Improving Consistency in Cardiovascular Disease Risk Assessment: Cross-Camera Adaptation for Retinal Images

Weiyi Zhang[1], Danli Shi[1,2], Mingguang He[1,2,3]
[1]School of Optometry, The Hong Kong Polytechnic University
[2]Research Centre for SHARP Vision (RCSV), The Hong Kong Polytechnic University
[3]Centre for Eye and Vision Research (CEVR)
wei-yi.zhang@connect.polyu.hk
{danli.shi, mingguang.he}@polyu.edu.hk

Abstract

This paper introduces a novel cross-camera domain adaptation method to address the challenges associated with achieving consistency and adaptability in cardiovascular disease (CVD) risk assessment using retinal images captured by conventional and portable cameras. The proposed method leverages an enhanced ordinal CVD risk classification approach to predict CVD risk levels, effectively capturing the ordinal relationship and implicit information embedded within retinal images. Additionally, a plug-and-play risk consistency loss is incorporated into the image translation model to ensure alignment in risk assessment between different image domains. Experimental evaluations on diverse datasets demonstrate the effectiveness and superiority of the proposed method in achieving consistent CVD risk assessment across various camera models. The results highlight the potential of the proposed approach to enhance early detection and intervention of CVD, utilizing the convenience and cost-effectiveness of portable retinal imaging technology. Overall, this research contributes to the field of computer-aided medical imaging by providing a robust and adaptable solution for CVD risk assessment, ultimately benefiting patients and healthcare providers in their efforts to combat CVD.

1. Introduction

Cardiovascular Diseases (CVD) are prevalent conditions requiring early detection and intervention for effective treatment and patient recovery. While medical guidelines, such as the WHO-CVD score [6], exist for predicting CVD risk based on patient physical data, they often require comprehensive and invasive examinations. Retinal image analysis offers a non-invasive approach to visualize microcirculation, and changes in retinal vessel calibers can indicate CVD risk factors [1]. Retinal cameras have emerged as a cost-effective alternative for efficient and convenient CVD risk assessment based on retinal images. Machine learning methods have also been employed for CVD risk evaluation using retinal images [18][11]. However, the commonly used retinal cameras in clinical practice and research, such as Topcon, Canon, Zeiss, and Heidelberg, are bulky desktop models, and high-quality retinal samples ideal for cardiovascular risk inference are primarily captured using these devices. Due to variations in specifications among different brands and models of retinal cameras [3][4], images obtained from low-cost portable retinal cameras may suffer from compromised image quality, such as overexposure of the optic disc, and exhibit significant stylistic differences compared to results obtained using large-scale devices [16]. These factors contribute to domain discrepancies and adaptability challenges in retinal camera-based research, consequently hindering the generalizability and robustness of machine learning-based CVD risk assessment tasks across different cameras.

To tackle camera domain variations in cross-camera CVD risk assessment tasks, prominent emphasis has been placed on feature and image alignment techniques in existing methods. For instance, Liu et al. [8] proposed a collaborative feature ensembling adaptation method for optic disc and cup segmentation, while Lei et al. [7] developed an unsupervised domain adaptation approach for the same task. Additionally, Yang et al. [17] introduced a camera adaptation method based on Residual-CycleGAN for diabetic retinopathy screening. Moreover, Ran et al. [12] leveraged a source-free active domain adaptation method to generate features of color fundus images by noise, and Wei et al. [15] explored the use of transformers to mitigate this issue. Despite their focus on the explicit quality of target domain images, requiring a substantial number of challenging-to-acquire training images, these approaches often overlook

the implicit image features and fall short in addressing the issue of consistency in risk assessment across different cameras.

In this paper, we propose a novel plug-and-play cross-camera domain adaptation method aimed at transforming images captured by portable cameras into images that closely resemble those acquired by standard clinical practice cameras. Our method ensures consistency in CVD risk assessment between the different image domains. Firstly, we introduce an enhanced ordinal CVD risk classification model based on retinal images, named Enhanced Ordinal CVD Risk (EOCRisk) classification. This model leverages ordinal loss to improve performance in predicting CVD risk levels, providing a more nuanced approach compared to traditional classification methods. Secondly, we design an insertable risk consistency loss that can be applied to any image translation framework. Through ablation experiments conducted on pix2pixHD [14] and CycleGAN [19], we demonstrate the effectiveness of the EOCRisk classification model and the risk consistency loss in improving the consistency of CVD risk assessment in different image translation tasks across portable and conventional retinal cameras. Furthermore, our method is validated on external datasets captured by three different portable cameras, showing promising results in addressing domain variations introduced by various cameras.

2. Methods

2.1. Enhanced Ordinal CVD Risk Classification

To address the task of CVD risk score prediction, we propose the Enhanced Ordinal CVD Risk Classification (EOCRisk). EOCRisk utilizes an ordinal multi-label classification method that transforms the original CVD scores into a 10-level ordinal scale, treating each level as a distinct class for multi-class classification. The core component of EOCRisk is the Weighted Logarithmic Loss ($Loss_{\mathrm{WL}}$) function, defined as follows:

$$Loss_{\mathrm{WL}}(p, l, im) = \mathrm{mean}\Bigg(-\sum_{i=1}^{N} \Bigg(\log(s(p)_{i,1}) \cdot l_i + \log(s(p)_{i,0}) \cdot (1 - l_i) \Bigg) \cdot im_i \Bigg), \tag{1}$$

where p refers to the raw, unprocessed output predictions, l indicates the CVD risk level for each sample, im contains the importance weights assigned to each sample, and N represents the number of samples in the batch. The term $s(p)_{i,1}$ denotes the probability for the i-th sample to be classified as a high-risk signal, while $s(p)_{i,0}$ represents the probability for the i-th sample to be classified as a low-risk signal. The variable l_i corresponds to the one-hot vector indicating the CVD risk level (ranging from 0 to 9) for the i-th sample, and im_i denotes the importance weight assigned to the i-th sample.

Explanatorily speaking, we use the softmax probabilities of high-risk and low-risk signals to calculate a weighted sum for the loss function. This involves taking the negative logarithms of these probabilities and weighting them based on the corresponding risk levels and importance weights assigned to each sample. By summing up these weighted values and taking the mean across the batch, we obtain the overall loss. This methodology ensures that the logarithmic probabilities are appropriately weighted, taking into account both the risk levels and the importance assigned to each sample. The final predicted risk level by EOCRisk is determined by the cumulative value of the predicted logits across the various risk signals, with higher values indicating a higher risk level. The overall network architecture is illuminated in Fig. 1.

EOCRisk effectively captures the consistency and ordinal relationship between risk levels while mitigating the impact of category imbalance. By leveraging this approach, we enhance the accuracy of CVD risk assessment across various cameras and facilitate implicit learning of relevant retinal image features associated with CVD diagnosis.

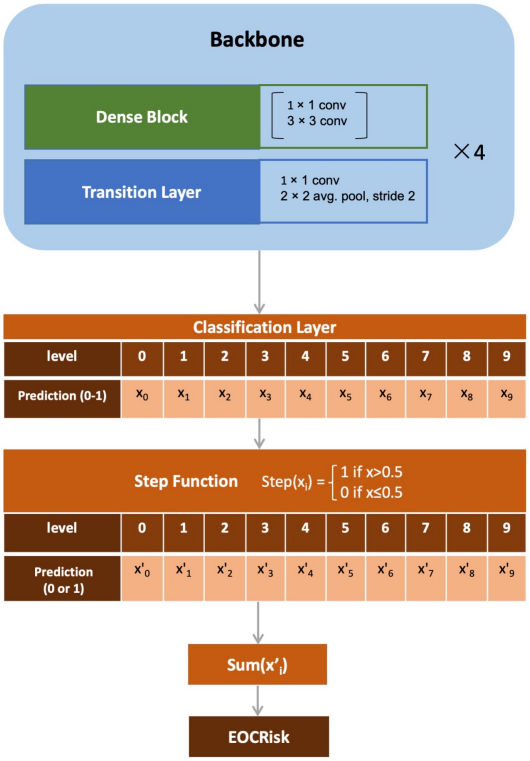

Figure 1. The network structure of the EOCRisk classification model and the calculation method for predicting the final risk values.

2.2. CVD Risk Consistency Loss

Building upon the aforementioned risk prediction model EOCRisk, we introduce a risk consistency loss that can be seamlessly integrated into any image translation framework. To illustrate this concept, we consider the pix2pixHD generative adversarial network architecture [14], as depicted in Fig. 2 (right). Specifically, we focus on the translation of images captured by an Optain camera to images resembling those from a Topcon camera.

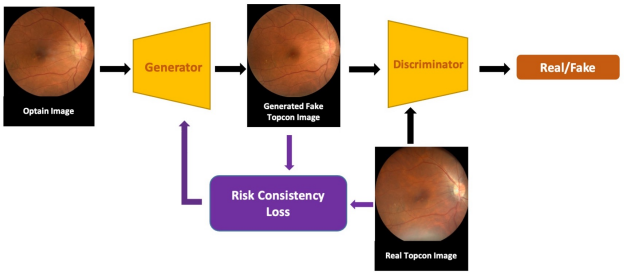

Figure 2. The model architecture of applying the risk consistency loss in pix2pixHD.

In the proposed approach, the generator network takes Optain camera images as input and synthesizes corresponding fake Topcon camera images. To quantify the consistency in CVD risk assessment between the two image domains, we employ the mean squared error (MSE) as a measure of agreement between the risk values of the target domain images and the generated fake images. This MSE-based metric captures the level of concordance in CVD risk evaluation. Mathematically, the risk consistency loss ($L_{\text{consistency}}$) is formulated as:

$$L_{\text{consistency}} = \frac{1}{N}\sum_{i=1}^{N}\left(\text{Risk}(x_{\text{target}}^{(i)}) - \text{Risk}(G(x_{\text{target}}^{(i)}))\right)^2, \tag{2}$$

where N represents the batch size, $x_{\text{target}}^{(i)}$ denotes the i-th target domain image, $G(\cdot)$ signifies the image translation function responsible for mapping target domain images to their fake counterparts, and $\text{Risk}(\cdot)$ represents the CVD risk assessment function. The loss term captures the squared difference between the risk value of a target domain image $x_{\text{target}}^{(i)}$ and the risk value of the corresponding generated fake image $G(x_{\text{target}}^{(i)})$. By averaging these squared differences across the batch, we obtain the overall risk consistency loss. During training, minimizing this loss aids in aligning the CVD risk assessment between the target domain and the generated fake images.

3. Experiments

3.1. Datasets

For the training of the EOCRisk assessment model, we utilized two datasets: the UK Biobank (UKB) Study [13] and a Chinese cohort dataset, with the aim of ensuring diversity in camera types and subjects' ethnicities. The UKB dataset consists of retinal images captured using the Topcon 3D OCT-1000 MKII camera, while the Chinese dataset comprises images obtained from the TRC-NW6S camera (Topcon, Tokyo, Japan). After quality assessment and screening, the final dataset included a total of 105,277 retinal images from the UKB development set and 30,092 retinal images from the Chinese dataset. The data were divided into train, validation, and test sets at a patient-level ratio of 3:1:1, with random allocation across the datasets.

During the camera adaptation experiment, we collected 1,292 pairs of images captured by the Topcon TRC-NW8 camera and the Optain portable camera. Out of these pairs, 65 pairs of images were set aside as the test set at the patient level, while the remaining pairs were allocated for training and validation, with a ratio of 4:1. The data does not contain information related to the subjects' CVD risk. For external validation, we acquired images of the right and left eyes of 11 subjects on three different brands of Optain, CenterVue-DRS, and SysEye portable fundus cameras, to assess the generalization capability of our proposed method.

3.2. Implementation details

During the training process of the EOCRisk assessment model, we selected Densenet169 [5] as the backbone architecture and initialized it with pre-trained weights from ImageNet [2]. The input images were consistently resized to a fixed dimension of 224x224. We employed a learning rate of 1e-5 with a batch size of 64 for optimization. To enhance the model's generalization capability, various data augmentation techniques were applied, including horizontal and vertical flipping, random zoom in and zoom out, and RandomErasing.

Regarding the training of the image translation model for camera adaptation, we set the weights for the risk consistency loss in pix2pixHD [14] and CycleGAN [19] to 3 and 4, respectively. Throughout the training process, the images underwent random horizontal/vertical flips and cropping and were then resized to a fixed size of 768x768 pixels. We utilized the Adam optimizer with optimization parameters $beta_1 = 0.5$ and $beta_2 = 0.999$. The learning rate was adjusted every 50 iterations using the PyTorch lr_scheduler, starting with an initial value of 0.0002. Each training session consisted of 200 epochs and was conducted on an NVIDIA GeForce RTX 3090 GPU.

Table 1. The performance comparison between a conventional regression model and our proposed EOCRisk classification model for predicting CVD risk.

Methods	R2-Score	MAE	RMSE	Explained Var.	Correlation
General Regression	0.300	1.85	2.34	0.500	0.682
EOCRisk (Ours)	**0.480**	**1.53**	**1.00**	**0.518**	**0.726**

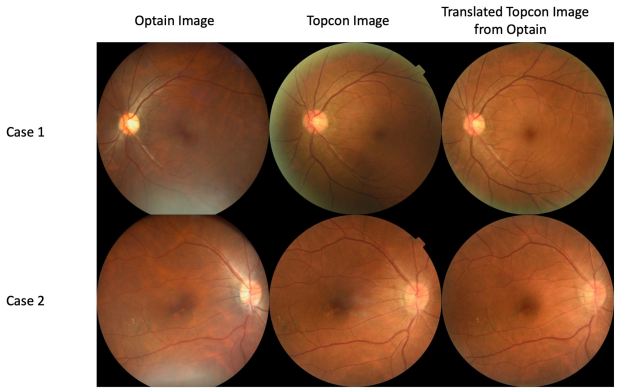

Figure 3. Two visualization examples of camera domain adaptation using pix2pixHD with our proposed EOCRisk consistency loss.

3.3. Evaluation criteria

Performance evaluation metrics are defined to assess the effectiveness of the proposed method in achieving consistency in CVD risk assessment. The metrics include the coefficient of determination (R2-score) [9], mean absolute error (MAE), root-mean-square error (RMSE), explained variance [10], and correlation coefficient. The R2 values indicate the improvement in the predictive power of the risk assessment models, while the RMSE and MAE values demonstrate the reduction in the prediction errors. The explained variance and correlation coefficients highlight the increased consistency and alignment between the risk values before and after applying the domain adaptation technique. Note that in the results below, the R2-score and explained variance may occasionally be negative. This implies that the model's performance is worse than that of a horizontal line fitted directly to the data or than simply using the mean of the observed outcomes as a predictor. Such negativity indicates exceptionally poor model performance.

3.4. Ablation studies

Firstly, we compared the performance of our proposed EOCRisk classification model with a conventional regression model when both models were trained using the same backbone and training scheme. The results of the CVD risk prediction can be seen in Table 1. The calculation of the ground-truth risk followed a recently calibrated WHO-CVD risk calculation method [6]. Our approach outperforms the regular regression model in all metrics, with an improvement of 18 percentage points in R2-Score.

In addition, we conducted ablation experiments on two image translation models to validate the effectiveness of the risk consistency loss and the superiority of the EOCRisk assessment model over the regular CVD risk regression model. The results of the ablation experiments are presented in Table 2. Specifically, pix2pixHD [14] was primarily applied to the translation from the image domain captured by the portable camera (Optain) to the image domain captured by the upright camera (Topcon), while CycleGAN [19] explored the translation in both directions. The findings indicate that performing domain translation using the two baseline models had limited impact in mitigating the disparities in CVD risk across different camera domains. However, with the introduction of the risk consistency loss, the consistency of CVD risk assessment after domain translation improved for both datasets. Notably, the risk loss based on the EOCRisk assessment model outperformed the loss based on the regular regression model. By utilizing pix2pixHD and incorporating the risk consistency loss with EOCRisk for camera adaptation, the R2-score and correlation coefficient for CVD risk between the two camera domains reached 0.6404 and 0.8275, respectively, while the MAE decreased to 0.8769, demonstrating the high efficacy of our approach. The visualization of this method is presented in Fig 3.

3.5. External validation

By comparing the CVD risk value indicators before and after applying our domain adaptation method for each combination of camera models (Optain, DRS, and SysEye), as summarized in Table 3, we observed improvements in the consistency of CVD risk assessment across the different camera domains after incorporating our risk consistency loss based on the EOCRisk assessment model. Specifically, for the Optain and DRS camera combination, the R2 score improved from -3.1700 to 0.3928, the RMSE decreased from 4.3684 to 1.1902, the MAE decreased from 3.9167 to 1.0833, and the correlation coefficient increased from -0.3892 to 0.6513. Similar improvements in CVD risk assessment consistency were observed for the Optain and SysEye, as well as the DRS and SysEye camera combinations. This demonstrates the effectiveness of our method in improving the consistency of CVD risk not only between the

Table 2. The disparities in CVD risk levels after applying different camera adaptation methods to datasets collected from two retinal cameras. The datasets are denoted as O for Optain data and T for Topcon data.

Models		Methods	R2	RMSE	MAE	Explained Var.	Correlation
Original Data			-0.3047	2.5115	1.9385	-0.2765	0.3782
Pix2pixHD[14]	O->T	Baseline	0.5990	1.3923	0.9846	0.6186	0.7853
		Baseline with Regression Risk Loss	0.6245	1.3473	0.9231	0.6253	0.7886
		Baseline with EOCRisk Loss	**0.6404**	**1.3185**	**0.8769**	**0.6463**	**0.8275**
Cycle GAN[19]	O->T	Baseline	-0.0094	2.0718	1.4615	0.0151	0.5150
		Baseline with Regression Risk Loss	0.0630	1.9962	1.3692	0.0924	0.5655
		Baseline with EOCRisk Loss	**0.3524**	**1.6595**	**1.0615**	**0.3769**	**0.6788**
	T->O	Baseline	0.0883	2.1231	1.6462	0.0888	0.4467
		Baseline with Regression Risk Loss	0.2781	1.8892	1.4769	0.2812	0.5790
		Baseline with EOCRisk Loss	**0.3372**	**1.8102**	**1.4615**	**0.3384**	**0.6256**

Table 3. Comprehensive external evaluation of the CVD risk assessment performance for three image domains, illustrating the comparison between the "before" scenario (without domain adaptation) and the "after" scenario (with domain adaptation using pix2pixHD with our proposed EOCRisk classification consistency loss).

Data	Methods	R2	RMSE	MAE	Explained Var.	Correlation
Optain &DRS	before	-3.1700	4.3684	3.9167	-2.3672	-0.3892
	after	0.3928	1.1902	1.0833	0.3958	0.6513
Optain &SysEye	before	-2.7500	4.5644	3.8333	-2.7450	-0.3942
	after	0.1981	1.5275	1.6667	0.2840	0.4436
DRS &SysEye	before	-0.6571	2.7538	2.0833	0.0121	0.4790
	after	0.3214	1.2583	1.0833	0.4673	0.6222

upright camera domain and the portable camera domain, but also across different portable camera models.

4. Discussion

Our proposed plug-and-play EOCRisk effectively mitigates the inconsistency in predicting CVD risk using retinal images across different cameras and demonstrates efficacy across various image translation networks. Existing camera adaptation methods for retinal images often prioritize visual fidelity across different cameras but overlook their implicit information. This limitation hampers the enhancement of implicit information consistency across cameras, especially when data availability is limited. In contrast, our approach addresses this challenge by focusing primarily on the implicit CVD risk information embedded within the images. As a result, our method yields superior performance and practical utility in CVD risk prediction.

However, it is important to note that our primary focus was on achieving consistency in CVD risk prediction and implicitly enhancing domain translation. We did not extensively explore the direct impact of our approach on the retinal images themselves. In future research, we aim to delve deeper into the interpretability of our method. Additionally, we plan to investigate how camera adaptation can address the generalization issues of other deep learning methods related to retinal images, such as specific lesion adaptation for tasks like diabetic retinopathy classification and age-related macular degeneration detection. This will provide a deeper understanding of our methodology and contribute to further advancements in the field.

5. Conclusion

In this study, we present a novel cross-camera domain adaptation method to enhance the consistency of cardiovascular disease (CVD) risk assessment using retinal images captured by conventional and portable cameras. Our approach

incorporates a risk consistency loss based on the enhanced ordinal CVD risk classification model, seamlessly integrating it into any image translation model and ensuring reliable CVD risk assessment across different cameras. Experimental results demonstrate the effectiveness and superiority of our proposed method.

References

[1] Patrick De Boever, Tijs Louwies, Eline Provost, Luc Int Panis, and Tim S Nawrot. Fundus photography as a convenient tool to study microvascular responses to cardiovascular disease risk factors in epidemiological studies. *JoVE (Journal of Visualized Experiments)*, (92):e51904, 2014. 1

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 3

[3] Federico Fantaguzzi, Andrea Servillo, Riccardo Sacconi, Beatrice Tombolini, Francesco Bandello, and Giuseppe Querques. Comparison of peripheral extension, acquisition time, and image chromaticity of optos, clarus, and eidon systems. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 261(5):1289–1297, 2023. 1

[4] Shuang He, Gabriella Bulloch, Liangxin Zhang, Yiyu Xie, Weiyu Wu, Yahong He, Wei Meng, Danli Shi, and Mingguang He. Cross-camera performance of deep learning algorithms to diagnose common ophthalmic diseases: A comparative study highlighting feasibility to portable fundus camera use. *Current Eye Research*, pages 1–7, 2023. 1

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 3

[6] Stephen Kaptoge, Lisa Pennells, Dirk De Bacquer, Marie Therese Cooney, Maryam Kavousi, Gretchen Stevens, Leanne Margaret Riley, Stefan Savin, Taskeen Khan, Servet Altay, et al. World health organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global Health*, 7(10):e1332–e1345, 2019. 1, 4

[7] Haijun Lei, Weixin Liu, Hai Xie, Benjian Zhao, Guanghui Yue, and Baiying Lei. Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(1):90–102, 2021. 1

[8] Peng Liu, Bin Kong, Zhongyu Li, Shaoting Zhang, and Ruogu Fang. Cfea: Collaborative feature ensembling adaptation for domain adaptation in unsupervised optic disc and cup segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22*, pages 521–529. Springer, 2019. 1

[9] Nico JD Nagelkerke et al. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991. 4

[10] Kevin E O'Grady. Measures of explained variance: Cautions and limitations. *Psychological Bulletin*, 92(3):766, 1982. 4

[11] Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature biomedical engineering*, 2(3):158–164, 2018. 1

[12] Jinye Ran, Guanghua Zhang, Fan Xia, Ximei Zhang, Juan Xie, and Hao Zhang. Source-free active domain adaptation for diabetic retinopathy grading based on ultra-wide-field fundus images. *Computers in Biology and Medicine*, page 108418, 2024. 1

[13] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3): e1001779, 2015. 3

[14] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 2, 3, 4, 5

[15] Qijie Wei, Jingyuan Yang, Bo Wang, Jinrui Wang, Jianchun Zhao, Xinyu Zhao, Sheng Yang, Niranchana Manivannan, Youxin Chen, Dayong Ding, et al. Supervised domain adaptation for recognizing retinal diseases from wide-field fundus images. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2298–2303. IEEE, 2023. 1

[16] Jo-Hsuan Wu, TY Alvin Liu, Wan-Ting Hsu, Jennifer Hui-Chun Ho, and Chien-Chang Lee. Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis. *Journal of medical Internet research*, 23(7):e23863, 2021. 1

[17] Dalu Yang, Yehui Yang, Tiantian Huang, Binghong Wu, Lei Wang, and Yanwu Xu. Residual-cyclegan based camera adaptation for robust diabetic retinopathy screening. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, pages 464–474. Springer, 2020. 1

[18] Weiyi Zhang, Zhen Tian, Fan Song, Pusheng Xu, Danli Shi, and Mingguang He. Enhancing stability in cardiovascular disease risk prediction: A deep learning approach leveraging retinal images. *Informatics in Medicine Unlocked*, 42: 101366, 2023. 1

[19] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 2, 3, 4, 5