

# SFCNEXT: A SIMPLE FULLY CONVOLUTIONAL NETWORK FOR EFFECTIVE BRAIN AGE ESTIMATION WITH SMALL SAMPLE SIZE

Yu Fu<sup>1,2</sup>, Yanyan Huang<sup>1</sup>, Shunjie Dong<sup>1</sup>, Yalin Wang<sup>3</sup>, Tianbai Yu<sup>1</sup>, Meng Niu<sup>4</sup> and Cheng Zhuo<sup>1,5\*</sup>

<sup>1</sup>Zhejiang University, Hangzhou, China

<sup>2</sup>Binjiang Institute of Zhejiang University, Hangzhou, China

<sup>3</sup>Arizona State University, Tempe, USA

<sup>4</sup>The First Hospital of Lanzhou University, Lanzhou, China

<sup>5</sup>Key Laboratory of Collaborative Sensing and Autonomous Unmanned Systems of Zhejiang Province, Hangzhou, China

## ABSTRACT

Deep neural networks (DNN) have been designed to predict the chronological age of a healthy brain from T1-weighted magnetic resonance images (T1 MRIs), and the predicted brain age could serve as a valuable biomarker for the early detection of development-related or aging-related disorders. Recent DNN models for brain age estimations usually rely too much on large sample sizes and complex network structures for multi-stage feature refinement. However, in clinical application scenarios, researchers usually cannot obtain thousands or tens of thousands of MRIs in each data center for thorough training of these complex models. This paper proposes a simple fully convolutional network (SFCNeXt) for brain age estimation in small-sized cohorts with biased age distributions. The SFCNeXt consists of Single Pathway Encoded ConvNeXt (SPEC) and Hybrid Ranking Loss (HRL), aiming to estimate brain ages in a lightweight way with a sufficient exploration of MRI, age, and ranking features of each batch of subjects. Experimental results demonstrate the superiority and efficiency of our approach.

**Index Terms**— Brain age estimation, Deep neural networks, Magnetic resonance images

## 1. INTRODUCTION

Brain development and aging, accompanied by complex biological and neuroanatomical changes, is an ongoing and lifelong process that is largely not fully understood [1, 2, 3, 4]. To capture the development and aging patterns, a brain age estimation model can be trained using healthy neuroimaging data to predict the brain age as closely as feasible to the actual chronological age. The discrepancy between the anticipated age and the actual chronological age is commonly referred to as the “brain age gap”, which has been associated with a variety of biological and cognitive characteristics [5, 6, 7].

Measuring the difference in brain age of patient groups may help assess the disease heterogeneity and improve disease risk screening [8].

T1-weighted magnetic resonance images (T1 MRIs), possess rich morphological information about the brain, have been frequently used for brain age estimations [5, 6, 9]. Whether a model can obtain a smaller mean absolute error (MAE), a larger Pearson correlation coefficient (PCC), and a larger Spearman’s rank correlation coefficient (SRCC) is crucial for determining its suitability for brain age estimation [6, 7, 9]. Researchers in the deep learning community have built different neural network backbones for brain age estimation with brain T1 MRIs, such as convolutional neural network (CNN)-based [9, 7], ResNet-based [1, 6] and Transformer-based [10]. However, most of these models own complex frameworks (e.g., many branches for information flows) and consume a lot of GPU resources, which may limit their deployment in clinical scenarios. Besides, a good performance of these models is largely depended on huge sample sizes (e.g., thousands or tens of thousands of MRIs), but researchers in a specific team often do not have sufficient MRIs to train, given the considerations of data sharing and privacy issues of multi-site data around the world [11].

To address the above issues, we propose a simple, fully convolutional network (SFCNeXt) for accurate and effective brain age estimation with a small sample size. The SFCNeXt consists of two technological parts: Single Pathway Encoded ConvNeXt (SPEC) and Hybrid Ranking Loss (HRL). The SPEC includes a backbone pathway that consists of a four-stage ConvNeXt, and a branch pathway adopts sex features. After the concatenation of the two pathways, a conformer encoder cascaded with a multilayer perceptron (MLP) is employed for adaptive feature fusion for brain age estimation. The HRL is a combination of mean square error (MSE) loss and fast differentiable ranking loss, which not only fits brain MRI features to brain ages but explores the soft ranking relationships of each batch of brain MRIs.

\*Corresponding author.

**Table 1.** Demographic information of brain age estimation datasets used in this study.

Site	$N_{img}$	Age Range	Age (Mean $\pm$ STD)	Gender (Male/Female)
ATT	13	20-25	22.23 $\pm$ 1.42	12/1
ATV	39	20-30	22.69 $\pm$ 2.20	29/10
CIN	39	20-67	38.69 $\pm$ 13.49	25/14
COI	119	20-72	50.87 $\pm$ 12.78	42/77
HKH	29	28-65	45.41 $\pm$ 9.53	12/17
HRC	49	23-68	41.69 $\pm$ 11.66	13/36
HUH	67	20-66	34.75 $\pm$ 12.97	29/38
KTT	75	19-57	28.89 $\pm$ 9.05	48/27
KUT	157	19-68	36.36 $\pm$ 13.18	92/65
SWA	101	19-55	28.44 $\pm$ 7.86	86/15
UTO	91	24-72	45.31 $\pm$ 14.11	31/60
Overall	779	19-72	37.61 $\pm$ 14.32	419/360

## 2. DATASETS AND PREPROCESSING

We evaluate the proposed SFCNeXt on a healthy cohort [12, 13], which includes T1-weighted MRIs from 11 consortium sites of Japan (see Table 1), with a total of 779 subjects aged 19-72 years. To our knowledge, this is a small-sized cohort relative to sample sizes adopted in mainstream brain age estimation models (e.g., tens of thousands of samples). The histogram matching and intensity normalization are performed for the harmonization of MRIs across datasets. All images are processed via a standard preprocessing pipeline with FSL 6.0 [14], which includes nonlinear registration to standard 2mm MNI space and brain extraction [15, 7]. After preprocessing, each MRI has a spatial size of  $91 \times 109 \times 91$  voxel with a spatial resolution of 2 mm<sup>3</sup>.

## 3. METHODOLOGY

### 3.1. Single Pathway Encoded ConvNeXt (SPEC)

The whole framework of the proposed SFCNeXt is displayed in Fig. 1 (a). The SPEC consists of a backbone pathway and a branch pathway. The backbone pathway is a four-stage ConvNeXt network (i.e., each stage includes several ConvNeXt blocks) for down-sampling and aggregating 3D brain MRI features in a cascade manner. At the end of the backbone pathway, a conformer encoder [16, 17] module is employed for the adaptive encoding these high-level brain MRI features. The branch pathway is a simple multilayer perceptron (MLP) that mainly aims to feed sex features (i.e., male or female) into the model. Although controversial in previous studies [18], we have found positive benefits of adding sex features in this study. After the concatenation of the two pathways, the brain age is obtained using the final MLP. The number of blocks in the four-stage ConvNeXt network is (1, 1, 3, 1), which significantly reduces the computing resources of the original ConvNeXt stage such as (3, 3, 9, 3) [19]. All of the ConvNeXt blocks in each stage of the SPEC use an overlapped working mode for 3D down-sampling (see Fig. 1 (b)), which can significantly mitigate the feature dam-

age problem caused by dividing each 3D brain into multiple non-overlapping 3D patches. Usually, a conformer encoder module (see Fig. 1 (a)) contains  $N$  conformer blocks, each of which contains two feed-forward (FF) modules sandwiching the multi-headed self-attention (MHSA) module and the convolution module (see Fig. 1 (c)). Each conformer block consists of 4 residual connection modules and a layer normalization operator.

### 3.2. Hybrid Ranking Loss (HRL)

Some previous studies have shown the benefits of ranking loss for improving the performance of brain age estimation [5, 7]. For example, [5] adopts Spearman’s rank correlation coefficient (SRCC) of chronological age to improve brain age estimation directly. For two subjects with ages  $y_i$  and  $y_j$ , the age difference loss can be seen as the MSE between the estimated brain age difference  $\hat{y}_i - \hat{y}_j$  and the true age difference  $y_i - y_j$ . Suppose the  $N_b$  denotes the number of paired samples  $(i, j)$  in each batch, we can list the MSE loss  $\mathcal{L}_{MSE}$  and age difference loss  $\mathcal{L}_d$  as follows, respectively:

$$\mathcal{L}_{MSE} = \frac{1}{N_b} \sum_{i=1}^{N_b} (\hat{y}_i - y_i)^2 \quad (1)$$

$$\mathcal{L}_d = \frac{1}{N_b} \sum_{i=1, j=1}^{N_b} ((\hat{y}_i - \hat{y}_j) - (y_i - y_j))^2 \quad (2)$$

where the sign of the  $\mathcal{L}_d$  indicates the positive or negative difference of two ages. SRCC rank correlation can be defined as the PCC between the rank values of two variables:

$$r_r = \frac{\text{cov}(\text{Rank}(\hat{y}), \text{Rank}(y))}{\sigma_{\text{Rank}(\hat{y})} \sigma_{\text{Rank}(y)}} \quad (3)$$

where  $\text{Rank}(y)$  is the rank operator,  $\text{cov}(\text{Rank}(\hat{y}), \text{Rank}(y))$  is the covariance of the two rank variables, and  $\sigma_{\text{Rank}(\hat{y})}$  and  $\sigma_{\text{Rank}(y)}$  are the standard deviations of the two rank variables. So the SRCC ranking loss can be simplified as:

$$\mathcal{L}_r = \sum_{i=1}^{N_b} (\text{Rank}(\hat{y}_i) - \text{Rank}(y_i))^2 \quad (4)$$

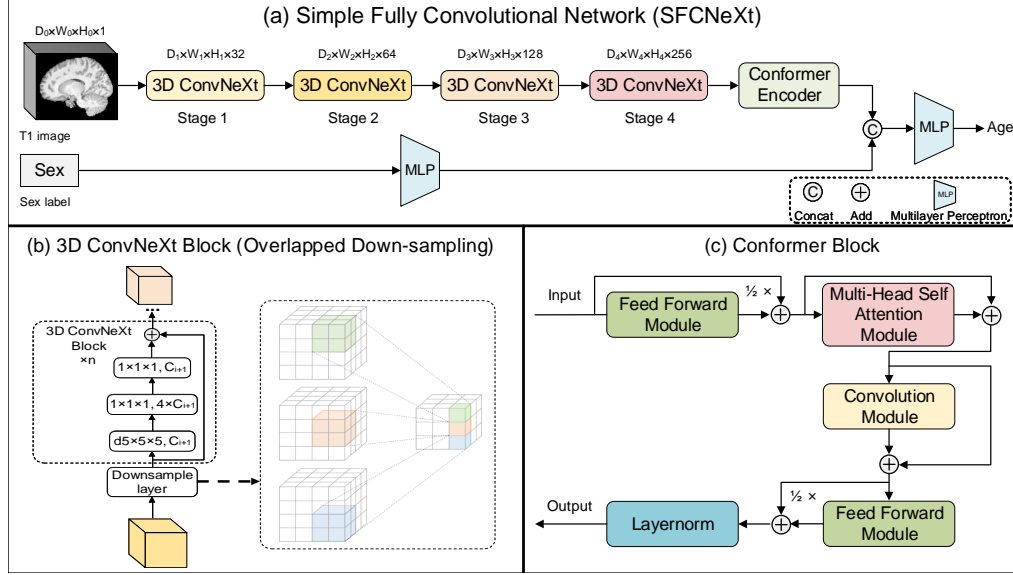
As shown by [20], computing SRCC is usually problematic: its derivatives are null or undefined, preventing gradient back-propagation. A simple method based on comparing pairwise distances between variables can obtain the approximation of SRCCs, but may take  $O(n^2)$  time [21], which is not suitable for our intention in SFCNeXt.

To reduce the computational complexity, this paper feats the ranking as projections onto a permutahedron. Let  $\mathbf{z}, \mathbf{w} \in \mathbb{R}^n$  and consider the linear program as:

$$\text{argmax}_{\mu \in \mathcal{P}(\mathbf{w})} \langle \mu, \mathbf{z} \rangle \quad (5)$$

then we can represent the  $r(\theta)$  by setting  $(\mathbf{z}, \mathbf{w}) = (-\theta, \rho)$ . Here the  $r(\theta)$  means rank values of  $\theta$  and the  $\mathcal{P}(\mathbf{w})$  denotes the permutahedron of  $\mathbf{w}$ .  $\theta$  and  $\rho$  denote permutations with linear program. Therefore, the  $\Psi$ -regularized soft ranking loss can be written as:

$$r_{\varepsilon\Psi}(\theta) := P_{\varepsilon\Psi}(-\theta, \rho) = P_{\Psi}(-\theta/\varepsilon, \rho) \quad (6)$$



**Fig. 1.** (a) The whole framework of the proposed SFCNeXt; (b) Overlapped 3D down-sampling; (c) Conformer block.

where  $0 < \varepsilon < \infty$ , and  $O(n \log n)$  time can be easily achieved, which facilitates our training. Finally, the total loss function  $\mathcal{L}$  is:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_1 \mathcal{L}_d + \lambda_2 \mathcal{L}_{r_\varepsilon \Psi}(\theta) \quad (7)$$

where the  $\lambda_1$  and  $\lambda_2$  are regularization parameters.

## 4. EXPERIMENTS AND RESULTS

To ensure the stability of the results, the healthy cohort is randomly partitioned into three subsets: the training set (80% of MRIs), the validation set (10% of MRIs), and the test set (10% of MRIs). The training-validation-test split is performed for total 10 times to conduct the 10-fold cross-validation procedure, which is commonly used in previous brain age estimation studies [1, 9]. All experiments are conducted using a single NVIDIA GeForce RTX 3090 GPU with PyTorch library, and the python version is 3.8.13. A group of recent 3D lightweight state-of-the-art (SOTA) models for brain age estimation, such as SFCN [9], 3D ResNet [9, 22] and their corresponding variants (e.g., with different layers, blocks, and optimizers), have been included in this study for comparison. For a fair comparison, all competing models are reproduced according to their released codes, and we further optimize them for better performance in our dataset. Sex labels are also used as covariates of these competing models for the improvement of performance.

### 4.1. Parameter evaluations of the SFCNeXt

First, we evaluate the performance of the proposed SFCNeXt using different loss functions (MAE or MSE), batch sizes (from 4 to 20), initial learning rates (ILR) (from  $1 \times 10^{-4}$

to  $1 \times 10^{-2}$ ), optimizers (Adam, AdamW or Adamax), attention heads (from 1 to 4) and conformer blocks (from 1 to 3). SFCNeXt (a) to SFCNeXt (p) in Table 2 depict SFCNeXt variants with different hyper-parameters. Firstly, in the top part of the Table 2, we find that adopting MSE as loss and 8 as batch size can achieve the best MAE, PCC, and SRCC (SFCNeXt (e)). Then, in the middle part of the Table 2, we note that combining the  $1 \times 10^{-4}$  as ILR and Adamax (SFCNeXt (m)) can reduce the MAE to  $3.683 \pm 0.009$ , increase the PCC to  $0.859 \pm 0.001$  and SRCC to  $0.817 \pm 0.001$ , when with 2 attention heads and 3 conformer blocks in the conformer encoder module. In the bottom part of Table 2, we vary the number of attention heads and Conformer blocks and demonstrate that 2 attention heads and 3 conformer blocks are the most suitable combinations for SFCNeXt. Here the SFCNeXt (m) and SFCNeXt (p) refer to the same model. Therefore, the final version of SFCNeXt adopts MSE as a loss function, 8 as batch size,  $1 \times 10^{-3}$  as the initial learning rate, Adamax as an optimizer, 2 attention heads and 3 conformer blocks.

### 4.2. Comparison of SFCNeXt vs. other lightweight SOTA models

Table 3 details the quantitative brain age estimation results for the test set by the SFCNeXt and its competing models (i.e. 3D ResNet, SFCN, and their corresponding variants). It is easy to note that the SFCNeXt outperforms all variants of 3D ResNet and SFCN. Fig. 2 depicts the scatter diagrams of the estimated brain ages versus the chronological brain ages based on the 3D ResNet18 (the best ResNet model), SFCN\_Adamax (the best SFCN model), and SFCNeXt. The SFCN\_Adam is found as the best model other than SFCNeXt. Compared with SFCN\_Adam, the SFCNeXt significantly reduces MAE by 9.7%, PCC by 1.5%, and SRCC by 2%. Also, the SFC-

NeXt network consumes 6971MB of GPU memory, which is comparable to other lightweight competing models.

**Table 2.** Performance of brain age estimation using SFCNeXt with different hyper-parameters (bold denotes better results).

Model	Loss	Batch Size	MAE (Mean±STD)	PCC (Mean±STD)	SRCC (Mean±STD)
SFCNeXt (a)	MAE	4	4.396±0.322	0.802±0.005	0.790±0.012
SFCNeXt (b)	MAE	8	4.363±0.066	0.832±0.008	0.811±0.004
SFCNeXt (c)	MAE	20	8.369±0.063	0.293±0.046	0.275±0.051
SFCNeXt (d)	MSE	4	4.514±0.116	0.814±0.011	0.772±0.007
<b>SFCNeXt (e)</b>	<b>MSE</b>	<b>8</b>	<b>4.013±0.041</b>	<b>0.852±0.014</b>	<b>0.797±0.004</b>
SFCNeXt (f)	MSE	20	5.489±0.131	0.774±0.017	0.781±0.003

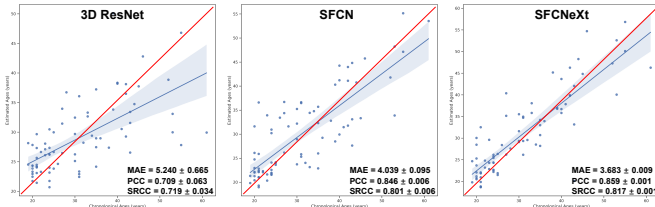
  

Model	ILR	Optimizer	MAE (Mean±STD)	PCC (Mean±STD)	SRCC (Mean±STD)
SFCNeXt (g)	$1 \times 10^{-3}$	Adam	4.013±0.041	0.852±0.014	0.787±0.004
SFCNeXt (h)	$2 \times 10^{-3}$	Adam	4.467±0.061	0.810±0.009	0.772±0.008
SFCNeXt (i)	$4 \times 10^{-3}$	Adam	6.497±1.986	0.671±0.184	0.658±0.145
SFCNeXt (j)	$1 \times 10^{-3}$	AdamW	4.398±0.322	0.802±0.005	0.790±0.012
SFCNeXt (k)	$2 \times 10^{-3}$	AdamW	4.426±0.143	0.802±0.017	0.771±0.026
SFCNeXt (l)	$4 \times 10^{-3}$	AdamW	4.369±0.186	0.824±0.013	0.775±0.041
<b>SFCNeXt (m)</b>	<b><math>1 \times 10^{-3}</math></b>	<b>Adamax</b>	<b>3.683±0.009</b>	<b>0.859±0.001</b>	<b>0.817±0.001</b>

Model	Attention Head	Conformer Block	MAE (Mean±STD)	PCC (Mean±STD)	SRCC (Mean±STD)
SFCNeXt (n)	2	1	3.919±0.075	0.854±0.006	0.814±0.002
SFCNeXt (o)	2	2	4.073±0.129	0.850±0.006	0.802±0.003
<b>SFCNeXt (p)</b>	<b>2</b>	<b>3</b>	<b>3.683±0.009</b>	<b>0.859±0.001</b>	<b>0.817±0.001</b>
SFCNeXt (q)	4	3	3.951±0.017	0.852±0.006	0.808±0.004

MAE Mean Absolute Error; MSE Mean Square Error; ILR: Initial Learning Rate



**Fig. 2.** The scatter diagrams of estimated brain ages by 3D ResNet18, SFCN.Adamax, and the SFCNeXt. The red lines in sub-graphs indicate the ideal estimation of  $y=x$  (i.e., the estimated brain age equals the chronological age), while the blue lines with CIs denote the actual estimation performance of each model. The smaller the angle between the two lines, the better the evaluation performance.

### 4.3. Ablation study of the SFCNeXt

To demonstrate the rationality of the entire SFCNeXt framework, we also perform ablation experiments for SFCNeXt versus its variants. For example, the "SFCNeXt (1)", "SFCNeXt (2)," and "SFCNeXt (3)" in Table 4 represent three variants of the SFCNeXt: "SFCNeXt w/o Sex Label," "SFCNeXt w/o Conformer Encoder" and "SFCNeXt Using Original ConvNeXt Stage (3, 3, 9, 3)". Results in Table 4 can help us understand the superiority of SFCNeXt relative to all of its variants quantitatively. We can note that the use of a conformer block (SFCNeXt (2)) can reduce the MAE by 0.363, increasing the PCC by 1.4% and SRCC by 2.9%. The sex label is also very important for the performance gain: without

**Table 3.** Brain age estimation results of SFCNeXt and other lightweight competing models.

Model	MAE (Mean±STD)	PCC (Mean±STD)	SRCC (Mean±STD)	GPU Consumption
3D ResNet10	5.817±0.385	0.664±0.062	0.715±0.043	4125MB
3D ResNet18	5.240±0.665	0.709±0.063	0.719±0.034	4497MB
3D ResNet34	5.744±0.576	0.682±0.063	0.712±0.057	5133MB
3D ResNet152	5.514±0.334	0.714±0.034	0.743±0.009	8879MB
SFCN.Adam	4.624±0.204	0.800±0.026	0.787±0.026	6431MB
SFCN.AdamW	4.332±0.262	0.823±0.020	0.799±0.038	6431MB
SFCN.Adamax	4.039±0.095	0.846±0.006	0.801±0.006	6431MB
<b>SFCNeXt</b>	<b>3.683±0.009</b>	<b>0.859±0.001</b>	<b>0.817±0.001</b>	<b>6971MB</b>

**Table 4.** Brain age estimation results of SFCNeXt and its corresponding variants.

Model	MAE (Mean±STD)	PCC (Mean±STD)	SRCC (Mean±STD)
SFCNeXt (1)	4.177±0.067	0.834±0.004	0.801±0.031
SFCNeXt (2)	4.046±0.086	0.847±0.005	0.794±0.012
SFCNeXt (3)	4.216±0.082	0.848±0.001	0.811±0.002
<b>SFCNeXt</b>	<b>3.683±0.009</b>	<b>0.859±0.001</b>	<b>0.817±0.001</b>

the sex label (see the SFCNeXt (1)), the MAE degenerates to 4.177, PCC decreases to 0.834, and SRCC decreases to 0.801. All ablation experiments have shown the rationality of adopting these modules in SFCNeXt itself.

## 5. CONCLUSION

This study proposes the SFCNeXt as a lightweight end-to-end neural network architecture for effective brain age estimation with T1 MRIs. We conduct widespread experiments on a multi-site small-sized healthy cohort with biased age distribution and achieve the MAE of 3.683, PCC of 0.859, and SRCC of 0.817. Firstly, we evaluate the different combinations of SFCNeXt's parameters, and justify the parameter settings of SFCNeXt. Secondly, we compare the SFCNeXt with current lightweight SOTA models, and the quantitative comparison results of SFCNeXt show its superiority in classical metrics, such as MAE, PCC, and SRCC. Thirdly, the ablation study is performed, and all results demonstrate the rationality and necessity of all modules adopted in the SFCNeXt. To our knowledge, this is the first lightweight model specially designed for small sample brain age estimation. In conclusion, the SFCNeXt consumes comparable GPU resources to other lightweight models, is easy to deploy, and is suitable for medical situations which own insufficient subjects. In the future, it can be implemented for MRIs of brain disease or disorder groups to predict the longitudinal development trajectory and help to create an early intervention strategy.

## 6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data available in open access by [12, 13]. Ethical approval was not required, as confirmed by the license attached with the open-access data.

## 7. ACKNOWLEDGEMENTS

This work was supported in part by the Zhejiang Provincial Innovation Team Project (No. 2020R01001) and in part by the Innovation Fund of the Department of Education of Gansu Province (2022B-023).

## 8. REFERENCES

- [1] Vishnu M Bashyam, Guray Erus, Jimit Doshi, Mohamad Habes, Ilya M Nasrallah, Monica Truelove-Hill, Dhivyaa Srinivasan, Liz Mamourian, Raymond Pomponio, Yong Fan, et al., “Mri signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide,” *Brain*, vol. 143, no. 7, pp. 2312–2324, 2020.
- [2] Maxwell L Elliott, Daniel W Belsky, Annchen R Knodt, David Ireland, Tracy R Melzer, Richie Poulton, Sandhya Ramrakha, Avshalom Caspi, Terrie E Moffitt, and Ahmad R Hariri, “Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort,” *Molecular psychiatry*, vol. 26, no. 8, pp. 3829–3838, 2021.
- [3] Yu Fu, Le Xue, Yi Liao, Lin Chen, Meng Niu, Hong Zhang, Mei Tian, and Cheng Zhuo, “Cross-modality generation of amyloid pet from fdg pet for alzheimer’s disease diagnosis,” in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 2674–2681.
- [4] Na An, Yu Fu, Jie Shi, Han-Ning Guo, Zheng-Wu Yang, Yong-Chao Li, Shan Li, Yin Wang, Zhi-Jun Yao, Bin Hu, et al., “Synergistic effects of apoe and clu may increase the risk of alzheimer’s disease: acceleration of atrophy in the volumes and shapes of the hippocampus and amygdala,” *Journal of Alzheimer’s Disease*, vol. 80, no. 3, pp. 1311–1327, 2021.
- [5] Jian Cheng, Ziyang Liu, Hao Guan, Zhenzhou Wu, Haogang Zhu, Jiyang Jiang, Wei Wen, Dacheng Tao, and Tao Liu, “Brain age estimation from mri using cascade networks with ranking loss,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3400–3412, 2021.
- [6] Sheng He, Diana Pereira, Juan David Perez, Randy L Gollub, Shawn N Murphy, Sanjay Prabhu, Rudolph Pienaar, Richard L Robertson, P Ellen Grant, and Yangming Ou, “Multi-channel attention-fusion neural network for brain age estimation: accuracy, generality, and interpretation with 16,705 healthy mris across lifespan,” *Medical Image Analysis*, vol. 72, pp. 102091, 2021.
- [7] Yu Fu, Yanyan Huang, Yalin Wang, Shunjie Dong, Le Xue, Xunzhao Yin, Qianqian Yang, Yiyu Shi, and Cheng Zhuo, “Otfpf: Optimal transport-based feature pyramid fusion network for brain age estimation with 3d overlapped convnext,” *arXiv preprint arXiv:2205.04684*, 2022.
- [8] Susumu Mori, Kengo Onda, Shohei Fujita, Toshiaki Suzuki, Mikimasa Ikeda, Khin Zay Yar Myint, Jun Hikage, Osamu Abe, Hidekazu Tomimoto, Kenichi Oishi, et al., “Brain atrophy in middle age using magnetic resonance imaging scans from japan’s health screening programme,” *Brain Communications*, vol. 4, no. 4, pp. fcac211, 2022.
- [9] Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith, “Accurate brain age prediction with lightweight deep neural networks,” *Medical image analysis*, vol. 68, pp. 101871, 2021.
- [10] Sheng He, Yanfang Feng, P Ellen Grant, and Yangming Ou, “Deep relation learning for regression and its application to brain age estimation,” *IEEE Transactions on Medical Imaging*, 2022.
- [11] Tonya White, Elisabet Blok, and Vince D Calhoun, “Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed,” *Human Brain Mapping*, vol. 43, no. 1, pp. 278–291, 2022.
- [12] Noriaki Yahata, Jun Morimoto, Ryuichiro Hashimoto, Giuseppe Lisi, Kazuhisa Shibata, Yuki Kawakubo, Hitoshi Kuwabara, Miho Kuroda, Takashi Yamada, Fukuda Megumi, et al., “A small number of abnormal brain connections predicts adult autism spectrum disorder,” *Nature communications*, vol. 7, no. 1, pp. 1–12, 2016.
- [13] Saori C Tanaka, Ayumu Yamashita, Noriaki Yahata, Takashi Itahashi, Giuseppe Lisi, Takashi Yamada, Naho Ichikawa, Masahiro Takamura, Yujiro Yoshihara, Akira Kunimatsu, et al., “A multi-site, multi-disorder resting-state magnetic resonance image database,” *Scientific data*, vol. 8, no. 1, pp. 1–15, 2021.
- [14] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith, “Fsl,” *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.
- [15] Stephen M Smith, “Fast robust automated brain extraction,” *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.
- [16] Yiping Lu, Zhuohan Li, Di He, Zhiqing Sun, Bin Dong, Tao Qin, Liwei Wang, and Tie-Yan Liu, “Understanding and improving transformer from a multi-particle dynamic system point of view,” *arXiv preprint arXiv:1906.02762*, 2019.
- [17] Yu Fu, Shunjie Dong, Yi Liao, Le Xue, Yuanfan Xu, Feng Li, Qianqian Yang, Tianbai Yu, Mei Tian, and Cheng Zhuo, “A resource-efficient deep learning framework for low-dose brain pet image reconstruction and analysis,” in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2022, pp. 1–5.
- [18] Ewelina Biskup, Frances-Catherine Quevenco, Maria Teresa Ferretti, and Antonella Santucci-Chadha, “Sex differences in brain metabolic activity: Beyond the concept of brain age,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 22, pp. 10630–10631, 2019.
- [19] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie, “A convnet for the 2020s,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976–11986.
- [20] Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga, “Fast differentiable sorting and ranking,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 950–959.
- [21] Tao Qin, Tie-Yan Liu, and Hang Li, “A general approximation framework for direct optimization of information retrieval measures,” *Information retrieval*, vol. 13, no. 4, pp. 375–397, 2010.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.