Coral-CVDs: A Consistent Ordinal Regression Model for Cardiovascular Diseases Grading

Zhuangzhi Gao[1,2,3], He Zhao[2], Zhongli Wu[3], Yuankai Wang[3], Gregory Yoke Hong Lip[3], Alena Shantsila[3], Eduard Shantsila[1], and Yalin Zheng[2,3(✉)]

[1] Department of Primary Care and Mental Health, University of Liverpool, Liverpool, UK
[2] Department of Eye and Vision Sciences, University of Liverpool, Liverpool, UK
yzheng@liverpool.ac.uk
[3] Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool, UK

Abstract. Cardiovascular diseases (CVDs) are the leading cause of death worldwide, emphasizing the critical need for early detection to improve treatment outcomes and prevent severe complications. Changes of the retina can be used to predict CVDs. Traditional methods often use color fundus photography (CFP) to classify CVDs risk levels into discrete categories. However, these methods typically treat the problem as a classification task, potentially overlooking the ordinal relationships among different risk levels. We propose Coral-CVDs to address this limitation by integrating these ordinal relationships into the classification models. This enhancement allows the model to better distinguish between the boundaries of adjacent risk levels. Additionally, we resolve the inconsistency issues present in traditional ordinal regression models and provide mathematical proofs to support our approach. We conducted a series of experiments using the UK Biobank data to validate our hypotheses and the results have shown its effectiveness.

Keywords: Ordinal regression · Medical grading · Cardiovascular diseases · Color fundus photography and UK Biobank

1 Introduction

Cardiovascular diseases (CVDs) is a leading cause of death worldwide [7]. It is crucial for clinicians to understand the future risks of CVDs and to provide patients with preventive advice, such as the intake of statins and change of lifestyles. The traditional risk assessment of CVDs, such as the Qrisk3 Score [12]

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-73119-8_8.

A. Bhavna et al. (Eds.): OMIA 2024, LNCS 15188, pp. 73–82, 2025.
https://doi.org/10.1007/978-3-031-73119-8_8

and the SEED Score [16], can help predict the future risk of CVDs. However, these assessment tools are limited by complex clinical practice, requiring individuals to have a certain level of medical knowledge and to complete numerous risk factor measurements. Clinicians need a simple and non-invasive tool. Color fundus photography (CFP) contains a wealth of biomarkers and has shown potential to predict ocular and systemic diseases [5,29], offering a promising approach.

Most previous studies have used CFPs to predict the risk of CVDs [6,15,22]. These studies selected risk level scores as ground truth and approached the problem as a classification task. However, conventional classification losses like multi-cross entropy usually overlook the inherent ordinal relationship between different labels. However, these labels often contain crucial ordinal relationships that can assist the classification model in discerning the boundaries between different labels. For instance, when assessing a patient with a high risk of cardiovascular diseases (CVDs), predicting "low" or "mild" risk would result in the same loss, despite the more significant clinical implications of distinguishing between "low" and "high" risks compared to differentiating "medium" from "high."

Ordinal regression (also referred to as ordinal classification or rank learning) can be viewed as an intermediary task bridging classification and regression, utilizing the ordinal relationships among different labels [17]. However, due to specific model design, it often suffers from inconsistency issues, as illustrated in Fig. 1. Inconsistency issues hinder model explanation during inference, yet clear, interpretable reasoning is crucial for artificial intelligence (AI) adoption in medical fields. In this paper, we tackle the issue of unclear boundaries among different labels. Inspired by Cao et al. [2], we propose a consistent ordinal regression model (Coral-CVDs) to leverage the ordinal relationships between different labels, aiming for more precise identification of various risk levels. We also resolve inconsistency issues and provide rigorous mathematical proofs. Our experiments have proved that our model achieves better performance than other models using CFP to predict the risk of CVDs. Our contributions can be summarized as follows:

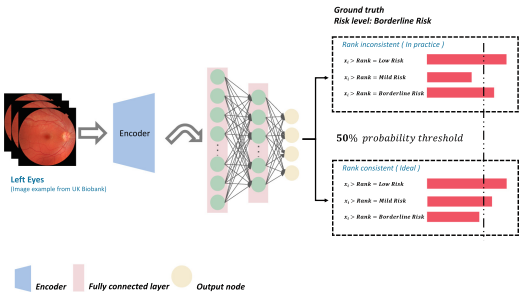

Fig. 1. Illustration depicting confidence scores that may arise among individual classifiers within ordinal regression. Ideally, the confidence scores should be consistent (below); however, in practice, they were found to be inconsistent (above).

1. We proposed a consistent ordinal regression model ***Coral-CVDs*** to use the ordinal relationship among different labels that has been overlooked in traditional methods for the prediction of risk of CVDs.
2. Our model addresses the inconsistency issues in ordinal regression model design by sharing weights, and provides rigorous mathematical proofs.
3. Our ordinal neural network architecture can be integrated with traditional classification models using CFP to predict the risk level of CVDs. Extensive experiments on the UK Biobank dataset validated our model's state-of-the-art performance and superior interpretability compared to other ordinal regression methods.

2 Related Work

CFP and the Risk Level of CVDs: Retinal vascular geometry provides insights into other vascular organ systems, such as the heart and kidneys [8,11,18]. There is a wealth of previous work investigating the use of CFP to predict CVDs [3,27]. Most of methods can be categorized into four steps: (1) selecting a specific ground truth dataset, often a pre-existing score that predicts the risk of CVDs, (2) training a classification model (3) selecting a threshold to hierarchically organize their predictions, and (4) comparing their model's performance with traditional assessments such as QRISK3 [16] and SEED [12] to validate it. In summary, previous research has primarily focused on leveraging AI to predict CVDs using CFP, rather than on methodological advancements. For example, [6,22] utilized coronary artery calcium score (CACS) as ground truth and employed EfficientNet [26] to classify the risk level of CVDs. Additionally, [15] utilized Ischemic CVDs score (ICVDs) as ground truth and applied Inception-ResNet-V2 [25] as classification model. These studies validated their effectiveness by comparing with approaches based on CVD risk factors like age, gender, and existing CVD assessment tools [28,30].

Ordinal Regression: Ordinal regression, as one of the techniques in order learning, involves estimating the rank of an instance directly through classifiers or regressors [17]. In machine learning, ordinal regression is often extended to multi-class classification using multiple binary classification subtasks to leverage the strengths of well-studied binary classifiers [9,14]. In this approach, each classifier predicts whether an instance's rank is above a certain level. The overall rank is then determined by combining these binary classification results. This idea has also been introduced into the field of computer vision. The K_th question is always formulated such that each input $x_i \in X$ is mapped to $K-1$ binary classifiers (representing $K-1$ sub-tasks), where K is the number of classes (or ranks). Each classifier is used to determine whether the rank of x_i is greater than the rank r, where $r \in [0, K-1]$. By combining the results of all binary classifiers, we can determine the predicted rank of x_i. Typical works using this approach include [4,19]. However, all these studies did not address the inconsistency issues in ordinal regression.

3 Methodology

The Coral-CVDs formulated $K-classes$ classification problem into $K-1$ binary classification sub-tasks. Specifically, in a $K-classes$ classification problem, the input is denoted as $x_i \in X$, where i is the number of input. The corresponding label $y_i \in [0, K]$. We extend the $K-classes$ classification problem to $K-1$ binary classification sub-tasks. Each sub-tasks determine whether the rank of input x_i is greater than a specific rank r, where r ranges from 0 to $K-1$. By aggregating the results of all binary classifiers, we obtain the prediction $y_i' \in [0, K)$.

3.1 Label Extension

We formulate our model as $K-1$ binary classification sub-problems, each label $y_i \in [0, 1, 2, \ldots, K]$ is transformed into $\{ y_i^0, y_i^1, \ldots, y_i^{(K-1)} \}$ and $y_i^r \in \{0, 1\}$. y_i^r means whether the rank of x_i is bigger than r as follows,

$$y_i^r = \begin{cases} 1, & \text{if rank}(x_i) > r, \text{ where } r \in [0, K-1], \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

For example, if we consider four levels of CVDs risk (low, mild, borderline, high risk level), the corresponding binary vector labels of a high risk level ($y_i = 3$) would be $[1, 1, 1]$. During model training, we utilize the extended binary labels, training a single network with $K-1$ binary classifiers in the output layer.

3.2 Coral-CVDs

Our proposed Coral-CVDs model architecture is illustrated in Fig. 2. It comprises three main modules. 1) Feature extraction modules: We utilize the pre-trained RETFound model [31], a self-supervised model trained on a large cohort of CFPs of left eyes, to effectively extract intrinsic features from CFPs, serving as our feature extractor. 2) Dimensionality reduction module: It incorporates several fully connected layers to gradually reduce the dimensionality of the features from high-dimensional to low-dimensional, to prevent information loss and enhance expressive capability. 3) Rank consistent module: It ensures consistency and facilitates ordinal regression, which is our main focus here.

3.3 Rank Consistent Module

As illustrated in Fig. 2, the K-class classification problem is formulated into m binary classification sub-tasks. In Coral-CVDs, the nodes in the penultimate layer are denoted by α, and $\mathcal{W}$ denotes the weight parameters of the neural network, where b represents the bias units. Here, all $\mathcal{W}$ share the same weights, and we employ $K-1$ learnable biases. The output of penultimate layer is defined as

$$\sum_{j=1}^{m} \mathcal{W}\alpha_j + b_k \quad \text{where} \quad k = K-1 \tag{2}$$

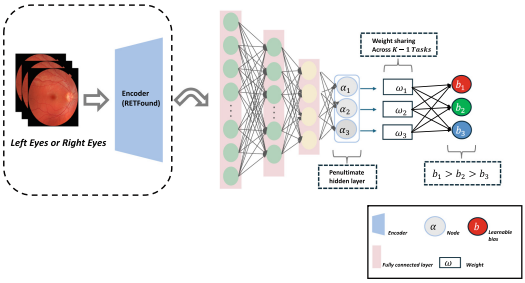

Fig. 2. Overview of the proposed Coral-CVDs model architecture. The Coral-CVDs model accepts either left or right fundus images as the input and employs RETFound as the feature encoder. In the penultimate layer, there are $K-1$ nodes representing $K-1$ tasks, with each task determining whether the risk of input x_i exceeds a specific risk level r. We use 3 nodes as there are 4 risk levels for our task. All nodes in this layer share the weights and utilize $K-1$ learnable biases to ensure consistency.

where m denotes the number of nodes in the penultimate layer. The last layer was $K-1$ binary classifiers and we employed σ represents the logistic sigmoid function so the output of last layer was

$$P(Task_k = 1) = \sigma(\sum_{j=1}^{m} \mathcal{W}\alpha_j + b_k) \tag{3}$$

In our task, we had 4 risk levels (low, mild, borderline and high risk level) leading to 3 sub-tasks where each task determines whether the risk level of the input x_i exceeds a specific threshold. If we want to keep consistent, it means that $P(\text{Task}_1 = 1) > P(\text{Task}_2 = 1) > P(\text{Task}_3 = 1)$, However, given that all nodes in the penultimate layer share the same weights, the question now is to demonstrate the rank monotonicity of the learnable bias b, such that $b_1 > b_2 > b_3$.

3.4 Theoretical Guarantees for Rank Consistency

We employ weighted cross entropy as the loss function, and the following theorem demonstrates that by minimizing the loss $\mathcal{L}$, the learned bias units of the output layer are non-increasing. We provide the complete proof in the supplementary materials.

3.5 Rank Label

To obtain the final results corresponding to the true classes or true ranks, we collect the predictions from all binary classifiers and transform them into true ranks. Based on the responses of the $K-1$ binary classifiers, the predicted rank label for an input x_i is given by:

$$y'i = 1 + \sum_{i=1}^{K-1} 1(P(Task_k = 1) > 0.5), \text{where} \quad y'_i \in [0, K] \tag{4}$$

In this function, we use a threshold of 50% probability to determine the output of the binary classifier.

4 Experiments

We utilize the CFPs of the UK Biobank dataset [1], a biomedical repository containing comprehensive health and genetic information from over half a million participants in the United Kingdom.

Color Fundus Photography. We establish the quality of CFPs by employing the method proposed in [10]. Subsequently, we utilize the approach proposed by Shen et al. [24] to improve the quality of CFP images that were regarded as 'reject' by [10]. These images are then re-graded, and all 'good' and 'usable' CFPs are used for the model development and evaluation. In total we obtain CFPs of 44,443 individuals, each with left and right fundus images.

Stratification of CVD Risk Levels: We investigate coronary heart disease, ischemic stroke, and transient ischemic attack as the focal points of our CVD research. Following [21], we calculate the QRISK3 score for each individual. QRISK3 scores range from 0 to 100, with higher scores indicating a greater risk of developing CVDs over the next 10 years. We then stratify the QRISK3 scores into four risk categories: low, middle, borderline, and high, as shown in Fig. 3.

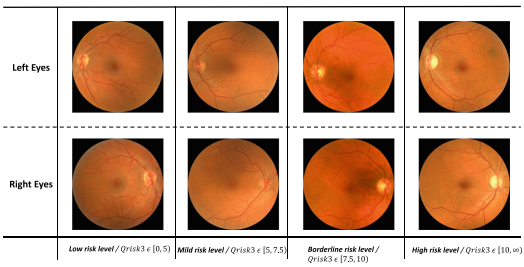

Fig. 3. Four example pairs of left and right eyes at varying risk levels for CVDs from the UK Biobank dataset.

After linking the CFP data with the CVD risk levels, we obtained data of 31,512 participants, each with images of both eyes. The dataset was then divided into training, testing, and validation datasets using a 7:2:1 ratio at individual level. Our data is imbalanced. For the **training dataset**, we had 9,097, 6,811, 3,442, and 2,708 instances for high, mild, borderline, and low risk levels respectively, while 2,600, 1,947, 984, and 774 instances for the **testing dataset**.

4.1 Implement Details and Evaluation Metrics

Implementation Details: We have found that incorporating feature extraction as part of model training and pre-extracting features before training yield similar results. For efficiency, we chose to pre-extract features and use them for model training. The network is trained end-to-end using the Adam optimizer [13] with an initial learning rate of 0.01 reduced by a factor of 0.5 every 50 epochs based on the minimum validation loss. We use a batch size of 256 and train for 500 epochs with early stopping on a Geforce *RTX 4090Ti* GPU.

Evaluation Metrics: We use both Mean Absolute Error (**MAE**) and Mean Squared Error (**MSE**) as metrics for ordinal regression problems. To evaluate our model's classification performance, we employ common classification metrics, including **Accuracy**, **Recall**, **Precision**, and $\mathbf{F}_1$ **score**.

5 Results

To evaluate the performance of the proposed Coral-CVD model, we compare with *CE-NN* as the baseline model that treat the task as a classification problem, and a classic ordinal regression models, namely *OR-NN* [20], at both classification and regression levels. We evaluate our model using data from both left and right eyes. Table 1 presents the quantitative results.

Table 1. Results of the CVDs risks level on the testing set based on Left and Right eyes respectively. ACC (%), Recall (%), Precision (%) and F1 (%) are used to report the classification performance whilst MAE (%), MSE (%) for the regression performance.

Eye	Methods	ACC (%)↑	Recall (%)↑	Precision (%)↑	F1 (%)↑	MAE↓	MSE↓
Left	CE-NN	**0.608**	0.428	0.406	0.380	0.707	1.526
Left	OR-NN [20]	0.544	**0.434**	**0.437**	**0.434**	**0.693**	1.269
Left	**Ours**	0.534	0.433	**0.437**	**0.434**	0.696	**1.249**
Right	CE-NN	**0.613**	0.431	0.436	0.380	0.705	1.537
Right	OR-NN [20]	0.549	0.446	**0.456**	**0.449**	**0.659**	**1.150**
Right	**Ours**	0.549	**0.448**	0.453	**0.449**	0.661	1.159

Table 1, shows that for using either left or right eyes, all the ordinal models exhibit superior performance to those of the traditional classification model (CE-NN) except for accuracy. Compared with the traditional ordinal regression model (OR-NN), our model exhibits similar performance and outperforms it in some metrics. For instance, regarding the left eye, our model and the OR-NN perform equally well in terms of precision and F1 score, while our model surpasses OR-NN in MSE. However, since our model achieves consistency, it offers better robustness and interpretability compared to the OR-NN.

In particular, we aim to address the traditional use of CFP for predicting the risk level of CVDs through a classification model, which often overlooks the ordinal relationship between the risk categories. As illustrated in Fig. 4, The CE-NN only identifies the samples with prominent features, resulting in most mild and borderline samples being misclassified as low and high risk levels. This explains why, despite achieving high accuracy, other metrics show poor performance. In medical applications, accurately identifying both diseased and non-diseased samples is crucial. Coral-CVDs achieves this effectively by incorporating ordinal relationships between different labels.

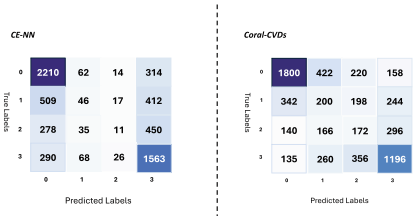

Fig. 4. Comparison of confusion matrices for CE-NN and Coral-CVDs.

6 Explanations for Coral-CVDs

We employ Gradient-weighted Class Activation Mapping (Grad-CAM) [23] to investigate the explanation of Coral-CVDs. It uses the gradients of the target class flowing into the specific layer to produce a coarse localization map highlighting important regions in the original image. We show a sample with a QRISK3 score of 19.519, indicating a high-risk level. The Grad-CAM visualizations for this sample at different risk levels highlight the areas of the input data that contribute most to the classification, in the supplementary materials.

7 Conclusion

In this paper, we introduced a consistent ordinal regression model: Coral-CVDs. Coral-CVDs, by incorporating the ordinal relationships between different labels into the model, helps the model better understand the boundary relationships between different risk levels. Additionally, by introducing a rank-consistent module, Coral-CVDs resolves the issue of rank inconsistency compared to traditional ordinal regression models. This work highlights the strengths of Coral-CVD in tackling classification problems under the ordinal regression framework and opens new avenue for wider applications.

References

1. UK Biobank - UK Biobank (2024), https://www.ukbiobank.ac.uk
2. Cao, W., Mirjalili, V., Raschka, S.: Rank consistent ordinal regression for neural networks with application to age estimation. Pattern Recognition Letters **140**, 325–331 (2020)
3. Chang, J., Lee, J., Ha, A., Han, Y.S., Bak, E., Choi, S., Yun, J.M., Kang, U., Shin, I.H., Shin, J.Y., et al.: Explaining the rationale of deep learning glaucoma decisions with adversarial examples. Ophthalmology **128**(1), 78–88 (2021)
4. 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)
5. Cheung, C.Y., Xu, D., Cheng, C.Y., Sabanayagam, C., Tham, Y.C., Yu, M., Rim, T.H., Chai, C.Y., Gopinath, B., Mitchell, P., et al.: A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. Nature biomedical engineering **5**(6), 498–508 (2021)
6. Cho, S., Song, S.J., Lee, J., Song, J., Kim, M.S., Lee, M., Lee, J.: Predicting coronary artery calcium score from retinal fundus photographs using convolutional neural networks. In: Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part I 19. pp. 599–612. Springer (2020)
7. Contributors, W.E.: Cardiovascular diseases, https://www.webmd.com/heart-disease/diseases-cardiovascular, accessed: 2023-08-17
8. Diaz-Pinto, A., Ravikumar, N., Attar, R., Suinesiaputra, A., Zhao, Y., Levelt, E., Dall'Armellina, E., Lorenzi, M., Chen, Q., Keenan, T.D., et al.: Predicting myocardial infarction through retinal scans and minimal personal information. Nature Machine Intelligence **4**(1), 55–61 (2022)
9. Frank, E., Hall, M.: A simple approach to ordinal classification. In: Machine Learning: ECML 2001: 12th European Conference on Machine Learning Freiburg, Germany, September 5–7, 2001 Proceedings 12. pp. 145–156. Springer (2001)
10. Fu, H., Wang, B., Shen, J., Cui, S., Xu, Y., Liu, J., Shao, L.: Evaluation of retinal image quality assessment networks in different color-spaces. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. pp. 48–56. Springer (2019)
11. Günthner, R., Hanssen, H., Hauser, C., Angermann, S., Lorenz, G., Kemmner, S., Matschkal, J., Braunisch, M.C., Küchle, C., Renders, L., et al.: Impaired retinal vessel dilation predicts mortality in end-stage renal disease. Circulation research **124**(12), 1796–1807 (2019)
12. Hippisley-Cox, J., Coupland, C., Brindle, P.: Development and validation of qrisk3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. bmj **357** (2017)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Li, L., Lin, H.T.: Ordinal regression by extended binary classification. Advances in neural information processing systems **19** (2006)
15. Ma, Y., Xiong, J., Zhu, Y., Ge, Z., Hua, R., Fu, M., Li, C., Wang, B., Dong, L., Zhao, X., et al.: Deep learning algorithm using fundus photographs for 10-year risk assessment of ischemic cardiovascular diseases in china. Science bulletin **67**(1), 17–20 (2022)

16. Majithia, S., Tham, Y.C., Chee, M.L., Nusinovici, S., Teo, C.L., Chee, M.L., Thakur, S., Soh, Z.D., Kumari, N., Lamoureux, E., et al.: Cohort profile: the singapore epidemiology of eye diseases study (seed). International journal of epidemiology **50**(1), 41–52 (2021)
17. McCullagh, P.: Regression models for ordinal data. Journal of the Royal Statistical Society: Series B (Methodological) **42**(2), 109–127 (1980)
18. McGeechan, K., Liew, G., Macaskill, P., Irwig, L., Klein, R., Klein, B.E., Wang, J.J., Mitchell, P., Vingerling, J.R., DeJong, P.T., et al.: Meta-analysis: retinal vessel caliber and risk for coronary heart disease. Annals of internal medicine **151**(6), 404–413 (2009)
19. Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal Regression with Multiple Output CNN for Age Estimation. pp. 4920–4928. Las Vegas, NV, USA (2016)
20. Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output cnn for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4920–4928 (2016)
21. Parsons, R.E., Liu, X., Collister, J.A., Clifton, D.A., Cairns, B.J., Clifton, L.: Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank. Heart **109**, 1690–1697 (2023)
22. Rim, T.H., Lee, C.J., Tham, Y.C., Cheung, N., Yu, M., Lee, G., Kim, Y., Ting, D.S., Chong, C.C.Y., Choi, Y.S., et al.: Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. The Lancet Digital Health **3**(5), e306–e316 (2021)
23. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: visual explanations from deep networks via gradient-based localization. International journal of computer vision **128**, 336–359 (2020)
24. Shen, Z., Fu, H., Shen, J., Shao, L.: Modeling and enhancing low-quality retinal fundus images. IEEE transactions on medical imaging **40**(3), 996–1006 (2020)
25. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017)
26. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
27. Ting, D.S.W., Wong, T.Y.: Eyeing cardiovascular risk factors. Nature Biomedical Engineering **2**(3), 140–141 (2018)
28. Tseng, R.M.W.W., Rim, T.H., Shantsila, E., Yi, J.K., Park, S., Kim, S.S., Lee, C.J., Thakur, S., Nusinovici, S., Peng, Q., et al.: Validation of a deep-learning-based retinal biomarker (reti-cvd) in the prediction of cardiovascular disease: data from uk biobank. BMC medicine **21**(1), 28 (2023)
29. Wagner, S.K., Fu, D.J., Faes, L., Liu, X., Huemer, J., Khalid, H., Ferraz, D., Korot, E., Kelly, C., Balaskas, K., et al.: Insights into systemic disease through retinal imaging-based oculomics. Translational vision science & technology **9**(2), 6–6 (2020)
30. Yi, J.K., Rim, T.H., Park, S., Kim, S.S., Kim, H.C., Lee, C.J., Kim, H., Lee, G., Lim, J.S.G., Tan, Y.Y., et al.: Cardiovascular disease risk assessment using a deep-learning-based retinal biomarker: a comparison with existing risk scores. European Heart Journal-Digital Health **4**(3), 236–244 (2023)
31. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (2023)