RDD-Net: Randomized Joint Data-Feature Augmentation and Deep-Shallow Feature Fusion Networks for Automated Diagnosis of Glaucoma

Yilin Tang[1], Min Zhang[1(✉)], and Jun Feng[2]

[1] School of Mathematics, Northwest University, Xi'an 710127, China
[2] School of Information Science&Technology, Northwest University, Xi'an 710127, China
dr.zhangmin@nwu.edu.cn

Abstract. Glaucoma is an irreversible eye disease that has become the leading cause of human blindness worldwide. In recent years, deep learning shows great potential for computer-aided diagnosis in clinics. However, the diversity in medical image quality and acquisition devices leads to distribution shifts that compromise the generalization performance of deep learning methods. To address this issue, many methods relied on deep feature learning combined with the employment of data-level augmentation or feature-level augmentation, respectively. these methods suffer from the limited search space of feature styles. Previous research indicated that introducing a diverse set of augmentations and domain randomization during training can expand the search space of feature styles. In this paper, we propose a Randomized joint Data-feature augmentation and Deep-shallow feature fusion method for automated diagnosis of glaucoma (RDD-Net). It consists of three main components: Data/Feature-level Augmentation (DFA), Explicit/Implicit augmentation (EI), and Deep-Shallow feature fusion (DS). DFA randomly selects data/feature-level augmentation statistics from a uniform distribution. EI involves both explicit augmentation, perturbing the style of the source domain data, and implicit augmentation, utilizing moments information. The randomized selection of different augmentation strategies broadens the diversity of feature styles. DS integrates deep-shallow features within the backbone. Extensive experiments have shown that RDD-Net achieves the SOTA effectiveness and generalization ability. The code is available at https://github.com/TangYilin610/RDD-Net.

Keywords: Glaucoma Diagnosis, Deep Learning, Domain Randomization, The Fundus Color Image.

1 Introduction

Glaucoma is a neurodegenerative disorder characterized by gradual damage to the optic nerve and retinal nerve fiber layers, which results in visual field deficits. Early glaucoma diagnosis and treatment are essential to prevent irreversible vision loss and ultimately blindness [1]. Motivated by the success of deep learning (DL) technology,

computer-aided diagnosis based on advanced convolutional neural networks (CNNs) has drawn extensive attention and achieved excellent performance, particularly for glaucoma disease with the fundus color images [2-10]. Currently, most CNN-based methods follow independent and identically distributed patterns. The acquisition of the fundus color images is inevitably influenced by various factors, including various devices and qualities. These factors compromise the generalization performance [10]. The performance significantly drops when these methods are applied in the unseen domains different from the training data which hinders the clinical practice [11]. It is a common problem known as domain generalization (DG).

Recently, many studies have explored the performance significantly drops in the unseen domains. Common approaches in mainstream research rely on deep feature learning combined with the employment of data-level augmentation or feature-level augmentation [12-17]. For example, DiMix proposed a combination of content-style disentanglement and image synthesis. This method improved performance when applied to unseen target domains [12]. TriD introduced original statistics to the augmented statistics and devised the statistics-randomization strategy to boost the robustness of the method [13]. CDDSA proposed an efficient contrastive domain disentanglement and style augmentation framework for generalizable medical image segmentation [14]. GDRNet learned robust features by simulating visual transformations and image degradations for diabetic retinopathy grading [15]. MoEx systematically regulated the amount of attention a network pays to the signal in feature moments to improve generalization capability [17]. The research on feature augmentation methods mainly focuses on improving feature representation. The data augmentation is typically based on empirical or heuristic methods. Despite their enhanced performance, these methods suffered from the limited search space of feature styles. This limitation may not effectively address the complexities encountered in real-world scenarios. Previous studies indicated that the diversity of feature styles and domain randomization during training can expand the feature search style space substantially [18,19].

Our main contributions are summarized as follows: (1) The proposed RDD-Net enables the method to learn robust features and boost the generalization performance of automated diagnosis of glaucoma. (2) A novel statistics randomization strategy for data augmentation is proposed. This strategy randomly samples the channel of data-feature level statistics from a uniform distribution. We introduce the feature statistics using the random method to the augmented statistics to expand the search space. The random method consists of explicit augmentation of perturbation of the style of the source domain data, and the implicit augmentation of utilizing moments information. (3) The domain-randomized augmented data and features are fed into the backbone network for automated diagnosis of glaucoma. The deep features and shallow features are collaborated in the backbone. The two publicly available benchmarks are used for the evaluations.

2 Methodology

The overview of Randomized joint Data-feature augmentation and Deep-shallow feature fusion networks (RDD-Net) is shown in Fig. 1. It consists of three main components: Data/Feature-level Augmentation (DFA), Explicit/Implicit augmentation (EI), and Deep-Shallow feature fusion (DS). DFA randomly selects data/feature-level augmentation statistics from a uniform distribution. EI involves both explicit augmentation, perturbing the style of the source domain data, and implicit augmentation, utilizing moments information. DS integrates deep-shallow features within the backbone.

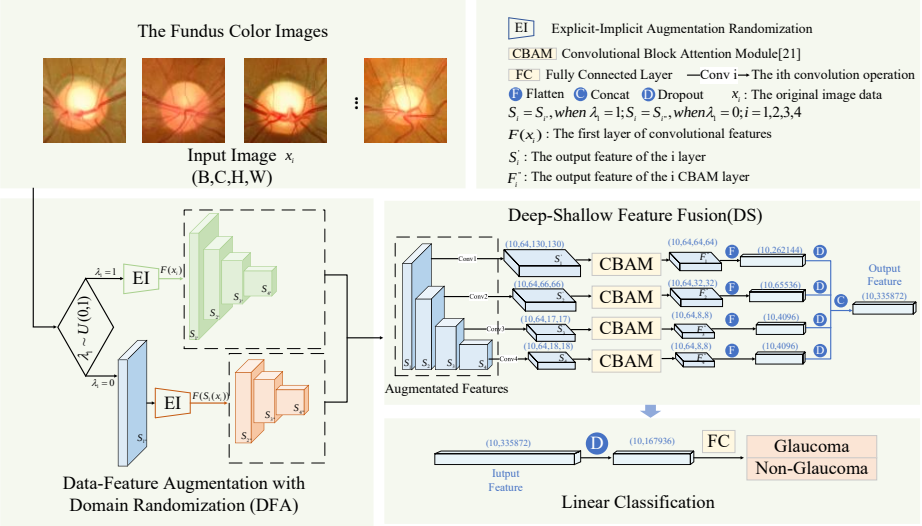

Fig. 1. Randomized joint data-feature augmentation and deep-shallow feature fusion networks. DFA randomly samples the channel of data-feature level statistics from a uniform distribution. EI consists of explicit augmentation and implicit augmentation to expand the search space of feature styles. DS deep features and shallow features are combined in the backbone. The fusion features are passed through a linear classifier to output the diagnostic results for the automated diagnosis of glaucoma.

2.1 Joint Data-Feature Augmentation with Domain Randomization

To enhance the diversity of feature styles. Joint data-feature augmentation with domain randomization is adopted, randomly sampling the **D**ata-level and the **F**eature-level **A**ugmentation statistics from a uniform distribution (see Fig. 1 (DFA)). Let $x \epsilon R^{B\times C\times H\times W}$ be the feature in a mini-batch, where B, C, H, and W respectively denote the mini-batch size, channel, height, and width. We denoted as $x_i \epsilon R^{B\times C}$, x_i represents the input features of the i-th batch in ResNeSt-50. $\lambda_1 \sim U(0,1)$. The output feature $F(x_\mathrm{i})$ is calculated as:

$$F(x_\mathrm{i}) = \lambda_1 R(x_\mathrm{i}) + (1-\lambda_1) R(S_1(x_\mathrm{i})) \tag{1}$$

where $S_1(x)$ is the first convolutional layer in the ResNeSt50 network. $R(x)$ is the output of the augmented feature.

2.2 Explicit-Implicit Augmentation Randomization

The random method consists of **E**xplicit augmentation of perturbation of the style of the source domain data, and the **I**mplicit augmentation of utilizing moments information (see Fig. 2). The explicit augmentation strategy randomly mixes augmented and original statistics across the channel-wise axis. The implicit augmentation strategy utilizes the moment information, which is the moments of the learned features from one training image is replaced with those from another. We randomly sample $\lambda_2 \sim U(0,1)$.

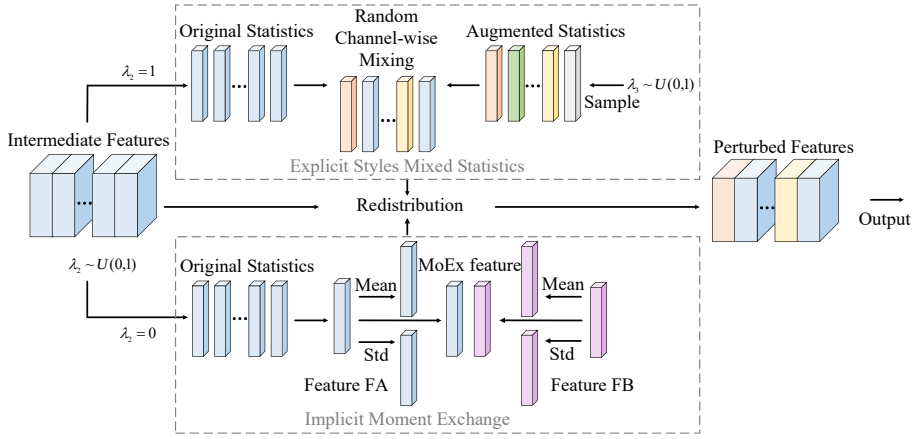

Fig. 2. Explicit-Implicit Augmentation Randomization Network. EI consists of explicit augmentation of perturbation of the style of the source domain data, and the implicit augmentation of utilizing moments information.

The overall process is summarized as:

$$R(y) = \lambda_2 Explicit\ (y) + (1-\lambda_2) Implicit\ (y), \quad \begin{cases} y = x_i, \lambda_1 = 1 \\ y = S_1(x_i), \lambda_1 = 0 \end{cases} \tag{2}$$

where y denotes the intermediate features.

The mixed feature statistics are applied to perturb the normalized x_i. Randomly sample the augmented statistics $\sigma(r), \mu(r) \in \mathrm{R}^{\mathrm{B}\times\mathrm{C}}$ from a uniform distribution: $\sigma(r) \sim U(0,1)$, $\mu(r) \sim U(0,1)$. Next, sample $P \in R^{B\times C}$ from the Beta distribution: $P \sim Beta(\alpha, \alpha)$. Utilize P as the probability to generate the Bernoulli distribution, from which $\lambda_i \in R^{B\times C}$ is sampled: $\lambda_3 \sim Bern(P)$, with α is empirically set to 0.1 [13].

$$Explicit(x_i) = \gamma_{mix} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + \beta_{mix} \tag{3}$$

$$\gamma_{mix} = \lambda_3 \sigma(r) + (1-\lambda_3)\sigma(x_i),\ \beta_{mix} = \lambda_3 \mu(r) + (1-\lambda_3)\mu(x_i) \tag{4}$$

The implicit strategy involves combining features from two images or feature sets by injecting the moments of images or features F_B into the feature representation of F_A:$F_A^{(B)} = \sigma_B \frac{F_A - \mu_A}{\sigma_A} + \mu_B$. Similar to Mixup and CutMix, it fuses features and labels across two training samples.

$$Implicit(x_i) = x_i \times scale + shift \tag{5}$$

$$shift = \mu(x_{index}) - \mu(x_i) \times scale \tag{6}$$

x_i is the input feature, x_{index} is a tensor used for index, $scale = \frac{\sigma(x_{index})}{\sigma(x_i)}$.

2.3 Deep-Shallow Feature Fusion Method

The domain-randomized augmented data and features are fed into the backbone network for automated diagnosis of glaucoma. We use ResNeSt50 [20] as the backbone. It can be observed that we not only utilize the deep features from the stage 4 S_4 output feature of ResNeSt-50 but also integrate shallow and intermediate features from stage 1 S_1, stage 2 S_2, and stage 3 S_3 (see Fig. 1 (DS)). We adjust the number of feature channels in each layer uniformly using convolutional operations and output the modified features S_i'. The output features F_i'' from the stage1, stage2, and stage3 of ResNeSt50 are fed into Convolutional Block Attention Module (CBAM) [21], followed by flatten activation. Then the features from different layers are fused with concatenation. The overall attention process can be summarized as follows:

$$F_i' = M_c(S_i') \otimes S_i' \tag{7}$$

$$F_i'' = M_s(F_i') \otimes F_i' \tag{8}$$

where $\otimes$ denotes element-wise multiplication. F_i'' is the final refined output. M_c is the channel attention. M_s is the spatial attention.

3 Experiments

3.1 Implementation Details

To evaluate the validity of our framework in addressing the challenge of low generalization performance attributed to distributional inconsistency, we devise two distinct sets of experiments.

In the first experiment, we utilize the Refuge2 public dataset [22] obtained from the acquisition of four fundus color image devices. Its train and validation subsets served as the source domains, while the partitioned test set functioned as the unseen domain, denoted as the target domain. In the second experiment, we assess four public datasets, Refuge2 [22], Harvard [23], ORIGA [24], and RIMONE [25], which contain 2000, 1544, 650, 783 samples, using a leave-one-domain approach to simulate real-world generalization challenges objectively and concisely.

We set the mini-batch size to 8 and adopted the SGD optimizer with a momentum of 0.99 for both tasks. The initial learning rate l_0 is set to 0.0001 and decays according to the polynomial rule $l_t = l_0 \times (1 - \frac{t}{T})^{0.9}$, where l_t is the learning rate at epoch 100 and T is the total number of epochs set to 100 for classification of glaucoma. For the task, the leave-one-domain-out strategy was used to evaluate the performance of each DG method, i.e., training on K-1 source domains and evaluating the left domain.

3.2 Comparison with the State-of-the-Arts

Result on Refuge2 Dataset. We compared it with other state-of-the-arts methods from glaucoma automatic diagnosis in Table 1.

Table 1. Comparison with state-of-the-arts glaucoma diagnosis methods. The best results are highlighted in **bold**.

Method	ACC (%)	AUC (%)
ConViT [26]	80.45	82.87
Swin [27]	81.95	82.32
DENet [28]	80.04	84.70
AGCNN [29]	81.20	82.16
ColNet [30]	79.69	85.36
MagNet [31]	83.20	77.52
CMSNET [32]	64.16	80.86
L2T-KT [33]	80.20	86.24
SeATrans [34]	86.96	88.47
G-RISK [35]	-	86.70
DS	88.00	89.19
RDD-Net	**93.88**	**94.90**

We compared the DS method with state-of-the-art diagnostic techniques. It has shown a significant improvement. The integration of DFA and EI strategies led to a substantial augmentation in performance. Consequently, we combined DFA and EI strategies with the DS approach to enhance the generalizability of glaucoma diagnosis methods.

Results on Cross-Dataset. We employed the same classification network and loss function to compare our RDD-Net with seven DG methods. DeepAll is the implementation setup where methods are trained on all source domain data and tested on unseen domains. By employing comparative experiments, our objective is to investigate the generalization ability and performance of these methods across different domains. The results are shown in Table 2.

Our method demonstrated the state-of-the-art performance in the cross-domain dataset experiments, as evidenced by the comparison with advanced experiments detailed in Table 2. These findings underscore the exceptional generalization capabilities of our method across diverse domains.

Table 2. Performance of DeepAll, our RDD-Net, and seven DG methods for automated diagnosis of glaucoma. The best results are highlighted in **bold**.

Method	Refuge2	Harvard	ORIGA	RIMONE	Average
DeepAll	88.70	76.88	74.15	80.00	79.93
Mame [36]	90.15	79.99	74.30	77.11	80.39
GDRNet [15]	88.25	80.63	65.23	85.57	79.92
Trid [13]	90.35	79.79	74.15	81.86	81.54
Mix [37]	89.40	79.14	66.31	81.86	79.18
CAB [38]	87.45	58.42	59.50	83.30	72.17
DRGen [39]	81.35	74.29	70.00	**87.21**	78.21
Fishr [40]	87.50	78.17	**77.69**	86.60	82.49
RDD-Net	**92.80**	**82.25**	75.85	84.54	**83.86**

Ablation Study of Proposed Components. To evaluate the effectiveness of our proposed components, we conducted an extensive ablation study under the cross-dataset and presented the ACC score achieved by different methods. The validity of the modules within the network is shown in Table 3.

Table 3. Results of cross-domain dataset ablation experiments (ACC(%)). The best results are highlighted in **bold**.

DS	DFA		EI		Refuge2	Harvard	ORIGA	RIMONE	Average
	Data	Feature	E	I					
-	-	-	-	-	90.35	81.93	56.92	82.27	77.87
√	-	-	-	-	91.85	**83.03**	59.07	84.12	79.52
√	-	√	√	-	92.10	81.54	59.46	85.36	79.62
√	√	-	√	-	92.20	82.25	66.62	**87.63**	82.18
√	√	√	√	-	92.40	82.57	71.12	84.33	82.61
√	√	√	√	√	**92.80**	82.25	**75.85**	84.54	**83.86**

Through a comparative analysis of experimental results with different combinations, it is evident that the simultaneous integration of DS, DFA, and EI shows the most effective outcome, achieving an accuracy of 83.86% in the proposed method. Overall, we indicated that each component in the proposed method for automatic glaucoma

diagnosis contributed positively to the method's performance, and their synergy in combination produced superior results.

Cross-Dataset Feature t-SNE Visualization Plot. To visualize the distribution of data post-enhancement as proposed in this paper, we employ the T-SNE technique to present the augmentation feature (see Fig. 3).

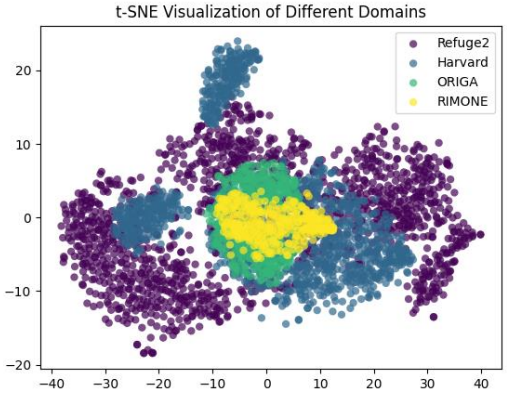

Fig. 3. The scatter plot of the statistics produced by our data augmentation strategy using 2D t-SNE on four datasets.

Fig.3 is the scatter plot of the statistics produced by our data augmentation strategy using 2D t-SNE on four datasets. Different colors denote the image statistic features from different datasets. From Fig.3, Refuge2 and Harvard produce better points separation. Meanwhile, as is shown in Tab.2, the ACC of Refuge2 and Harvard are improved significantly compared with the SOTAs. It indicates that broadening the feature search space is vital to improve the generalization.

4 Conclusion

In this paper, we proposed the RDD-Net to expand the search space of feature styles for automated diagnosis of glaucoma. RDD-Net consists of three main components: Data/Feature-level Augmentation (DFA), Explicit/Implicit augmentation (EI), and Deep-Shallow feature fusion (DS). DFA randomly selects data/feature-level augmentation statistics from a uniform distribution. EI involves both explicit augmentation, perturbing the style of the source domain data, and implicit augmentation, utilizing moments information. The randomized selection of different augmentation strategies broadens the diversity of feature styles. DS integrates deep-shallow features within the backbone. Extensive experiments have shown that RDD-Net achieves the SOTA effectiveness and generalization ability.

Acknowledgment. We would like to thank Dr. Xiao Zhang, Dr. Ying Huang, Dr. Feihong Liu and Dr. Hansheng Li for their valuable comments. This work was financially supported by the Key R&D Plan of Shaanxi Province, China (No.2023-YBSF-304).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Gaibullaeva, N. N.: The role of clinical examination in early diagnosis of glaucoma. In: Health and Medical Sciences **4**(3), 333–337 (2021)
2. Deperlioglu, O., et al.: Explainable framework for glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. In: Future Generation Computer Systems **129**, 152–169 (2022)
3. Vali, M., et al.: Differentiating glaucomatous optic neuropathy from non-glaucomatous optic neuropathies using deep learning algorithms. In: American Journal of Ophthalmology, **252**, 1–8 (2023)
4. Ghorui, A., et al.: Deployment of CNN on color fundus images for the automatic detection of glaucoma. In: International Journal of Applied Science and Engineering **20**(1), 1–9 (2023)
5. Velpula, V. K., et al.: Automatic glaucoma detection from fundus images using deep convolutional neural networks and exploring networks behaviour using visualization techniques. In: SN Computer Science **4**(5), 487 (2023)
6. Shoukat, A., et al.: Automatic diagnosis of glaucoma from retinal images using deep learning approach. In: Diagnostics **13**(10), 1738 (2023)
7. Guo, J. M., et al.: A study of the interpretability of fundus analysis with deep learning-based approaches for glaucoma assessment. In: Electronics **12**(9), 2013 (2023)
8. Saha, S., et al.: A fast and fully automated system for glaucoma detection using color fundus photographs. In: Scientific Reports13(1), 18408(2023)
9. Thanki, R.: A deep neural network and machine learning approach for retinal fundus image classification. In: Healthcare Analytics **3**, 100140 (2023)
10. Li, T., et al.: Applications of deep learning in fundus images: A review. In: Medical Image Analysis **69**, 101971 (2021)
11. Zhou, K., et al.: Domain generalization: A survey. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
12. Kim, H., et al.: DiMix: Disentangle-and-Mix based domain generalizable medical image segmentation. In: Greenspan, H., et al. (eds.) Medical Image Computing and Computer Assisted Intervention–MICCAI 2023, LNCS, vol. 14222, pp. 242–251. Springer, Cham (2023)
13. Chen, Z., et al.: Treasure in Distribution: A domain randomization based multi-source domain generalization for 2D medical image segmentation. In: Greenspan, H., et al. (eds.) Medical Image Computing and Computer Assisted Intervention–MICCAI 2023, LNCS, vol. 14223, pp. 89–99. Springer, Cham (2023)
14. Ran, G., et al.: CDDSA: Contrastive domain disentanglement and style augmentation for generalizable medical image segmentation. In: Medical Image Analysis **89**, 102904 (2023)
15. Che, H., et al.: Towards generalizable diabetic retinopathy grading in unseen domains. In: Greenspan, H., et al. (eds.) Medical Image Computing and Computer Assisted Intervention–MICCAI 2023, LNCS, vol. 14224, pp. 430–440. Springer, Cham (2023)

16. Galappaththige, C. J., et al.: Generalizing to unseen domains in diabetic retinopathy classification. In: Winter Conference on Applications of Computer Vision 2023, pp. 7685–7695 (2023)
17. Li, B., et al.: On feature normalization and data augmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021, pp. 12378–12387 (2021)
18. Gokhale T, et al.: Improving diversity with adversarially learned transformations for domain generalization. In: Winter Conference on Applications of Computer Vision 2023, pp. 434–443 (2023)
19. Lu, J., et al.: Multi-feature fusion for enhancing image similarity learning. In: IEEE Access **7**, 167547–167556 (2019)
20. Zhang, H., et al.: Resnest: Split-attention networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022, pp. 2736–2746 (2022)
21. Woo, S., et al.: Cbam: Convolutional block attention module. In: European Conference on Computer Vision 2018, pp. 3–19 (2018)
22. Fang, H., et al.: REFUGE2 Challenge: A treasure trove for multi-dimension analysis and evaluation in glaucoma screening. In: arXiv preprint arXiv:2202.08994 (2022)
23. Ahn, J. M., et al.: A deep learning model for the detection of both advanced and early glaucoma using fundus photography. In: PloS one **13**(11), e0207982 (2018)
24. Zhuo, Z., et al.: Origa-light: An online retinal fundus image database for glaucoma analysis and research, In: IEEE Eng. in Med. and Bio. Soc. pp. 3065–3068 (2010)
25. Fumero, F., et al.: RIM-ONE: An open retinal image database for optic nerve evaluation, In: International symposium on computer-based medical systems–CBMS, pp. 1–6 (2011)
26. d'Ascoli, S., et al.: Convit: Improving vision transformers with soft convolutional inductive biases. In: International Conference on Machine Learning, pp. 2286–2296 (2021)
27. Liu, Z., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: IEEE/CVF International Conference on Computer Vision 2021, pp. 10012–10022 (2021)
28. Fu, H., et al.: Disc-aware ensemble network for glaucoma screening from fundus image. In: IEEE transactions on medical imaging **37**(11), 2493–2501 (2018)
29. Li, L., et al.: Attention based glaucoma detection: A large-scale database and CNN model. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, pp. 10571–10580 (2019)
30. Yi, Z., et al.: Collaborative learning of semi-supervised segmentation and classification for medical images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019, pp. 2079–2088 (2019)
31. Gupta, S., et al.: Mag-net: Multi-task attention guided network for brain tumor segmentation and classification. In: Srirama, S.N., Lin, J.CW., Bhatnagar, R., Agarwal, S., Reddy, P.K. (eds) International Conference on Big Data Analytics 2021, LNCS, vol. 13147, pp. 3–15. Springer, Cham (2021)
32. Zhou, Y., et al.: Multitask learning for segmentation and classification of tumors in 3d automated breast ultrasound images. In: Medical Image Analysis **70**, 101918 (2021)
33. Wu, J., et al.: Leveraging undiagnosed data for glaucoma classification with teacher-student learning. In: Martel, A.L., et al. (eds) Medical Image Computing and Computer-Assisted Intervention–MICCAI 2020, LNCS, vol. 12261, pp. 731–740. Springer, Cham (2020)
34. Wu, J., et al.: SeATrans: Learning segmentation-assisted diagnosis model via transformer. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds) Medical Image Computing and Computer Assisted Intervention–MICCAI 2022, LNCS, vol. 13432, pp. 677-687 Springer, Cham (2022)
35. Hemelings, R., et al.: A generalizable deep learning regression model for automated glaucoma screening from fundus images. In: NPJ Digital Medicine **6**(1), 112 (2023)

36. Li, C., et al.: Domain generalization on medical imaging classification using episodic train-ing with task augmentation. In: Computers in Biology and Medicine **141**, 105144 (2022)
37. Zhou, K., et al.: Domain generalization with MixStyle. In: International Conference on Learning Representations–ICLR 2021 (2021)
38. He, A., et al.: CabNet: category attention block for imbalanced diabetic retinopathy grading. In: IEEE Trans. Med. Imaging **40**(1), 143153 (2020)
39. Atwany, M., et al.: DRGen: domain generalization in diabetic retinopathy classification. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds) Medical Image Computing and Computer-Assisted Intervention–MICCAI 2022, LNCS, vol. 13432, pp. 635–644. Springer, Cham (2022)
40. Rame, A., et al.: Fishr: Invariant gradient variances for out-of-distribution generalization. In: International Conference on Machine Learning, pp. 18347–18377 (2022)