 (2021) 551–555.
[31] B. Jani, C. Rajkumar, Ageing and vascular ageing, Postgrad. Med. 82 (2006) 357–362.
[32] F.U.S. Mattace-Raso, T.J.M. Van Der Cammen, A. Hofman, et al., Arterial stiffness and risk of coronary heart disease and stroke, Circulation 113 (2006) 657–663.
[33] W. Williamson, A.J. Lewandowski, N.D. Forkert, et al., Association of cardiovascular risk factors with MRI indices of cerebrovascular structure and function and white matter hyperintensities in young adults, JAMA 320 (2018) 665–673.
[34] National Health Insurance Service, General health screening: status of average height distribution by gender, age, city, and province (2022), Statistics Korea (2024). https://kosis.kr/statHtml/statHtml.do?orgId=350&tblId=DT_35007_N130&conn_path=I2. (Accessed 7 May 2024).
[35] National Health Insurance Service, General health screening: status of average weight distribution by gender, age, city, and province (2022), Statistics Korea (2024). https://kosis.kr/statHtml/statHtml.do?orgId=350&tblId=DT_35007_N132&conn_path=I2. (Accessed 7 May 2024).