


RESEARCH

Open Access



# Clinical information prompt-driven retinal fundus image for brain health evaluation

Nuo Tong<sup>1,2</sup>, Ying Hui<sup>3</sup>, Shui-Ping Gou<sup>1</sup>, Ling-Xi Chen<sup>2</sup>, Xiang-Hong Wang<sup>4</sup>, Shuo-Hua Chen<sup>5</sup>, Jing Li<sup>6</sup>, Xiao-Shuai Li<sup>7</sup>, Yun-Tao Wu<sup>5</sup>, Shou-Ling Wu<sup>5</sup>, Zhen-Chang Wang<sup>7\*</sup>, Jing Sun<sup>7\*</sup> and Han Lv<sup>7\*</sup> 

## Abstract

**Background** Brain volume measurement serves as a critical approach for assessing brain health status. Considering the close biological connection between the eyes and brain, this study aims to investigate the feasibility of estimating brain volume through retinal fundus imaging integrated with clinical metadata, and to offer a cost-effective approach for assessing brain health.

**Methods** Based on clinical information, retinal fundus images, and neuroimaging data derived from a multicenter, population-based cohort study, the KaiLuan Study, we proposed a cross-modal correlation representation (CMCR) network to elucidate the intricate co-degenerative relationships between the eyes and brain for 755 subjects. Specifically, individual clinical information, which has been followed up for as long as 12 years, was encoded as a prompt to enhance the accuracy of brain volume estimation. Independent internal validation and external validation were performed to assess the robustness of the proposed model. Root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) metrics were employed to quantitatively evaluate the quality of synthetic brain images derived from retinal imaging data.

**Results** The proposed framework yielded average RMSE, PSNR, and SSIM values of 98.23, 35.78 dB, and 0.64, respectively, which significantly outperformed 5 other methods: multi-channel Variational Autoencoder (mcVAE), Pixel-to-Pixel (Pixel2pixel), transformer-based U-Net (TransUNet), multi-scale transformer network (MT-Net), and residual vision transformer (ResViT). The two- (2D) and three-dimensional (3D) visualization results showed that the shape and texture of the synthetic brain images generated by the proposed method most closely resembled those of actual brain images. Thus, the CMCR framework accurately captured the latent structural correlations between the fundus and the brain. The average difference between predicted and actual brain volumes was 61.36 cm<sup>3</sup>, with a relative error of 4.54%. When all of the clinical information (including age and sex, daily habits, cardiovascular factors, metabolic factors, and inflammatory factors) was encoded, the difference was decreased to 53.89 cm<sup>3</sup>, with a relative error of 3.98%. Based on the synthesized brain MR images from retinal fundus images, the volumes of brain tissues could be estimated with high accuracy.

<sup>†</sup>Nuo Tong and Ying Hui contributed equally to this work.

\*Correspondence:

Zhen-Chang Wang

cjr.wzhch@vip.163.com

Jing Sun

samantha\_sunjing@163.com

Han Lv

chrislvhan@126.com

Full list of author information is available at the end of the article



**Conclusions** This study provides an innovative, accurate, and cost-effective approach to characterize brain health status through readily accessible retinal fundus images.

*Trial registration No.* NCT05453877 (<https://clinicaltrials.gov/>).

**Keywords** Retinal fundus image, Brain volume, Brain health, Magnetic resonance imaging, Deep learning, Eye and brain connection

## Background

The human brain is the highest command center of the nervous system and regulates the functions of memory, language, and movement through a complex and multilevel neural network. The optimal integrity of brain structure, including an appropriate brain volume, is considered as a key aspect of brain health, while a prominent reduction in brain volume (brain atrophy) is an important indicator of brain degeneration [1–4].

Conventionally, brain tissue volumes are quantitatively measured based on brain magnetic resonance imaging (MRI) data. However, the high cost and time-consuming nature of MRI techniques have always been a great concern [5–8]. Conducting large-scale brain MRI collection within the general population is often impractical. Furthermore, in resource-poor or remote settings without MRI equipment and specific scenarios such as the aerospace field, MRI-based measurements are far from reach. This issue has long troubled the scientific community and will negatively impact medical equity and humanity's endeavors in aerospace exploration. Thus, further investigation is warranted to explore a cost-effective alternative method for estimating brain volume to evaluate brain health status within the general population.

Retinal fundus images are widely utilized in ophthalmologic practice. In comparison to brain MRI examination, the acquisition of retinal fundus images is characterized by lower costs, fewer preparation requirements, and shorter time. Moreover, in contrast to a heavy MRI scanner, the equipment for retina image collection is lightweight and portable. The retina and brain are intricately interconnected in terms of embryonic development and physiological anatomy [9, 10]. Thus, the retina may serve as a distinctive window into the brain's structures, functioning as a cost-effective tool for evaluating brain health.

Recent studies have explored the eye-brain connections through an image-based deep learning approach. Xie et al. [11] identified the degenerative changes in the geometrical parameters of retinal microvasculature in Alzheimer's disease (AD) and mild cognitive impairment (MCI) individuals, thus offering valuable insights for the diagnosis and clinical decision-making of AD and MCI. Additionally, Zhao et al. [12] investigated the cross-organ

connections between the eyes and brain at both phenotypic and genetic levels, and further observed that retinal imaging traits exhibited significant predictive power for neuropsychiatric and neurological disorders. However, a general limitation of these studies is their exclusive focus on extracting quantitative features and genetic phenotypes from the original images, resulting in the underutilization of retinal and brain imaging data, while the high-dimensional information embedded within the images remains insufficiently explored.

In this study, we aim to propose a novel cross-modal correlation representation (CMCR) framework to elucidate the intricate co-degenerative relationships between the eyes and brain using retinal fundus images and brain MRI data obtained from a large-scale population-based cohort study. Clinical metadata are employed as prompts to enhance the accuracy of brain volume estimation from retinal fundus images. We hypothesize that, based on biological connections between the eyes and brain, the volumes of brain tissues can be accurately estimated using retinal fundus images together with clinical information.

## Methods

### Data source

This study was embedded in the KaiLuan Study (KLS), an ongoing multicenter prospective cohort study conducted in Northern China [13]. More than 170,000 individuals aged 18–98 years have been included since the launch in 2006 [14]. Standardized clinical data collection, including demographic questionnaires, physical examinations, and laboratory tests, has been performed biennially across 11 local hospitals since baseline. From 2020 onwards, a subset of participants was voluntarily enrolled in the Multimodality MEdical imaging sTudy bAsed on the KaiLuan Study (META-KLS) for the collection of brain MRI and retinal fundus images [15]. Between 2020 and 2022, these participants underwent baseline examinations including both brain MRI and retinal fundus imaging. Specifically, the age and sex distribution of the participants were similar to that of the overall KLS population [16]. The standardized protocol for clinical information collection, retinal/brain imaging, and automated processing pipeline has been previously published [15].

The inclusion criteria for this study included: 1) subjects who had completed both baseline brain MRI and retinal fundus scans between 2020 and 2022; 2) subjects who did not have a clinical diagnosis of cardiovascular diseases, neurological disorders, neurodegenerative diseases, or neuropsychiatric disorders. The exclusion criteria were as follows: 1) participants with missing identifiable demographic information, such as age and sex; 2) poor imaging quality that was not suitable for analysis. Specifically, for the retinal fundus imaging, images with significant artifacts and blurry structures were excluded; 3) incomplete brain MRI data or retinal imaging data; 4) participants with a diagnosis of cancer history. Finally, a total of 755 adult participants were eligible for analysis.

Brain MRI data were acquired using a GE Healthcare 3.0 T MRI system (General Electric 750W, Milwaukee, Wisconsin, USA) equipped with an 8-channel phased array head coil at Kailuan General Hospital. Total brain volume was calculated based on high-resolution T1-weighted imaging (T1WI) sequences obtained through three-dimensional BRAin Volume imaging (3D-BRAVO). The acquisition parameters for 3D-BRAVO-T1WI have been previously described in detail [15]. After standardized preprocessing steps, including reorientation, segmentation, and spatial normalization, automated analysis was performed using the CAT 12 toolbox in SPM 12 (<http://www.neuro.uni-jena.de>) to calculate gray matter and white matter volumes for all subjects. Total brain volume was computed as the sum of the gray and white matter volumes.

Additionally, to validate the generalizability and clinical applicability of the proposed method, an external validation dataset was obtained from Beijing Friendship Hospital, Capital Medical University. The inclusion and exclusion criteria were consistent with those specified for the META-KLS subjects. There were 142 subjects with both retinal fundus and 3D T1-weighted images, with the same parameters as META-KLS. The age and sex information were also collected. An overview of the study design was illustrated in Fig. 1, and the data processing steps were presented in Fig. 2.

This study has been approved by the Medical Ethics Committee of Kailuan General Hospital (2021002) and Beijing Friendship Hospital, Capital Medical University (2020-P2-155-01). All participants provided written informed consent, and the imaging data utilized in this research were anonymized.

### Details of characteristics

To investigate the impact of individual clinical information on brain volume, the key factors that affect brain health were categorized into 5 types, including demographic information, daily habits, cardiovascular factors,

metabolic factors, and inflammatory factors (Additional file 1: Table S1). The multi-type clinical metadata of subjects, collected retrospectively between 2006 and 2018, were utilized in this study for training (70%) and testing (30%) (Additional file 1: Table S2).

### Functional overview of the brain volume quantification system

The proposed CMCR network is designed to model the intricate co-degenerative relationships between retinal fundus images and brain MRI data. The architecture consists of two main components: the cross-modal degenerative representation network and the clinical information prompt-driven retinal fundus images translation network, as shown in Fig. 3. The cross-modal degenerative representation network is the core component of the CMCR framework. It is responsible for capturing the degenerative features from both retinal fundus images and brain MRI data and establishing a shared degenerative representation space. The network comprises 3 key modules: the brain encoder, the eye encoder, and the cross-modal degenerative fusion module. The prompt-driven brain degeneration inference model leverages the pre-trained cross-modal degenerative representation network to predict brain images from retinal fundus images. By incorporating clinical information as prompts, this model recalibrates the degenerative features derived from retinal images, thereby enhancing the accuracy of brain volume estimation.

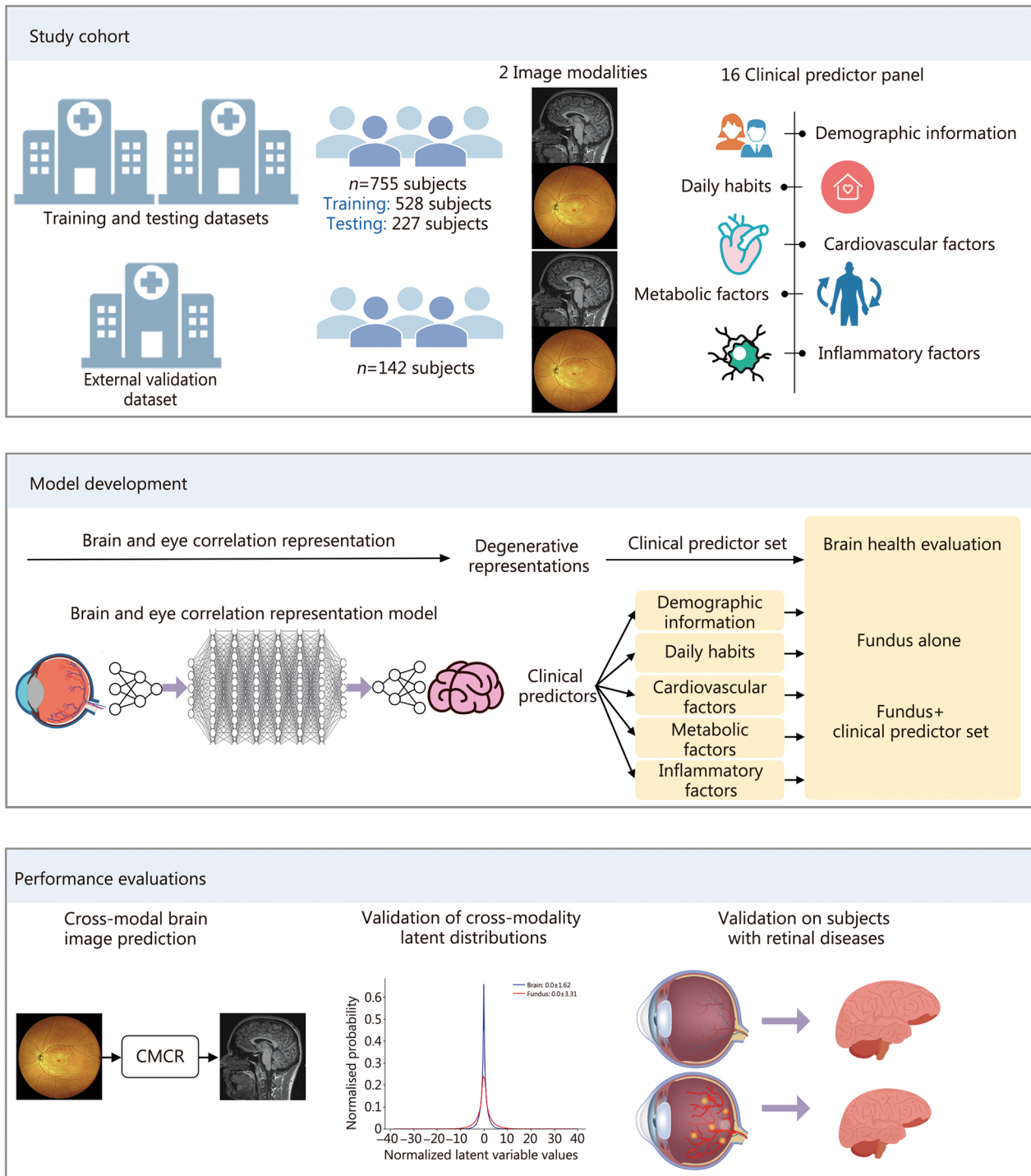
### Comparison methods and performance evaluation metrics

To evaluate the performance of the proposed brain and eye degeneration correlation representation network, several image synthesis methods are utilized for comparison, including multi-channel Variational Autoencoder (mcVAE) [17], Pixel-to-Pixel (Pixel2pixel) [18], transformer-based U-Net (TransUNet) [19], multi-scale transformer network (MT-Net) [20], and residual vision transformer (ResViT) [21]. The implementation details and hyperparameter tuning ranges for each comparative method are summarized in Additional file 1: Table S3.

Three quantitative evaluation metrics are utilized to evaluate the synthesis performance of different models, including: 1) root mean square error (RMSE); 2) peak signal-to-noise ratio (PSNR); and 3) structural similarity index measure (SSIM). The mathematical formulations for RMSE and PSNR are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (X_{ij} - \hat{X}_{ij})^2} \quad (1)$$

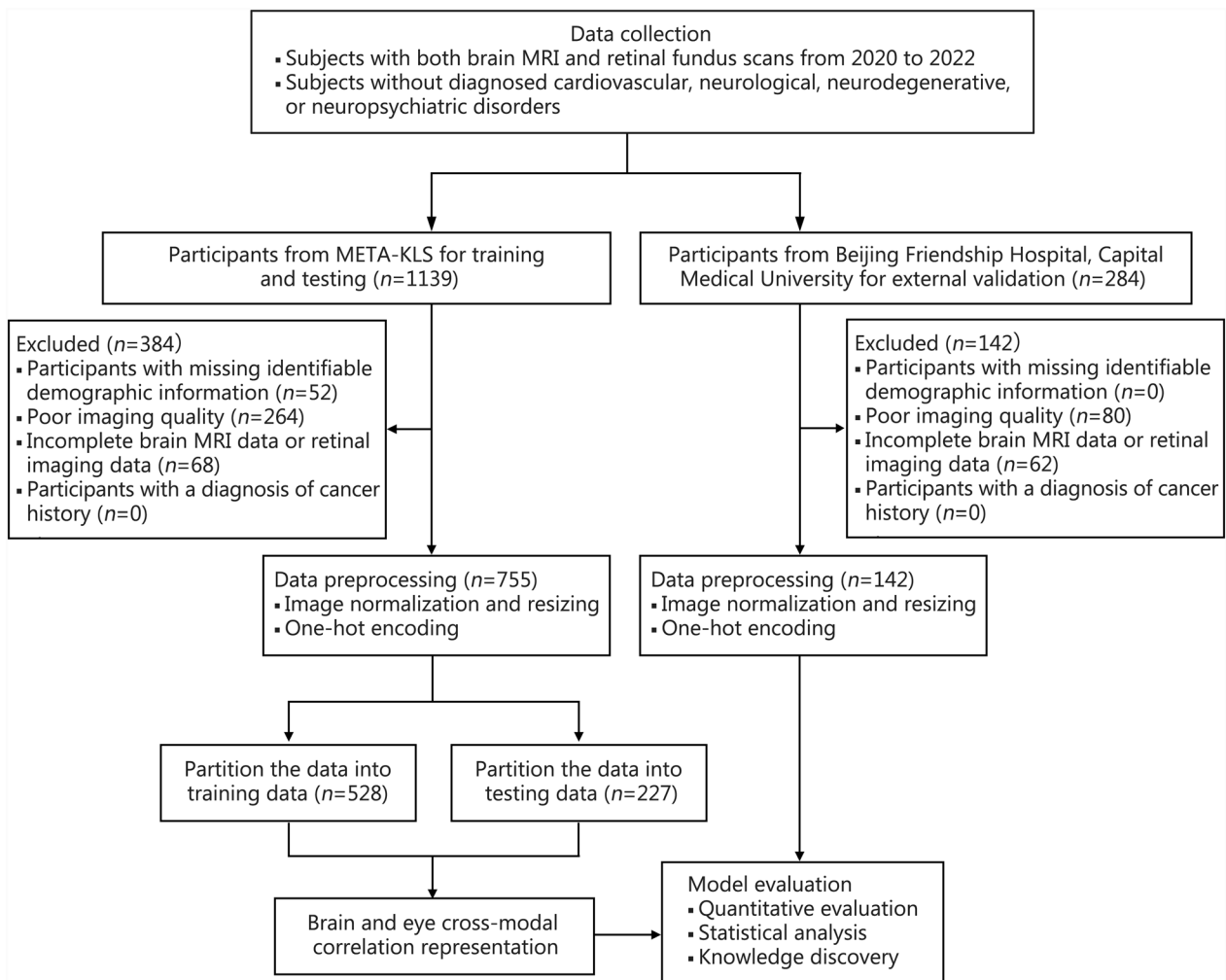
$$\text{PSNR} = 20 \log_{10} \frac{\text{Max}(x)}{\text{RMSE}} \quad (2)$$



**Fig. 1** Study overview. CMCR cross-modal correlation representation

where  $M$  and  $N$  represent the size of the images,  $i$  and  $j$  symbolize the coordinate position,  $x$  and  $\hat{x}$  represent real voxel and predicted voxel. The PSNR

indicates the ratio of the peak signal energy to the average noise energy, and it is inversely proportional to the RMSE. The formulations for SSIM:



**Fig. 2** Data processing steps. MRI magnetic resonance imaging

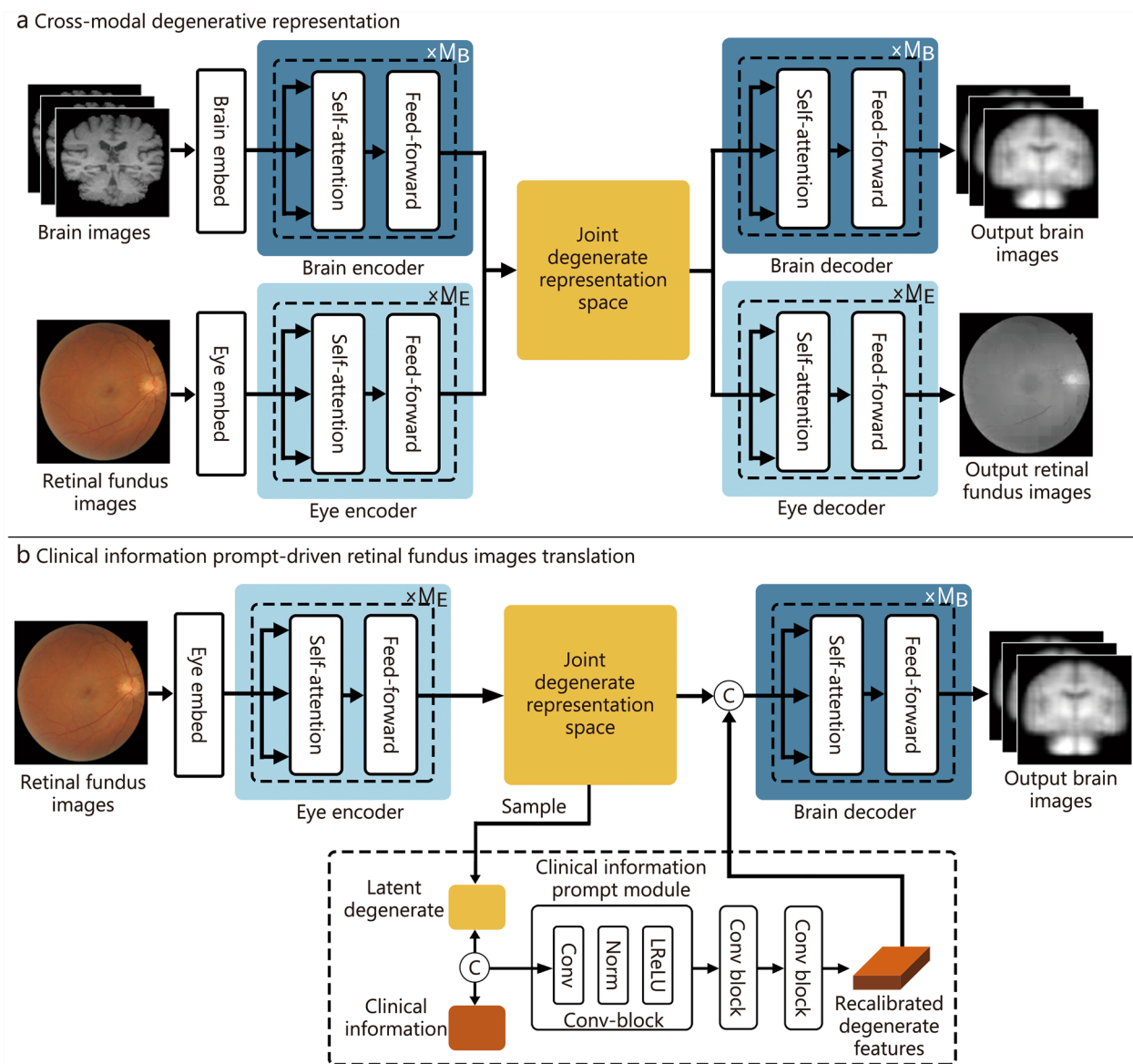
$$SSIM = 1 - \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

where  $\mu_x$  and  $\mu_y$  represent the mean value in the real and predicted images,  $\sigma_x$  and  $\sigma_y$  represent the standard deviation in the real and predicted images, and  $\sigma_{xy}$  represents the covariance.  $c_1$  and  $c_2$  represent constants to avoid system errors caused by a denominator of 0. SSIM is a perception-based computational model that focuses on the fuzzy changes of the structural information in human perception.

**Statistical analysis**

All statistical analyses were performed using Python (v 3.8.3) with the SciPy library (v 1.10.1) for statistical tests and the NumPy library (v 1.22.4) for numerical computations. Continuous variables were presented as mean  $\pm$  standard deviation (SD) for normally distributed data or median (interquartile range, IQR) for non-normal distributions, with normality assessed using Shapiro-Wilk tests. Categorical variables were reported as  $n$  (%).

To comprehensively evaluate the performance of the proposed CMCR framework and assess the statistical



**Fig. 3** The overall architecture of the proposed brain and eye correlation representation framework. The framework consisted of two main components: the cross-modal degenerative representation network and the clinical information prompt-driven retinal fundus images translation network. The output images represented the synthetic brain images generated from the retinal fundus images. These synthetic brain images were used to estimate brain tissue volume, providing a non-invasive and cost-effective approach to assess brain health.  $xM_B$  and  $xM_E$  represent  $M_B$  and  $M_E$  identical modules, respectively. Conv convolutional layer, Norm normalization layer, LReLU leaky rectified linear unit

significance of improvements over comparative methods, multiple statistical analyses were conducted. Quantitative metrics, including RMSE, PSNR, and SSIM, were computed to systematically assess the accuracy and quality of synthetic brain images. For brain volume estimation, both absolute differences (in  $cm^3$ ) and relative errors (percentage difference from ground truth) were reported to provide a detailed evaluation. Paired Student's *t*-test (or Wilcoxon signed-rank tests for non-normal data) were applied to compare the CMCR framework against

each baseline method under identical experimental conditions, including scenarios with and without integrated clinical information. Before conducting paired Student's *t*-test, we confirmed that 1) observations were matched (same subjects under CMCR or other comparing methods and baseline conditions) and 2) paired differences were normally distributed (Shapiro-Wilk test,  $P > 0.05$ ).

Pearson correlation coefficients (*r*) were calculated to quantitatively assess the linear relationship between the predicted and actual brain volumes in the external

**Table 1** Comparative performance analysis of cross-modal brain image synthesis (mean ± SD,  $n = 227$ )

Methods	RMSE	PSNR (dB)	SSIM
mcVAE	232.67 ± 43.67**	28.42 ± 2.38**	0.42 ± 0.08**
Pixel2pixel	210.23 ± 39.48**	29.35 ± 2.06**	0.50 ± 0.06**
TransUNet	198.55 ± 38.61**	30.12 ± 1.92**	0.53 ± 0.05**
MT-Net	172.63 ± 30.28**	32.54 ± 1.27**	0.54 ± 0.05**
ResViT	165.58 ± 27.87**	33.56 ± 1.10*	0.58 ± 0.04**
CMCR	98.23 ± 18.92	35.78 ± 1.01	0.64 ± 0.04

\* $P < 0.05$ , \*\* $P < 0.01$  vs. CMCR. *RMSE* root mean square error, *PSNR* peak signal-to-noise ratio, *SSIM* structural similarity index measure, *mcVAE* multi-channel Variational Autoencoder, *Pixel2pixel* Pixel-to-Pixel, *TransUNet* transformer-based U-Net, *MT-Net* multi-scale transformer network, *ResViT* residual vision transformer, *CMCR* cross-modal correlation representation

validation dataset. Bland-Altman plots were used to visually evaluate the agreement between predicted and actual brain volumes, as well as to identify any potential

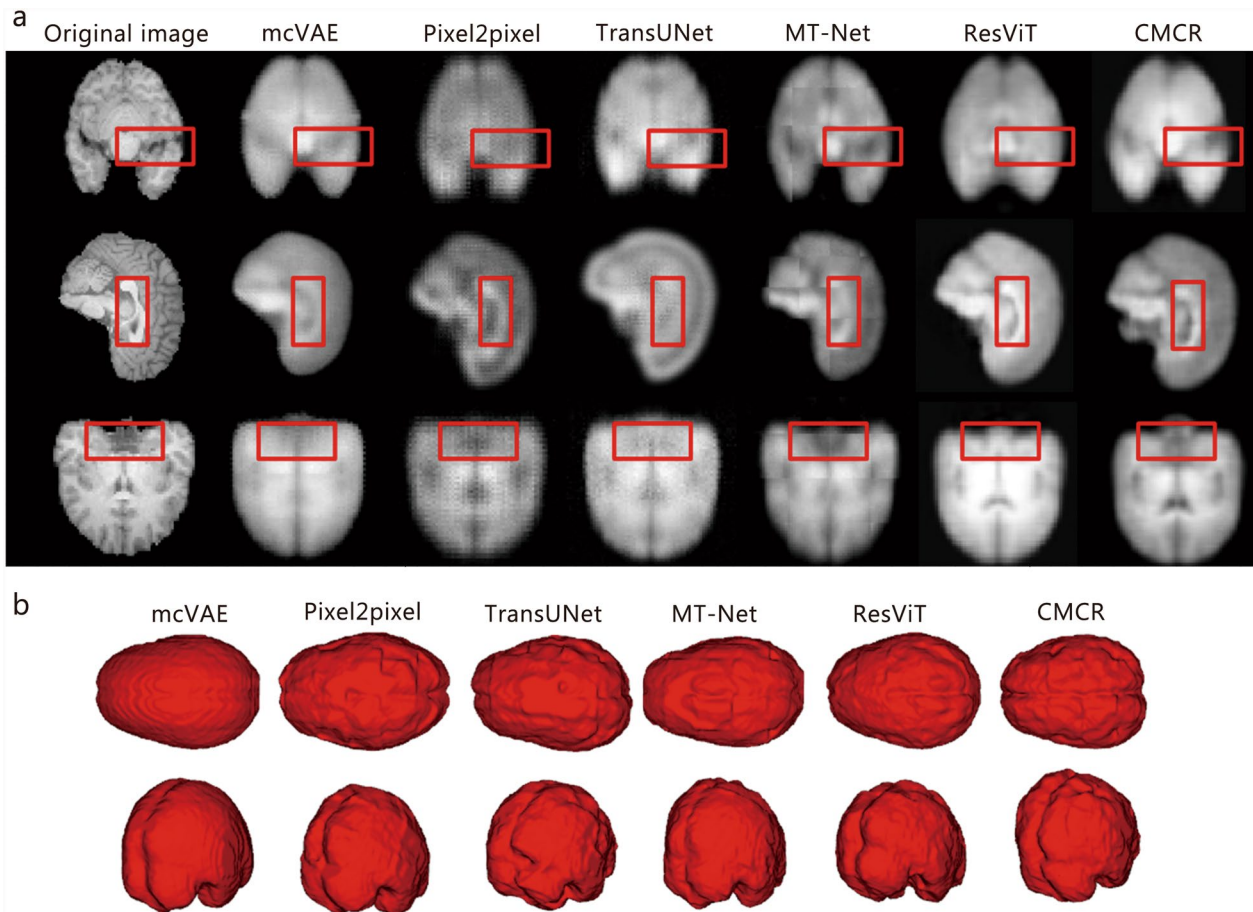
systematic bias or limits of agreement.  $P < 0.05$  (two-tailed) was considered statistically significant.

**Results**

**Cross-modal brain image synthesis**

The quantitative metrics for cross-modal brain image synthesis from retinal fundus images using this method and other comparative methods are shown in Table 1. Specifically, the proposed CMCR framework achieves an RMSE of 98.23, a PSNR of 35.78 dB, and a SSIM of 0.64. CMCR achieves statistically significant improvements across all metrics, demonstrating superior performance in cross-modal brain image generation from retinal fundus images. It consistently outperforms existing methods in accuracy, signal-to-noise ratio, and structural similarity.

The 2D and 3D comparisons of the predicted brain images generated by various methods are presented in Fig. 4. All 6 methods can roughly predict the brain



**Fig. 4** Synthetic brain images in 2D **a** and 3D **b** generated from retinal fundus images using the CMCR method and other comparative methods. The red boxes highlight the regions exhibiting significant improvements achieved by the CMCR framework. 2D two-dimensional, 3D three-dimensional, mcVAE multi-channel Variational Autoencoder, Pixel2pixel Pixel-to-Pixel, TransUNet transformer-based U-Net, MT-Net multi-scale transformer network, ResViT residual vision transformer, CMCR cross-modal correlation representation

**Table 2** Differences between the estimated brain volumes and the ground truths (mean  $\pm$  SD,  $n = 227$ )

Clinical information	Absolute differences (cm <sup>3</sup> )						Relative errors (%)					
	mcVAE	Pixel2pixel	TransUNet	MT-Net	ResViT	CMCR	mcVAE	Pixel2pixel	TransUNet	MT-Net	ResViT	CMCR
N/A	144.26 $\pm$ 26.32**	132.57 $\pm$ 26.47**	119.25 $\pm$ 21.47**	88.26 $\pm$ 13.57**	72.34 $\pm$ 11.49**	61.36 $\pm$ 9.87	10.67 $\pm$ 1.95**	9.80 $\pm$ 1.96**	8.82 $\pm$ 1.59**	6.53 $\pm$ 1.00**	5.35 $\pm$ 0.85**	4.54 $\pm$ 0.73
With demographic information	137.58 $\pm$ 24.37**	123.54 $\pm$ 22.54**	116.51 $\pm$ 17.54**	85.34 $\pm$ 12.34**	70.21 $\pm$ 10.92**	59.57 $\pm$ 9.24	10.17 $\pm$ 1.80**	9.13 $\pm$ 1.67**	8.61 $\pm$ 1.30**	6.31 $\pm$ 0.91**	5.19 $\pm$ 0.81**	4.40 $\pm$ 0.68
With daily habits	129.76 $\pm$ 20.14**	115.49 $\pm$ 19.84**	110.24 $\pm$ 14.23**	78.21 $\pm$ 9.34**	67.26 $\pm$ 9.84*	57.93 $\pm$ 7.62	9.59 $\pm$ 1.49**	8.54 $\pm$ 1.47**	8.15 $\pm$ 1.05**	5.78 $\pm$ 0.69**	4.97 $\pm$ 0.73*	4.28 $\pm$ 0.56
With cardiovascular factors	127.54 $\pm$ 18.67**	113.69 $\pm$ 15.56**	108.34 $\pm$ 13.69**	77.22 $\pm$ 9.89**	65.64 $\pm$ 8.26**	56.98 $\pm$ 7.53	9.43 $\pm$ 1.38**	8.41 $\pm$ 1.15**	8.01 $\pm$ 1.01**	5.71 $\pm$ 0.73**	4.85 $\pm$ 0.61**	4.21 $\pm$ 0.56
With metabolic factors	124.78 $\pm$ 18.02**	106.28 $\pm$ 13.51**	103.25 $\pm$ 12.37**	74.32 $\pm$ 8.74**	62.63 $\pm$ 7.54**	55.75 $\pm$ 6.97	9.23 $\pm$ 1.33**	7.86 $\pm$ 1.00**	7.63 $\pm$ 0.91**	5.49 $\pm$ 0.65**	4.63 $\pm$ 0.56**	4.12 $\pm$ 0.52
With inflammatory factors	120.58 $\pm$ 13.34**	104.41 $\pm$ 12.43**	92.58 $\pm$ 10.67**	72.21 $\pm$ 8.42**	60.29 $\pm$ 8.69*	53.89 $\pm$ 6.51	8.92 $\pm$ 0.99**	7.72 $\pm$ 0.92**	6.84 $\pm$ 0.79**	5.34 $\pm$ 0.62**	4.46 $\pm$ 0.64*	3.98 $\pm$ 0.48

The CMCR method has been compared in performance with other methods at every step using paired Student's *t*-test. \* $P < 0.05$ , \*\* $P < 0.01$  vs. CMCR. The results are reported in absolute differences (cm<sup>3</sup>) and relative errors (%). *N/A* not including any clinical information, *mcVAE* multi-channel Variational Autoencoder, *Pixel2pixel* Pixel-to-Pixel, *TransUNet* transformer-based U-Net, *MT-Net* multi-scale transformer network, *ResViT* residual vision transformer, *CMCR* cross-modal correlation representation

outline. However, more precise contours and local details are significantly compromised by the other 5 comparative methods. The proposed CMCR network not only accurately delineates the cerebral sulcus but also generates rich texture details, which are essential for brain volume calculation. Thus, these quantitative results, in conjunction with superior 2D and 3D visualizations, provide strong evidence that the CMCR framework more accurately captures the latent structural correlations between the fundus and the brain.

#### Validation of cross-modality latent distributions

In addition to quantitatively evaluating the quality of the generated images, the discrepancies in the latent degenerate features of the brain and eye in the joint degenerate space are also evaluated. To validate that the latent distributions of brain and retinal fundus images are effectively encoded by the proposed network, the distributions are illustrated in Additional file 1: Fig. S1. The results suggest that brain and retinal fundus images show consistent distributions in latent degenerate space, thereby validating the feasibility of quantifying brain volumes using retinal fundus images.

The peak differences in the distribution of brain and retinal fundus images in Additional file 1: Fig. S1 arise from the inherent differences in the modality-specific characteristics of the two imaging types. While both brain MRI and retinal fundus images are utilized to assess degenerative changes, they capture distinct aspects of tissue structure and pathology. The brain MRI provides high-resolution 3D structural information, capturing detailed volumetric and morphological alterations in brain tissues. The degenerative features extracted from brain MRI are typically more globally distributed across the brain, leading to a broader and more uniform distribution in the latent space. In contrast, the 2D retinal fundus images primarily capture microvascular and structural changes in the retina. The degenerative features derived from retinal fundus images are often more localized, focusing on specific regions such as the optic disc or retinal vasculature. This localized nature can result in a more concentrated distribution in the latent space. These differences in the spatial and structural characteristics of the two modalities contribute to the observed peak differences in their latent space distributions. However, despite these differences, the shared degenerative representation space established by the CMCR framework effectively captures the underlying degenerative relationships between the fundus and brain, as evidenced by the consistent alignment of their distributions.

#### Cross-modal brain volume quantification

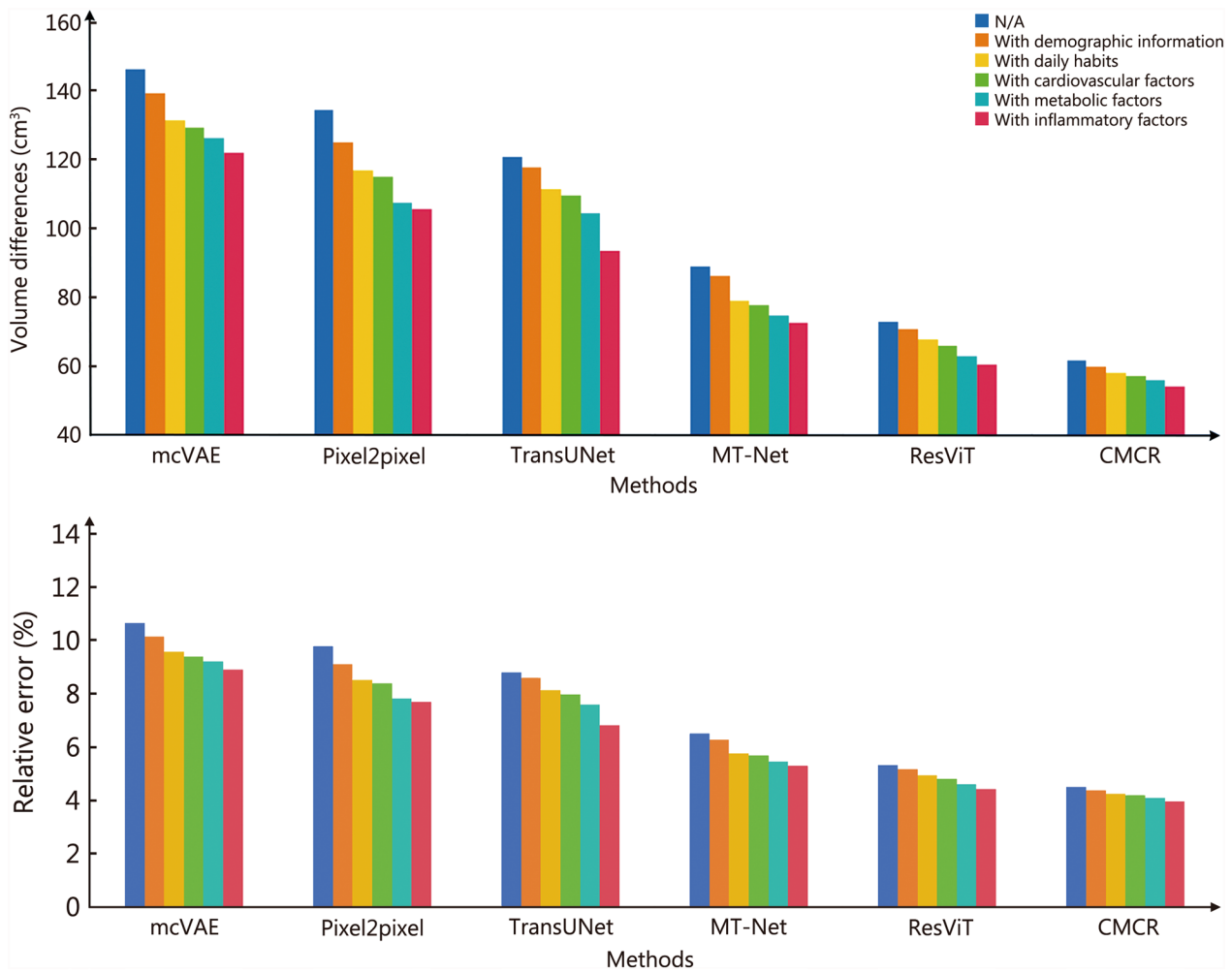
The brain volumes derived from the predicted brain images are calculated and utilized as an important metric to characterize brain health status. In the experiments, the differences between the estimated brain volumes and the ground truths are shown in Table 2. The average difference between estimated brain volumes derived from generated brain images and the actual brain volumes is  $61.36 \text{ cm}^3$ , with a relative error of 4.54%. Compared with other methods, our proposed method exhibits the smallest error in predicting brain volumes. The quantitative evaluation results are consistent with the visualization results in Fig. 4.

To further investigate the contributions of clinical information to the proposed model, 5 categories of clinical information are incrementally encoded and integrated with the degenerate latent code. It can be observed that as more categories of clinical information are integrated, the estimated brain volumes are closer to the ground truths. With the incremental incorporation of age and sex, daily habits, cardiovascular factors, metabolic factors, and inflammatory factors, the differences in estimated brain volume were reduced by 1.79, 3.43, 4.38, 5.61, and  $7.47 \text{ cm}^3$ , respectively. When all of the clinical information was encoded, the difference was decreased to  $53.89 \text{ cm}^3$ , with a relative error of 3.98%. Additionally, our method exhibits significantly superior performance over other methods.

The bar charts in Fig. 5 illustrate the contributions of various types of clinical information in predicting brain volumes using retinal fundus images. It is evident that age and sex, and daily habits exert a more significant influence compared to other clinical factors. More importantly, the predictive results have achieved a considerably high accuracy without the inclusion of invasive examination data or less commonly used parameters. The Bland-Altman plots presented in Additional file 1: Fig. S2 illustrate the differences between estimated brain volumes and ground truths, with and without the incorporation of clinical information.

#### Validation on subjects with retinal diseases

To evaluate the generalizability of the proposed CMCR across diverse clinical conditions, brain volume estimation results for subjects diagnosed with common retinal diseases were presented in Table 3 and Additional file 1: Fig. S3. Specifically, the retinal diseases diagnosed by ophthalmologists included age-related macular degeneration, fundus changes in high myopia, diabetic retinopathy, vitreoretinal interface disorders, and retinal



**Fig. 5** Bar charts illustrating the contributions of various types of clinical information in predicting brain volumes using retinal fundus images. N/A not including any clinical information, mcVAE multi-channel Variational Autoencoder, Pixel2pixel Pixel-to-Pixel, TransUNet transformer-based U-Net, MT-Net multi-scale transformer network, ResViT residual vision transformer, CMCR cross-modal correlation representation

**Table 3** Quantitative evaluation results obtained from subjects with and without retinal diseases (cm<sup>3</sup>, mean ± SD, n = 227)

Subjects condition	mcVAE	Pixel2pixel	TransUNet	MT-Net	ResViT	CMCR
Normal fundus	100.62 ± 12.67**	93.45 ± 14.36**	88.06 ± 9.67**	69.29 ± 7.98**	58.49 ± 6.24*	48.96 ± 5.69
Abnormal fundus	140.38 ± 14.34**	114.56 ± 15.64**	100.68 ± 11.21**	82.20 ± 10.26**	69.92 ± 6.69*	60.21 ± 7.36

The CMCR method has been compared in performance with other methods using paired Student’s *t*-test. \**P* < 0.05, \*\**P* < 0.01 vs. CMCR. mcVAE multi-channel Variational Autoencoder, Pixel2pixel Pixel-to-Pixel, TransUNet transformer-based U-Net, MT-Net multi-scale transformer network, ResViT residual vision transformer, CMCR cross-modal correlation representation

microvascular abnormalities. For subjects with normal fundus images, the CMCR method achieves the lowest brain volume estimation error (48.96 cm<sup>3</sup>), outperforming all comparative methods. For subjects with abnormal fundus images, our method also achieves the lowest error (60.21 cm<sup>3</sup>). Subjects with retinal

diseases exhibit slightly larger brain volume differences compared to those without retinal diseases. Despite the increased differences, the CMCR framework still demonstrates superior performance, highlighting its robustness in handling image variability induced by retinal pathology. The findings provide cutting-edge

**Table 4** Quantitative evaluation results obtained from the external validation set ( $n = 142$ )

Methods	Brain volume differences (cm <sup>3</sup> , mean ± SD)	Pearson's $r$	P-value
mcVAE			
N/A	183.28 ± 28.63**	0.6299	0.007
With demographic information	175.24 ± 25.64**	0.6598	0.007
Pixel2pixel			
N/A	176.54 ± 24.36**	0.6785	0.006
With demographic information	162.76 ± 22.94**	0.6951	0.006
TransUNet			
N/A	146.68 ± 19.34**	0.7084	0.006
With demographic information	139.62 ± 17.37**	0.7121	0.008
MT-Net			
N/A	106.74 ± 15.97**	0.7326	0.008
With demographic information	95.47 ± 14.09**	0.7569	0.009
ResViT			
N/A	99.32 ± 12.67**	0.7463	0.008
With demographic information	90.91 ± 11.93**	0.7782	0.008
CMCR			
N/A	77.14 ± 9.82	0.8274	–
With demographic information	71.24 ± 8.20	0.8451	–

The performance of our proposed method is compared against that of other methods using paired Student's  $t$ -test, both with and without the inclusion of age and sex. \* $P < 0.01$  vs. CMCR. "-" indicates that statistical comparison is not applicable because the CMCR model serves as the reference method in all paired  $t$ -tests. N/A not including demographic information, mcVAE multi-channel Variational Autoencoder, Pixel2pixel Pixel-to-Pixel, TransUNet transformer-based U-Net, MT-Net multi-scale transformer network, ResViT residual vision transformer, CMCR cross-modal correlation representation

evidence demonstrating the model's robust generalizability across a wide spectrum of clinical conditions.

### External validation

External validation experiments are conducted using an external dataset collected from Beijing Friendship Hospital, Capital Medical University. The dataset includes 142 subjects with matched retinal fundus and brain T1 images. Only age and sex information are available for analysis.

The quantitative evaluation results on the external dataset are presented in Table 4. Due to variations in the quality of retinal fundus images from different centers, the discrepancy between estimated brain volume and the actual brain volume has increased (Additional file 1: Fig. S4). Nevertheless, the proposed CMCR still exhibits the best performance and generalizability in estimating brain volumes.

To evaluate the stability of the proposed method, Pearson correlation between the predicted and actual brain volumes, as well as a paired Student's  $t$ -test between our method and other methods, is conducted on the external dataset. It is noteworthy that the proposed CMCR exhibits the highest Pearson's  $r$  values among all methods, indicating robust correlations between predicted and actual brain volumes. In summary, all analyses suggest that the proposed CMCR holds promise as a reliable and

cost-effective clinical tool for assessing brain volumes through retinal fundus images.

### Time complexity analysis

The time complexity is analyzed to evaluate the computational efficiency and scalability of our CMCR framework. The framework consists of several key components, each contributing to the overall time complexity.

#### Brain encoder and eye encoder

Both the brain encoder and eye encoder utilize a Vision Transformer (ViT) architecture. The time complexity is primarily determined by the number of patches and the dimensionality of the embeddings. Each transformer layer involves self-attention and feed-forward operations. The time complexities for self-attention are  $O((N^B)^2 \cdot D)$  and  $O((N^E)^2 \cdot D)$ , respectively, where  $D$  is the dimensionality of the embeddings. For  $M_B$  and  $M_E$  transformer layers in brain and eye encoders, the total time complexities are  $O(M_B \cdot ((N^B)^2 \cdot D + N^B \cdot D^2))$  and  $O(M_E \cdot ((N^E)^2 \cdot D + N^E \cdot D^2))$ , respectively.

#### Cross-modal degenerative fusion module

The cross-modal degenerative fusion module integrates the degenerative features from the brain encoder and eye encoder using a cross-modal attention mechanism.

The cross-modal attention mechanism involves computing attention scores between the brain and eye features. The time complexity for cross-modal attention is  $O(N^B \cdot N^E \cdot D)$ , where  $N^B$  and  $N^E$  are the number of patches for the brain and eye images, respectively.

#### **Brain decoder and eye decoder**

The brain decoder and eye decoder employ transformer-based modules to map high-level degenerative representations to lower semantic representations and perform low-level reconstruction. Similar to the encoders, the time complexities for the transformer layers in brain and eye decoders are  $O(M_B \cdot ((N^B)^2 \cdot D + N^B \cdot D^2))$  and  $O(M_E \cdot ((N^E)^2 \cdot D + N^E \cdot D^2))$ , respectively. The reconstruction process involves upsampling and convolutional operations, with the time complexities of  $O(N^B \cdot D)$  and  $O(N^E \cdot D)$ , respectively. The overall time complexity of our CMCR framework are the sum of the time complexities of its components. The overall time complexity can be approximated as:

$$O(M_B \cdot ((N^B)^2 \cdot D + N^B \cdot D^2) + M_E \cdot ((N^E)^2 \cdot D + N^E \cdot D^2) + N^B \cdot N^E \cdot D + N^B \cdot D + N^E \cdot D)$$

where  $M_B$  and  $M_E$  is the number of transformer layers in the brain and eye encoders/decoders, respectively;  $N^B$  and  $N^E$  is the number of patches for the brain and eye images, respectively, and  $D$  is the dimensionality of the embeddings.

In addition to demonstrating robust performance across diverse clinical settings, our time complexity analysis provides important insights into the computational feasibility of the CMCR framework. By analytically decomposing the time complexity of each core component, specifically, the ViT-based brain and eye encoders/decoders, as well as the cross-modal attention fusion module, we have validated that the model achieves a strong balance between representational capacity and computational efficiency.

This significance is particularly pronounced given the intended application of our model in large-scale population health screening and resource-constrained clinical settings. The use of transformer-based modules, while inherently powerful, is often associated with substantial computational costs. Our detailed complexity evaluation confirms that the CMCR framework remains tractable for practical deployment, facilitating its integration into real-world scenarios where Graphics Processing Unit (GPU) resources may be limited or batch processing is required.

Furthermore, the modular analysis of computational burden provides a roadmap for further optimization and scalability. Future research could explore lightweight

model variants, such as hybrid CNN-transformer structures or sparse attention mechanisms, to further reduce latency while maintaining performance.

Collectively, these findings highlight the potential of CMCR not only as a clinically accurate and biologically interpretable system but also as a computationally efficient solution for accessible and cost-effective brain health evaluation.

#### **Discussion**

Brain health is an evolving major public health concern that has attracted growing interest from both the scientific community and broader society. Brain volume measurement for assessing the integrity of the brain's macrostructure is a critical approach in the regular screening of brain health. In this study, we developed a cross-modal deep learning framework integrating retinal images and brain MR images to learn a combined representation of eye-brain structural correlations. This framework demonstrated that brain MR images can

be synthesized across modalities from retinal fundus images, and that retinal fundus images, along with individual clinical metadata, possess significant potential for accurately estimating brain tissue volume. The study presents an innovative, accurate, and cost-effective approach for characterizing brain health status through readily accessible retinal fundus images.

The design and hypothesis of our study are determined based on the well-established physiological [22], pathological [23], and genetic [12] connections between the eyes and brain. Therefore, we proposed a cross-modal eye-brain degenerative correlation representation model to effectively and quantitatively evaluate the degenerative status of both the eyes and brain. This strategy fully capitalizes on the pathophysiological associations between the eyes and brain while enhancing the interpretability of the results. Specifically, the volume of brain tissue can be accurately estimated based on synthesized brain MR images generated from retinal imaging data. In terms of algorithms, our framework addresses the limitations of previous work by fully leveraging the high-dimensional information embedded within retinal and brain imaging data, rather than focusing exclusively on quantitative features or genetic phenotypes. Specifically, the CMCR framework introduces several innovative aspects compared to existing methods that are widely used for multi-type data integration.

- (1) Unlike existing methods that handle separate specific imaging data as independent inputs, our framework explicitly models the degenerative inter-relationship between retinal and brain imaging modalities. This is accomplished through a Variational Spatial-Transformer Auto-Encoder (VTAE), which captures high-dimensional degenerative features from both retinal fundus and brain images, enabling a more comprehensive and interpretable representation of brain health.
- (2) We incorporate clinical information as prompts to recalibrate the degenerative features extracted from retinal images. This approach dynamically adjusts the latent representations based on diverse categories of personal information, significantly enhancing the accuracy of brain volume estimation and providing a more individualized assessment of brain health.
- (3) Multi-modal masked modeling is employed in the training of the proposed cross-modal degenerative representation network, where patches from both retinal and brain images are randomly masked. This technique drives the model to learn robust cross-modal representations by reconstructing the masked regions, enhancing its capability to capture latent degenerative relationships even with incomplete or noisy data.
- (4) Our framework exhibits robust generalization performance across diverse datasets, as demonstrated by the outcomes of rigorous internal and external validation.

Our systematic work presents numerous advantages over existing research on eye-brain joint analysis, providing a significant complement to the exploration of eye-brain correlation. Firstly, we innovatively utilized original retinal fundus images for model training and fully exploited the high-dimensional information inherent in retinal imaging and the interrelationship between retinal and brain structures. This tailored model extends beyond the conventional focus on quantitative features or genetic phenotypes, representing a significant advancement in current eye-brain joint analysis. Secondly, the research represents a substantial advancement in the evaluation of brain health across a broader demographic. Although brain tissue atrophy is widely recognized as an early indicator of neurological disorders and neurodegenerative diseases [24–26], MRI-based measurements exhibit limited generalizability in resource-limited settings, such as rural areas and aerospace exploration, where access to advanced neuroimaging examinations, real-time radiological interpretation, and quantitative analysis remains restricted. Therefore, the study introduces an innovative approach to assess the status of

brain health without the necessity of neuroimaging data acquisition, thus facilitating the regular screening of brain health in the general population while minimizing medical financial burdens. The estimated smaller brain volume may serve as an early alert for further medical consultation, facilitating timely referrals for additional examinations and promoting the early diagnosis and treatment of neurological diseases. Thirdly, we also conducted a comprehensive analysis of the model's performance in subjects with retinal diseases. The robustness of prediction results demonstrated the reliable generalizability of the proposed model across diverse clinical conditions, thereby enhancing its clinical applicability and proficiency in addressing various common challenges encountered in clinical settings. Lastly, the participants of the KLS were recruited across 11 hospitals in Tangshan city, located at the central area of the Bohai Sea Gulf region [15]. The findings can be effectively generalized to diverse populations within the Western Pacific regions.

There were two primary considerations for incorporating multi-category clinical information to refine the prediction model. Firstly, a substantial body of prior research has demonstrated that personal metadata, including physical examination results [27–29], lifestyle factors [30–32], cardiovascular factors [33–35], metabolic indicators [36–38], and inflammatory markers [39], are significantly correlated with brain health. Integrating this medical knowledge into the model is not only a crucial enhancement for interdisciplinary research in medicine and engineering but also an essential prerequisite for the application of the model in clinical settings. Secondly, diverse categories of clinical data were incrementally encoded into the proposed model, progressing from the most readily available to less commonly used factors. Specifically, basic demographic information such as chronological age and sex was prioritized for inclusion. Subsequently, lifestyle factors such as smoking and alcohol consumption can be easily obtained through questionnaire interviews. Cardiovascular factors, such as blood pressure, lipid profiles, and fasting blood glucose levels, are assessed through invasive or non-invasive methods during physical examinations by well-trained clinicians. Finally, metabolic and inflammatory markers, including creatinine, uric acid, and neutrophils, are less frequently acquired in physical examinations or in population-based cohorts and were therefore encoded in the final step. The experimental results demonstrated that the sequential integration of subject metadata progressively improved prediction accuracy and achieved more accurate estimates of brain volume. The outcome validated the significantly positive contribution of multi-category clinical information to our model and validated the research hypothesis that the volumes of brain tissues can be accurately estimated using retinal fundus

images together with clinical information. Therefore, the integration of diverse categories of clinical information demonstrates substantial value in enabling more precise assessment of individual brain health.

Specifically, the availability and accessibility of clinical information need to be carefully considered for model interpretation in clinical settings. As illustrated in Table 2, brain tissue volume could be estimated with considerable accuracy using age, sex, and lifestyle information, in the absence of data from invasive assessments or less commonly examined parameters. This finding significantly enhances the practicality of implementing our proposed system into clinical settings for routine brain health assessments.

For the validation set, only age and sex are available, while the other clinical features used in training are missing. Theoretically, incorporating the full set of clinical data could achieve better performance. However, considering the practicality of future clinical applications, the clinical features that can be included in practical applications may be limited. Moreover, it was significantly difficult to obtain an optimal validation dataset with the full set of clinical data. Since age and sex informations were readily available, we strategically tried to incorporate only the information in the validation analysis. The proposed model demonstrated reliable high accuracy within the validation dataset, thereby laying the groundwork for future practical implementation.

There are several limitations in our study that we aim to address in future research. First, given the cross-sectional design of this study, our research establishes a shared degenerative representation between retinal fundus images and brain images at the baseline level. Future research should further investigate the predictive capacity of retinal images for the progression of brain atrophy using longitudinal eye-brain imaging data. Such longitudinal validation could provide critical insights into the temporal relationships between retinal and brain degeneration, potentially facilitating the identification of early markers of neurodegeneration and enabling personalized diagnosis and intervention. Second, while brain macrostructural volume is a crucial neurobiomarker of brain health status [16, 29, 40], it cannot definitively indicate the incidence of any specific neurological diseases. The estimated brain volume results may serve as an alert signal rather than a definitive diagnostic criterion. Moreover, future research is warranted to extend the investigations into distinct neuroanatomical regions, particularly those with significant pathological implications, such as the hippocampus associated with AD. Third, although the training and testing datasets utilized in this study exhibit differences in demographic information such as age and sex, these clinical variables have been adjusted during the

model development. External validation was also conducted to examine the impact of dataset discrepancies on model performance. Lastly, it may also be necessary to develop improved algorithms for predicting other subtle neuroimaging markers in asymptomatic individuals, such as the detection of cerebral microbleeds or the quantification of white matter hyperintensity that is closely associated with cerebral small vessel disease [41, 42].

## Conclusions

In this study, we proposed a novel tailored cross-modal eye and brain degenerative correlation representation framework to capture the latent degenerative relationships between the eyes and brain, and estimate the brain tissue volume from retinal fundus images and clinical information. The study provides an innovative, reliable, and cost-effective approach to accurately characterize brain health status through readily accessible retinal fundus images.

## Abbreviations

2D	Two-dimensional
3D	Three-dimensional
3D-BRAVO	Three-dimensional BRAIn VOLUME imaging
AD	Alzheimer's disease
CMCR	Cross-modal correlation representation
GT	Ground truth
KLS	KaiLuan Study
MCI	Mild cognitive impairment
mcVAE	Multi-channel variational autoencoder
META-KLS	Multi-modality MEical imaging sTudy bAsed on the KaiLuan Study
MRI	Magnetic resonance imaging
MT-Net	Multi-scale transformer network
Pixel2pixel	Pixel-to-pixel
PSNR	Peak signal-to-noise ratio
ResViT	Residual vision transformer
RMSE	Root mean square error
SD	Standard deviation
SSIM	Structural similarity index measure
T1WI	T1-weighted imaging
TransUNet	Transformer-based U-Net
VTAE	Variational Spatial-Transformer Auto-Encoder

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40779-025-00630-2>.

**Additional file 1. Table S1** The categories of clinical information.

**Table S2** Comparison of clinical characteristics between training and testing datasets ( $n = 755$ ). **Table S3** The implementation details of the comparison methods. **Fig. S1** Comparison of the latent space variance: distribution of the latent variables obtained using the proposed model on retinal fundus images and brain images. **Fig. S2** Bland-Altman plots for estimated brain volumes and ground truths, without **a** and with **b** the incorporation of clinical information. **Fig. S3** Scatter plots of the estimated brain volume and the actual brain volume for subjects with normal and abnormal fundus conditions. **Fig. S4** Bland-Altman plots for estimated brain volumes and ground truths, without **a** and with **b** the incorporation of clinical information on the external validation set.

### Acknowledgements

The authors would like to thank all the study investigators, technicians, and collaborators for their invaluable contributions of time and expertise, which were crucial to the successful completion of this study. An AI-assisted language polishing tool, Youdao AI web version, was utilized to enhance the readability and grammatical accuracy of this manuscript. The authors have critically reviewed and verified all suggestions provided by the AI tool.

### Authors' contributions

NT, YH, SPG, LXC, XHW, JS, and HL conceived and designed the study. NT, JS, XHW, HL, and ZCW developed the overall research methodology. NT, XHW, and LXC designed the deep learning models and performed the statistical analysis. Figures and tables were prepared by NT, JS, and XHW with assistance from SPG and HL. NT and JS drafted the manuscript. NT, JS, and HL contributed to revisions of the manuscript. YH, SHC, JL, XSL, YTW, SLW, and ZCW contributed to the acquisition and analysis of data. All authors read and approved the final manuscript for publication.

### Funding

This work was supported by the National Natural Science Foundation of China (62522119 and 62372358), the Beijing Natural Science Foundation (7242267), the Beijing Scholars Program ([2015] 160), the Natural Science Basic Research Program of Shaanxi (2023-JC-QN-0719), and the Guangdong Basic and Applied Basic Research Foundation (2022A1515110453).

### Availability of data and materials

Clinical data will be available for other research groups whose proposed use of the data has been approved by an independent review committee identified for this purpose. Requests for data should be directed to the principal investigator, Dr. Zhen-Chang Wang (cjr.wzhch@vip.163.com).

### Declarations

#### Ethics approval and consent to participate

This study has been approved by the Medical Ethics Committee of Kailuan General Hospital (2021002) and Beijing Friendship Hospital, Capital Medical University (2020-P2-155-01). All participants provided written informed consent, and the imaging data utilized in this research were anonymized.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China. <sup>2</sup>Guangzhou Institute of Technology, Xidian University, Guangzhou 510555, China. <sup>3</sup>Department of Radiology, Kailuan General Hospital, Tangshan 063000, Hebei, China. <sup>4</sup>Shenzhen Bay Laboratory, Shenzhen 518132, Guangdong, China. <sup>5</sup>Department of Cardiology, Kailuan General Hospital, Tangshan 063000, Hebei, China. <sup>6</sup>Department of Radiology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing 102218, China. <sup>7</sup>Department of Radiology, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China.

Received: 26 November 2024 Accepted: 10 July 2025

Published online: 06 August 2025

### References

- Optimizing brain health across the life course. WHO position paper. Geneva: World Health Organization, 2022 Licence: CC BY-NC-SA 3.0 IGO. <https://creativecommons.org/licenses/by-nc-sa/3.0/igo/>.
- Lin Y, Chen C, Ma Z, Sabor N, Wei Y, Zhang T, et al. Emulation of brain metabolic activities based on a dynamically controllable optical phantom. *Cyborg Bionic Syst.* 2023;4:0047.
- Wang Y, Pan Y, Li H. What is brain health and why is it important?. *BMJ.* 2020;371:m3683.
- Shi Z, Deng R, Yuan Q, Mao Z, Wang R, Li H, et al. Enzyme commission number prediction and benchmarking with hierarchical dual-core multi-task learning framework. *Research.* 2023;6:0153.
- Tozlu C, Jamison K, Gu Z, Gauthier SA, Kuceyeski A. Estimated connectivity networks outperform observed connectivity networks when classifying people with multiple sclerosis into disability groups. *NeuroImage: Clinical.* 2021;32:102827.
- Zhang H, Zhou QQ, Chen H, Hu XQ, Li WG, Bai Y, et al. The applied principles of EEG analysis methods in neuroscience and clinical neurology. *Mil Med Res.* 2023;10(1):67.
- Ke XY, Hou W, Huang Q, Hou X, Bao XY, Kong WX, et al. Advances in electrical impedance tomography-based brain imaging. *Mil Med Res.* 2022;9(1):10.
- Andrews MJ, Salat DH, Milberg WP, McGlinchey RE, Fortier CB. Poor sleep and decreased cortical thickness in veterans with mild traumatic brain injury and post-traumatic stress disorder. *Mil Med Res.* 2024;11(1):51.
- Bales TR, Lopez MJ, Clark J. Embryology, Eye. 2023. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
- Cornel E, Holt C. Precocious pathfinding: retinal axons can navigate in an axonless brain. *Neuron.* 1992;9(6):1001–11.
- Xie J, Yi Q, Wu Y, Zheng Y, Liu Y, Macerollo A, et al. Deep segmentation of OCTA for evaluation and association of changes of retinal microvasculature with Alzheimer's disease and mild cognitive impairment. *Br J Ophthalmol.* 2024;108(3):432–9.
- Zhao B, Li Y, Fan Z, Wu Z, Shu J, Yang X, et al. Eye-brain connections revealed by multimodal retinal and brain imaging genetics. *Nat Commun.* 2024;15(1):6064.
- Wang C, Yuan Y, Zheng M, Pan A, Wang M, Zhao M, et al. Association of age of onset of hypertension with cardiovascular diseases and mortality. *J Am Coll Cardiol.* 2020;75(23):2921–30.
- Zhao M, Song L, Sun L, Wang M, Wang C, Yao S, et al. Associations of type 2 diabetes onset age with cardiovascular disease and mortality: the Kailuan study. *Diabetes Care.* 2021;44(6):1426–32.
- Sun J, Hui Y, Li J, Zhao X, Chen Q, Li X, et al. Protocol for multi-modality medical imaging study based on Kailuan study (META-KLS): rationale, design and database building. *BMJ Open.* 2023;13(2):e067283.
- Lv H, Sun J, Zhang T, Hui Y, Li J, Zhao X, et al. Associations of serum uric acid variability with neuroimaging metrics and cognitive decline: a population-based cohort study. *BMC Med.* 2024;22(1):256.
- Diaz-Pinto A, Ravikumar N, Attar R, Suinesiaputra A, Zhao Y, Levelt E, et al. Predicting myocardial infarction through retinal scans and minimal personal information. *Nat Mach Intell.* 2022;4(1):55–61.
- Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; p. 5967–76.
- Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: transformers make strong encoders for medical image segmentation. *arXiv: Computer Vision and Pattern Recognition.* 2021.
- Li Y, Zhou T, He K, Zhou Y, Shen D. Multi-scale transformer network with edge-aware pre-training for cross-modality MR image synthesis. *IEEE Trans Med Imaging.* 2023;42(11):3395–407.
- Dalmaz O, Yurt M, Çukur T. ResViT: residual vision transformers for multimodal medical image synthesis. *IEEE Trans Med Imaging.* 2022;41(10):2598–614.
- Ooi TL, He ZJ. Sensory eye dominance: relationship between eye and brain. *Eye Brain.* 2020;12:25–31.
- Baksh BS, Garcia JC, Galor A. Exploring the link between dry eye and migraine: from eye to brain. *Eye Brain.* 2021;13:41–57.
- Rahayel S, Tremblay C, Vo A, Zheng YQ, Lehericy S, Arnulf I, et al. Brain atrophy in prodromal synucleinopathy is shaped by structural connectivity and gene expression. *Brain.* 2022;145(9):3162–78.
- Uchida Y, Nishimaki K, Soldan A, Moghekar A, Albert M, Oishi K. Acceleration of brain atrophy and progression from normal cognition to mild cognitive impairment. *JAMA Netw Open.* 2024;7(10):e2441505.
- Wen J, Fu CHY, Tosun D, Veturi Y, Yang Z, Abdulkadir A, et al. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. *JAMA Psychiat.* 2022;79(5):464–74.

27. Sun J, Wang L, Gao Y, Hui Y, Chen S, Wu S. Discovery of high-risk clinical factors that accelerate brain aging in adults: a population-based machine learning study. *Research*. 2024;7:0500.
28. Hamer M, Batty GD. Association of body mass index and waist-to-hip ratio with brain structure: UK Biobank study. *Neurology*. 2019;92(6):e594–600.
29. Sun J, Zeng N, Hui Y, Li J, Liu W, Zhao X, et al. Association of variability in body size with neuroimaging metrics of brain health: a population-based cohort study. *Lancet Reg Health West Pac*. 2024;44:101015.
30. Tan ZS, Spartano NL, Beiser AS, DeCarli C, Auerbach SH, Vasan RS, et al. Physical activity, brain volume, and dementia risk: the Framingham study. *J Gerontol A Biol Sci Med Sci*. 2017;72(6):789–95.
31. Boller B, Mellah S, Ducharme-Laliberté G, Belleville S. Relationships between years of education, regional grey matter volumes, and working memory-related brain activity in healthy older adults. *Brain Imaging Behav*. 2017;11(2):304–17.
32. Luhar RB, Sawyer KS, Gravitz Z, Ruiz SM, Oscar-Berman M. Brain volumes and neuropsychological performance are related to current smoking and alcoholism history. *Neuropsychiatr Dis Treat*. 2013;9:1767–84.
33. Deng YT, Li YZ, Huang SY, Ou YN, Zhang W, Chen SD, et al. Association of life course adiposity with risk of incident dementia: a prospective cohort study of 322,336 participants. *Mol Psychiatry*. 2022;27(8):3385–95.
34. Chuang SY, Wang PN, Chen LK, Chou KH, Chung CP, Chen CH, et al. Associations of blood pressure and carotid flow velocity with brain volume and cerebral small vessel disease in a community-based population. *Transl Stroke Res*. 2021;12(2):248–58.
35. Ma Y, Yilmaz P, Bos D, Blacker D, Viswanathan A, Ikram MA, et al. Blood pressure variation and subclinical brain disease. *J Am Coll Cardiol*. 2020;75(19):2387–99.
36. Hakala JO, Pahkala K, Juonala M, Salo P, Kahonen M, Hutri-Kahonen N, et al. Repeatedly measured serum creatinine and cognitive performance in midlife: the cardiovascular risk in young Finns study. *Neurology*. 2022;98(22):e2268–81.
37. Rajagopalan P, Refsum H, Hua X, Toga AW, Jack CR Jr, Weiner MW, et al. Alzheimer's disease neuroimaging initiative Mapping creatinine- and cystatin C-related white matter brain deficits in the elderly. *Neurobiol Aging*. 2013;34(4):1221–30.
38. Tang X, Song ZH, Cardoso MA, Zhou JB, Simo R. The relationship between uric acid and brain health from observational studies. *Metab Brain Dis*. 2022;37(6):1989–2003.
39. Jiang L, Cai X, Yao D, Jing J, Mei L, Yang Y, et al. Association of inflammatory markers with cerebral small vessel disease in community-based population. *J Neuroinflammation*. 2022;19(1):106.
40. Lv H, Zeng N, Li M, Sun J, Wu N, Xu M, et al. Association between body mass index and brain health in adults: a 16-year population-based cohort and mendelian randomization study. *Health Data Sci*. 2024;4:0087.
41. Duering M, Biessels GJ, Brodtmann A, Chen C, Cordonnier C, de Leeuw FE, et al. Neuroimaging standards for research into small vessel disease—advances since 2013. *Lancet Neurol*. 2023;22(7):602–18.
42. Wardlaw JM, Smith C, Dichgans M. Small vessel disease: mechanisms and clinical implications. *Lancet Neurol*. 2019;18(7):684–96.