Retinal Age Estimation with Temporal Fundus Images Enhanced Progressive Label Distribution Learning

Zhen Yu[1,2,3], Ruiye Chen[5,6], Peng Gui[3,4], Lie Ju[2,3,7], Xianwen Shang[5,6], Zhuoting Zhu[5,6], Mingguang He[5,6], and Zongyuan Ge[2,3,8(✉)]

[1] Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
[2] AIM for Health Lab, Monash University, Victoria, Australia
zongyuan.ge@monash.edu
[3] Monash Medical AI, Monash University, Victoria, Australia
[4] School of Computer Science, Wuhan University, Hubei, China
[5] Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia
[6] Ophthalmology, Department of Surgery, University of Melbourne, Melbourne, Australia
[7] Faculty of Engineering, Monash University, Melbourne, Australia
[8] Faculty of IT, Monash University, Melbourne, Australia
https://mmai.group

Abstract. Retinal age has recently emerged as a reliable ageing biomarker for assessing risks of ageing-related diseases. Several studies propose to train deep learning models to estimate retinal age from fundus images. However, the limitation of these studies lies in 1) both of them only train models on snapshot images from single cohorts; 2) they ignore label ambiguity and individual variance in the modeling part. In this study, we propose a progressive label distribution learning (LDL) method with temporal fundus images to improve the retinal age estimation on snapshot fundus images from multiple cohorts. First, we design a two-stage LDL regression head to estimate adaptive age distribution for individual images. Then, we eliminate cohort variance by introducing ordinal constraints to align image features from different data sources. Finally, we add a temporal branch to model sequential fundus images and use the captured temporal evolution as auxiliary knowledge to enhance the model's predictive performance on snapshot fundus images. We use a large retinal fundus image dataset which consists of ∼130k images from multiple cohorts to verify our method. Extensive experiments provide evidence that our model can achieve lower age prediction errors than existing methods.

Keywords: Retinal age estimation · label distribution learning · temporal fundus image

1 Introduction

Population ageing is a huge health burden worldwide as the risk of morbidity and mortality increases exponentially with age [2]. However, great heterogeneity

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14226, pp. 629–638, 2023.
https://doi.org/10.1007/978-3-031-43990-2_59

exists across individuals with the same chronological age, indicating chronological age poorly reflects intra-individual variation [10]. A quest for biomarkers that can accurately determine individual-specific, age-related risk of adverse outcomes has been embarked upon. Among the countless potential candidate ageing biomarkers [4,6,12], retinal age has been verified to be one of the most reliable indicators with the advantages of being rapid, non-invasive, and cost-effective [5,16,17].

With the advent of technology, deep learning (DL) algorithms have found great applications in retinal age prediction. For example, Liu et al. [9] developed a convolutional neural network (CNN) to estimate retinal age with label distribution learning (LDL) on 12k fundus images from healthy Chinese populations. Zhu et al. [17] trained a CNN regression model on the UK Biobank cohort consisting of ∼70k fundus images. The limitations of these studies include: 1) the use of a single source of data in these studies has underestimated the complexity of data variance in real-world scenarios, which limits the generalizability of retinal age prediction. 2) only snapshot databases are used in these studies and failure track a detailed trail of age-specific changes. 3) outputting a single value with direct regressions [17] ignores the ambiguity of age labels, and using fixed label distribution [9] as ground truth did not consider individual variations. Tackling these shortcomings will improve the generalizability as well as reduce technical errors in the age prediction algorithm, providing a more reliable retinal age estimate.

Therefore, in this study, we present an attempt to provide a novel accurate estimate of retinal age by learning adaptive age distribution from multiple cohorts with temporal fundus images available. Instead of learning a model using fixed label distribution as ground truth, we formulate the age estimation as a two-stage LDL task and give an adaptive distribution estimate for individual fundus images. As learning the LDL model with images from different data sources can harm the consistency and ordinality of embedding space, we introduce ordinal constraints to align the image features from different domains. Moreover, to leverage the temporal knowledge from the fundus image sequence, we add a temporal branch to capture the temporal evolution and use this auxiliary information to enhance the predictive performance of our model on snapshot images. We verify our method on a large retinal fundus dataset which consists of approximately 130k images of healthy subjects from the UKB cohort and Chinese cohorts. Extensive experiments prove that our model can achieve lower age prediction errors on multiple cohorts.

2 Method

2.1 Progressive Label Distribution Learning

As shown in Fig. 1, we formulate the retinal age estimation as a two-stage label distribution learning process. In the first stage, the model uses global features to predict a coarse age distribution on roughly discretized age labels. Each coarse age prediction is associated with a query vector corresponding to an age group.

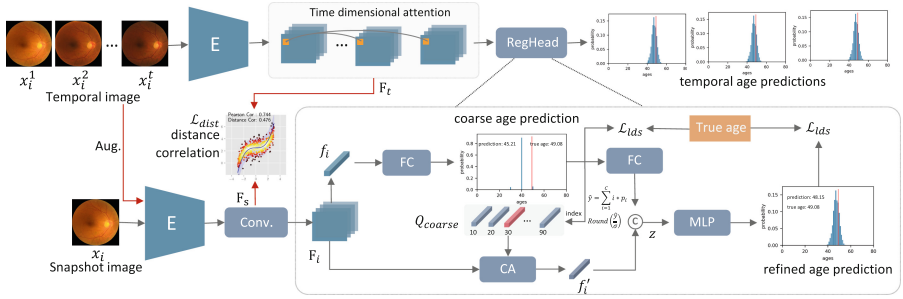


Fig. 1. Overview of the proposed method. The regression is formulated as a two-stage label distribution learning problem and the model further uses temporal knowledge to improve the snapshot learning.

Then, the model performs class attention [13] between the age group query and spatial features to generate fine-level features which are further combined with the coarse age prediction to give refined age predictions.

Formally, given a dataset with N images $\mathcal{X} = \{x_i\}_{i=1}^{N}$, the corresponding age labels $\mathcal{Y} = \{y_i\}_{i=1}^{N}$ range in $[a, b]$. The image encoder first transforms an input image x_i into a spatial feature $\mathbf{F}'_i \in \mathbb{R}^{H \times W \times C}$, then a convolutional projection layer maps $\mathbf{F}'_i$ into the base representation $\mathbf{F}_i \in \mathbb{R}^{H \times W \times D}$ and the averaged feature $f_i \in \mathbb{R}^{1 \times D}$ for the following age distribution learning. We discretize the age classes as $\hat{y}_i = \mathrm{R}\left(y_i/\delta_d\right) * \delta_d$, where $\mathrm{R}\left(\cdot\right)$ denotes the round operator and δ_d is the age bin for tuning the discretization degree. Therefore, the total discretized age class number is $C_{\delta_d} = \mathrm{R}\left(\frac{|b-a|}{\delta_d}\right)$. For the coarse-level age estimation, we set a large $\delta_d = 10$ which determines the age group queries as $\mathbf{Q}_{coarse} \in \mathbb{R}^{C_{\delta_d} \times D}$. Then, we use an FC layer with softmax applied on the f_i to calculate the coarse age distribution $p_i \in \mathbb{R}^{1 \times C_{\delta_d}}$. Different from the previous study [9] using fixed label distribution as ground truth, we directly learn the distribution from training data with discretized age labels:

$$\mathcal{L}_{lds} = \frac{1}{N}\sum_{i=1}^{N} -log\left(p_{i,\hat{y}_i}\right) + \frac{\alpha}{2N}\sum_{i=1}^{N}\left(y_i - m_i\right)^2 + \frac{\beta}{N}\sum_{i=1}^{N}\sum_{c=1}^{C_{\delta_d}} p_{i,c} * \left(y_i - m_i\right)^2 \quad (1)$$

where $m_i = \sum_{c=1}^{C_{\delta_d}} p_{i,c} * \hat{y}_c$ is the expected value of the learned distribution p_i, The first term is the cross-entropy loss which helps the model converges in an early training stage, the last two terms encourage the learned distribution to be centered and concentrated at the true age labels.

In the refining stage, the mean value of coarse age distribution m_i is used to select the age group query from Q_{coarse} to involve the computation of fine-level feature:

$$f' = GAP\left(A\left(Q_{coarse}[R\left(\frac{m_i}{\delta_d}\right)], \mathbf{F}_i; \theta_a\right)\right) \quad (2)$$

where $GAP\left(\cdot\right)$ is the global averaged pooling and $A\left(\cdot\right)$ denote the attention function with θ_a as the parameters. The key and value vectors in the atten-

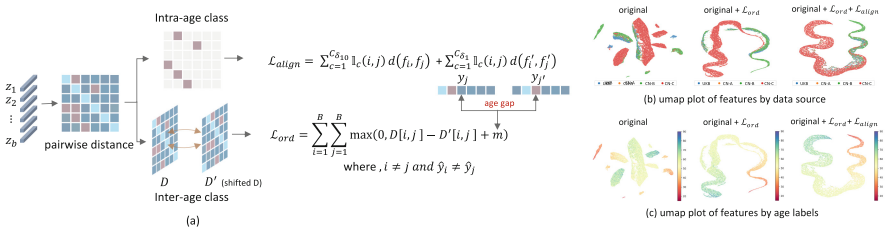


Fig. 2. Illustration of domain-aware ordinal feature alignment. The left figure shows the constraint for the inter-age class and intra-age class; The right two figures denote the feature visualization result.

tion function come from $\mathbf{F}_i$. Finally, we concatenate the f' with the mapped coarse age distribution as the final feature embedding to predict the fine-level age distribution on a small age bin of $\delta_d = 1$:

$$z = \text{concat}\left(f', \text{f}\left(p_i \odot \hat{y}_{1,\ldots,C_{\delta_d}}; \theta_f\right)\right) \tag{3}$$

$$p_i' = \text{softmax}\left(\text{mlp}\left(z; \theta_m\right)\right) \tag{4}$$

where $\text{f}(\cdot)$ denotes an FC layer with parameters θ_f, $\text{mlp}(\cdot)$ represents a multi-layer perceptron with one hidden layer and the parameter is θ_m. The training loss is the same with Eq. 1.

2.2 Cross-Domain Ordinal Feature Alignment

Although existing studies [7,14] show that formulating regression as a classification task to learn the label distribution yields better performance, the ordinal information of age relations is lost in feature space. Moreover, when the training data comes from distinct data sources, the domain variance further damages the coherence of the learned features. In Fig. 2 (b) and Fig. 2 (c), we visualize the intermediate feature learned on our fundus image dataset. As can be seen, in Fig. 2 (c) the original model produces scattered and inconsistent features for ordinal age labels, while in 2 (b) the features exhibit a clear gap for different data sources.

To address the above issues, we propose to introduce ordinal constraints in the label distribution learning and perform feature alignment to eliminate the domain variance. The key idea of imposing ordinal constraints in embedding space is to construct a set of triplets and enforce the feature distance to be consistent with the relative age gap. Specifically, for each batch of input data $\{x_1, ..., x_B\}$, we first compute their pairwise feature distance which outputs a distance matrix $D \in \mathbb{R}^{B \times B}$. Then, we construct feature triplets and calculate the distance gap by subtracting shifted distance matrix D' from the original D. In this case, each sample will have a chance to serve as the anchor to be compared with other samples. We formulate the ordinal constraint as following margin loss:

$$\mathcal{L}_{ord} = \sum_{i=1}^{B}\sum_{j=1}^{B} \max\left(0, D\left[i,j\right] - D'\left[i,j\right] + m\right), \text{s.t.} i \neq j, \text{and} \hat{y}_i \neq \hat{y}_j \tag{5}$$

where $D\left[\cdot\right]$ denotes metric of Euclidean distance, m is a dynamic margin depends on the relative age difference gap between $|\hat{y}_i - \hat{y}_j|$ and $|\hat{y}_i - \hat{y}_{j'}|$. To align features from different data domains, we directly select samples from same class and push them closer in the embedding space by minimizing the intra-class distance on both coarse-level features and fine-level features:

$$\mathcal{L}_{align} = \sum_{c=1}^{C_{\delta_{10}}} \mathbb{I}_c\left(i,j\right) d\left(f_i, f_j\right) + \sum_{c=1}^{C_{\delta_1}} \mathbb{I}_c\left(i,j\right) d\left(f'_i, f'_j\right) \tag{6}$$

where the $\mathbb{I}_c\left(i,j\right)$ is an indicator function.

2.3 Co-Learning with Temporal Fundus Images

Compared to merely learning from single snapshot images, temporal data capturing more aging information can further boost retinal age prediction. However, in practice, temporal fundus data can be limited because the individuals are often lost to follow-up. Directly learning a temporal model on these small data usually cause poor generalization. Therefore, we propose to co-train our model on limited temporal imaging data and large-scale snapshot imaging data. Our aim is to use the auxiliary knowledge from temporal data to enhance the performance of our model tested on snapshot images.

As illustrated in Fig. 1, the temporal branch consists of an image encoder, a time dimensional attention module (TDA), and a regression head. The temporal image encoder and the regression head are the same as that of the snapshot branch. We first input a fundus image sequence into the temporal branch and extract temporal features $\mathbf{F}_t \in \mathbb{R}^{T\times D}$ from the TDA module which performs time dimensional attention to capturing the correlation of local regions across temporal images. At the same time, we input augmented sequential fundus images into the snapshot branch which outputs feature $\mathbf{F}_s$[1]. Inspired by [15], we encourage the snapshot features to preserve similar relations in temporal features by optimizing the distance correlation loss:

$$\mathcal{L}_{dist} = 1 - \mathcal{R}^2\left(\mathbf{F}_s, \mathbf{F}_t\right) \tag{7}$$

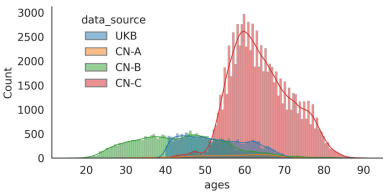
where $\mathcal{R}^2\left(\cdot\right)$ denotes the distance correlation and the detailed definition refers to [15]. We simple sum $\mathcal{L}_{lds}$, $\mathcal{L}_{ord}$, $\mathcal{L}_{align}$ and $\mathcal{L}_{dist}$ as the final loss.

[1] We omit the i in feature notions for simplicity.

3 Experiment and Results

3.1 Dataset and Implementation

Dataset : We include four datasets in our experiment: the UK Biobank cohort, CN-A, CN-B and CN-C. The UK Biobank is a publicly available prospective cohort with over 50,000 UK residents recruited in 2006[2] 45-degree fundus images were introduced in 2009 for the study subjects. CN-A was a cross-sectional study recruiting participants in eye hospitals. Another database was from the historical data collected in general hospitals, named CN-B. CN-C is an ongoing prospective cohort study that enrolled a total of 4,939 participants in 2009–2010. Participants were invited to take part in annual follow-up assessments including fundus images.



(a) Age distribution of different cohorts.

Dataset		N_{patient}	N_{image}	Age range
UKB cohort		10891	18909	40-70
Chinese cohort	CN-A	673	4005	31-79
	CN-B	12796	26628	15-91
	CN-C	4663	85298	27-91
Combined dataset		29003	133895	15-91

(b) Demographics of UKB cohort and Chinese cohort.

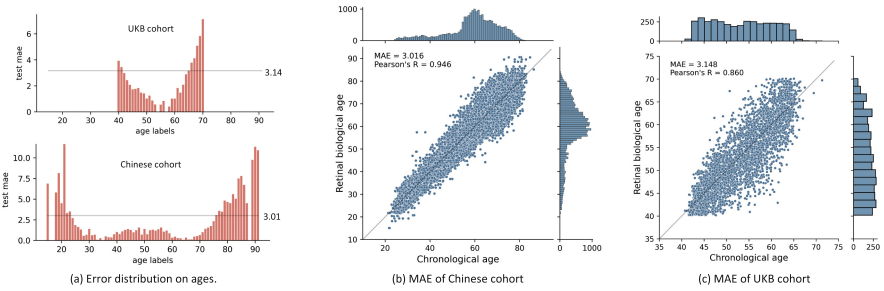
Fig. 3. Summary of retinal datasets used in this study.

Training details : As the biological age is normally developed and assessed in healthy populations where biological age is considered equal to chronological age, the model here is trained on 133895 selected snapshot images of healthy subjects without any report of systemic diseases from the four datasets (shown in Fig. 3). The temporal data is a subset of the snapshot data and consists of 2937 sequences with an average length of 5. We split the dataset into training, validation, and testing set with a ratio of 7:1:2. The standard data augmentation techniques such as random resized cropping, color transformation, and flipping are equally used in all experiments. Each image is resized to a fixed input size of 320 $\times$ 320. We use ReseNet-50 [3] as the image encoder for all models and train them using ADAM optimizer with a batch size of 100 and a training epoch of 45 with early stopping. The initial learning rates are set to 1×10^{-5} and 3×10^{-4} for the backbone layers and newly added layers, respectively. We divide the learning rate by 10 every 15 epochs.

[2] https://biobank.ndph.ox.ac.uk/showcase/browse.cgi.

Table 1. Comparison of the proposed method with existing stuides.

Method	UKB cohort		Chinese cohort		All data	
	MAE	Pearson's R	MAE	Pearson's R	MAE	Pearson's R
Direct regression	3.64	0.820	3.44	0.921	3.47	0.918
Classification	3.51	0.831	3.37	0.923	3.39	0.921
Mean-Variance [11]	3.44	0.829	3.13	0.941	3.28	0.936
Ranking-coral [1]	3.41	0.831	3.21	0.939	3.24	0.935
POE-Reg [8]	3.56	0.836	3.14	0.941	3.20	0.936
POE-CLS	3.23	0.818	3.13	0.942	3.18	0.937
PLDL	3.15	0.854	3.07	0.945	3.08	0.942
PLDL (with temp)	**3.14**	**0.859**	**3.01**	**0.946**	**3.03**	**0.943**



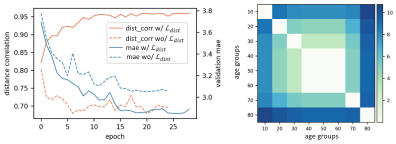
(a) Error distribution on ages. (b) MAE of Chinese cohort (c) MAE of UKB cohort

Fig. 4. Error distribution and MAE result on different cohorts.

Evaluation metrics : Consistent with previous studies, we consider the mean absolute error (MAE) and the Pearson correlation as measures for assessing the performance of models.

3.2 Quantitative Results

Comparative Study: We then compare our model with existing popular regression methods which include both direct regression method, classification-



(a) Training curves and similarity matrix of age group queries.

Component	Pearson's R	MAE
coarse stage	0.939	3.23
+ refined stage	0.941	3.11
+ $\mathcal{L}_{ord}$	0.941	3.09
+ $\mathcal{L}_{align}$	0.942	3.08
+ temporal	0.943	3.03

(b) Ablation results on the proposed model.

Fig. 5. Result of ablation study on the proposed method.

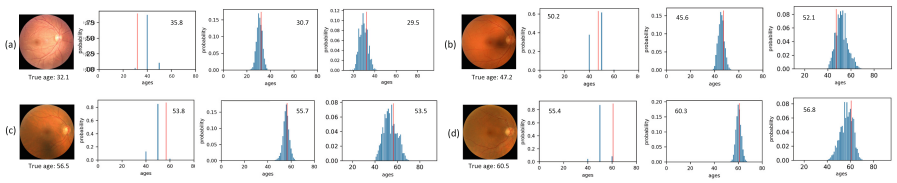


Fig. 6. Show cases of estimated age distributions. For each sample, the left two distributions are coarse prediction and refined prediction from our model, respectively. The rightmost figure denotes the result from the baseline classification model.

based methods [8,11], and ranking-based method [1]. The Mean-var improves the classification model by adding concentration regularization, the ranking-based method explicitly introduces ordinal information by combing a set of binary classifiers, and the POE methods model uncertainty with probabilistic embeddings. Table 1 shows the detailed comparison results. We denote our method as PLDL. It can be seen that the classification model outperforms the direct regression method on all the data cohorts. This observation is consistent with previous studies [11,14]. The POE-CLS is the best-performed model in all baselines, however, the performance is inferior to our model. When trained only with the snapshot images, our method achieves an MAE of 3.08 and Pearson's R of 0.942.

Ablation Study: Here, we give the ablation results of our model to illustrate how different components affect the final performance. Figure 5b shows how the performance changed when adding different components in the proposed method. As can be seen, with only the coarse stage prediction, the model produces an MAE of 3.23 and Pearson's R of 0.939. Performing the refined age stage improves the MAE by ~ 0.1. Introducing ordinal feature alignment gives a margin improvement in age prediction performance, but the feature space shows a clear improvement (see Fig. 2). At last, modelling the temporal fundus images improves the MAE from 3.08 to 3.03. In Fig. 5a, we give the learning curve of distance correlation and the MAE results on the validation set. As we can see, the snapshot features show a high correlation with the feature from the temporal branch and the MAE also becomes lower when optimizing the $\mathcal{L}_{dist}$. In Fig. 4, we give the detailed MAE distribution over different age labels on each cohort. The results indicate that the model produces high MAE on the tail ages in each cohort. Therefore, a future step to improve our model would be to consider the imbalanced learning techniques or group-wise analysis to reduce the MAE bias.

3.3 Visualization Results

In Fig. 6, we illustrate the estimated age distribution for some fundus image samples by our model and the baseline classification model. It can be seen that the refined age distributions are more accurate than the coarse prediction due to

more precise discretization. Compared to the estimated age distributions from the baseline model, our method shows a more concentrated age distribution. In Fig. 5a, we visualize the similarity matrix by computing the pair-wise cosine distance between the age group queries. It can be seen that the query vectors for age groups 40∼50, 50∼60, and 60∼70 exhibit a very high similarity which implies that these groups may share more common ageing features.

4 Conclusion

In this study, we present a novel accurate modeling of retinal age prediction. Our model is capable of learning adaptive age distribution from multiple cohorts and leveraging temporal knowledge learned from sequencing images to improve age prediction on snapshot image modeling. Our model demonstrated improved performance in four independent datasets, with an overall MAE much lower than previously proposed algorithms.

References

1. Chen, S., Zhang, C., Dong, M., Le, J., Rao, M.: Using ranking-CNN for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5183–5192 (2017)
2. Cheng, X., et al.: Population ageing and mortality during 1990–2017: a global decomposition analysis. PLoS Med. **17**(6), e1003138 (2020)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
4. Horvath, S., Raj, K.: DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat. Rev. Genet. **19**(6), 371–384 (2018)
5. Hu, W., et al.: Retinal age gap as a predictive biomarker of future risk of Parkinson's disease. Age and Ageing **51**(3), afac062 (2022)
6. Lee, J., et al.: Deep learning-based brain age prediction in normal aging and dementia. Nature Aging **2**(5), 412–424 (2022)
7. Li, Q., et al.: Unimodal-concentrated loss: Fully adaptive label distribution learning for ordinal regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20513–20522 (2022)
8. Li, W., Huang, X., Lu, J., Feng, J., Zhou, J.: Learning probabilistic ordinal embeddings for uncertainty-aware regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13896–13905 (2021)
9. Liu, C., et al.: Biological age estimated from retinal imaging: a novel biomarker of aging. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11764, pp. 138–146. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32239-7_16
10. Lowsky, D.J., Olshansky, S.J., Bhattacharya, J., Goldman, D.P.: Heterogeneity in healthy aging. J. Gerontol. Series A: Biomed. Sci. Med. Sci. **69**(6), 640–649 (2014)
11. Pan, H., Han, H., Shan, S., Chen, X.: Mean-variance loss for deep age estimation from a face. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5285–5294 (2018)

12. Peretz, L., Rappoport, N.: Deviation of physiological from chronological age is associated with health. In: Challenges of Trustable AI and Added-Value on Health, pp. 224–228. IOS Press (2022)
13. Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G., Jégou, H.: Going deeper with image transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 32–42 (2021)
14. Zhang, S., Yang, L., Mi, M.B., Zheng, X., Yao, A.: Improving deep regression with ordinal entropy. arXiv preprint arXiv:2301.08915 (2023)
15. Zhen, X., Meng, Z., Chakraborty, R., Singh, V.: On the versatile uses of partial distance correlation in deep learning. In: Computer Vision-ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVI. pp. 327–346. Springer (2022). https://doi.org/10.1007/978-3-031-19809-0_19
16. Zhu, Z., et al.: Association of retinal age gap with arterial stiffness and incident cardiovascular disease. Stroke **53**(11), 3320–3328 (2022)
17. Zhu, Z., et al.: Retinal age gap as a predictive biomarker for mortality risk. British J. Ophthalmol. 107(4), 547–554 (2022)