DR Detection Using Optical Coherence Tomography Angiography (OCTA): A Transfer Learning Approach with Robustness Analysis

DR detection using Optical Coherence Tomography Angiography (OCTA): a transfer learning approach with robustness analysis

Rayna Andreeva *[1], Alessandro Fontanella *[1], Ylenia Giarratano[2], and Miguel O. Bernabeu[2]

[1] School of Informatics, University of Edinburgh, Edinburgh, UK
[2] Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK

Abstract. OCTA imaging is an emerging modality for the discovery of retinal biomarkers in systemic disease. Several studies have already shown the potential of deep learning algorithms in the medical domain. However, they generally require large amount of manually graded images which may not always be available. In our study, we aim to investigate whether transfer learning can help in identifying patient status from a relatively small dataset. Additionally, we explore if data augmentation may help in improving our classification accuracy. Finally, for the first time, we propose a validation of our model on OCTA images acquired with a different device. OCTA scans from three different groups of participants were analysed: diabetic with and without retinopathy (DR and NoDR, respectively) and healthy subjects. We used the convolutional neural network architecture VGG16 and achieved 83.29% accuracy when classifying DR, NoDR and Controls. Our results demonstrate how transfer learning enables fairly accurate OCTA scan classification and augmentation based on geometric transformations helps in improving the classification accuracy further. Finally, we show how our model maintains consistent performance across OCTA imaging devices, without any re-training.

Keywords: Optical coherence tomography angiography · Transfer learning · OCTA devices · Diabetic retinopathy

1 Introduction

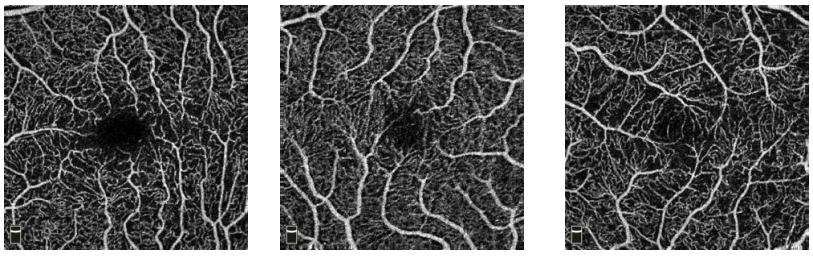
There is an estimated number of more than 20 million people in the UK suffering from at least one long-term condition [19]. If the current trend continues, this figure is projected to increase by more than 20% in the next 5 years [5]. The result of the surge would be to further aggravate the economic, social and human burden on the National Health Service (NHS) [19]. Some of these long-term conditions could be mitigated if early detection was in place to encourage prevention methods such as changes in lifestyle and diet. Hence, there is a need for

* These authors contributed equally

the discovery of biomarkers in the early stages of disease and therefore the latest Artificial Intelligence (AI) technologies could be utilised for identification of people who can benefit from preventative therapies, improve patients outcomes and reduce costs if implemented in clinical practice.

A potential source for disease biomarkers are the changes in microvasculature which have been linked to multiple pathological conditions like diabetes, chronic kidney disease (CKD) and Alzheimer's disease. In particular, the only place in the human body where it is possible to observe the blood vessels in a non-invasive manner with a simple instrument is the retina [19]. A number of recent studies have identified the potential of retinal imaging as a tool for early detection of systemic disease [28]. Indeed, the changes in retinal microvasculature which can be detected on the scans are indicators not only of eye disease, but also for disease of the body. Diabetic retinopathy (DR), neurodegenerative disease, cardiovascular disease and CKD are some of the diseases which have been found to leave a footprint on the retina, often prior to the development of clinically identifiable symptoms [19, 3, 10]. Hence, focusing on the information provided by structural changes in retinal blood vessels can be useful for early diagnosis and better medical treatment. In this work we focus on DR, a diabetes complication that may cause vision loss to the patient.

Optical coherence tomography angiography (OCTA) has emerged recently in the retina imaging domain with the advantage of being a non-invasive and rapid imaging modality. It provides *in vivo* scans of multiple layers of the retina and an insight into the microvasculature by constructing a map of the blood flow. Quantifiable features can be extracted from the OCTA images which are valuable biomarkers for various disease. Studies have identified the usefulness of candidate biomarkers for distinguishing between healthy and DR eyes. Examples include foveal avascular zone (FAZ) area, FAZ contour irregularity [13, 25], vessel caliber, fractal dimension, tortuosity, density and geometric features of the vascular network [1, 14, 2, 22]. Moreover, vessel density has been useful for identifying CKD [27] and both vessel density and perfusion density for Alzheimer's disease [31, 30]. Example of Controls, DR and NoDR images are provided in Figure 1, where the NoDR label refers to diabetic patients without retinopathy.



(a) Control image (b) NoDR image (c) DR image

Fig. 1: OCTA scans from Control, DR and NoDR patients

Deep learning as a subfield of machine learning has shown remarkable results in image classification tasks. The main advantage of deep learning models is that hand-crafted features are not required. In fact, features are extracted automatically in the process, saving time from doing feature engineering and removing the need for identifying disease biomarkers in advance. However, such systems require large amounts of labeled data for training. As a new imaging modality, OCTA datasets are usually small in size. Therefore, an approach known as transfer learning has been adopted, which has been established to have strong performance, especially when dealing with domains with limited data [24, 32, 9].

Machine learning methods have shown promising results in the quest to improve medical evaluations and patient outcomes. However, these models sometimes fail to replicate their results in real world clinical settings [4], where interoperator variability and data quality issues are more common than in highly controlled laboratory environments. In particular, the OCTA technology is based on proprietary algorithms and no standard has emerged yet for image generation, and as such different manufacturers of OCTA devices exploit different algorithms [6]. As a result, the various OCTA devices on the market differ in quality, resolution and size of the images they generate. Moreover, even images collected with the same device may present shifts in their distribution, due to interoperator variability or cooperation of patients during the examination [20, 12]. Hence, a validation of the generalisation ability of the model is required to verify clinical robustness and reliability [21].

In this study we investigate the feasibility of using deep learning for determining patient status in DR, a disease with a known vascular footprint, using a small cross-sectional dataset of OCTA images. First, we investigated whether transfer learning can address the issue related to the limited size of our dataset and achieve competitive performance in disease classification. Secondly, we explored if geometric transformations improve classification performance. Thirdly, we validated our model on a dataset from a different OCTA device to test its robustness. Novel contributions of this study are the ability to independently classify diabetic eyes with and without retinopathy and the validation of the consistency in the classification accuracy of our model on a dataset composed of OCTA images collected with a different imaging device.

2 Methods

2.1 Datasets and imaging devices

The first dataset in the study, NHS Lothian, consists of three groups: diabetic with and without retinopathy (13 DR an 13 NoDR, respectively) and 31 age- and gender-matched healthy subjects (Controls). From each patient we considered both left and right eye whenever available, in order to increase the size of the dataset. Therefore, a total of 51 images from diabetic participants (26 DR and 25 NoDR) and 56 Control images were analysed. The scans were captured by a commercial Optovue RTVue XR Avanti system (Optovue Inc., Fremont, CA).

In this study, only images of the superficial layer with 3×3 field of view (FOV) and 304×304 pixel resolution were analysed.

The second dataset used is a publicly available dataset of OCTA images, OCTAGON [8]. The dataset consists of 144 healthy and 69 diabetic OCTA images, captured using the DRI OCT Triton system (Topcon Corp, Tokyo, Japan). As in the previous analysis we used images of 3×3 FOV superficial depth level with 320×320 pixel resolution of 36 controls and 19 diabetic subjects. Scans of both eyes of the patients were used whenever they were available.

2.2 Data augmentation and transfer learning

Several studies have shown the effectiveness of deep learning architectures on imaging tasks. However, these networks usually require a large amount of labeled data to avoid overfitting. In our work, we are dealing with particularly small datasets. A possible solution to this problem is data augmentation [26]. Several augmentation techniques are possible, but in our work we focused on geometric transformations on the input space. For DR classification, we selected zoom in the range $[0.8, 1.2]$ and rotations up to $40°$ as the most effective transformations. Furthermore, we performed online data augmentation, meaning that each training batch is augmented at every epoch during training, removing the constraints on the memory requirements. Performances are then computed by averaging results from a 5-fold cross-validation. To tackle the issues related to the limited size of our datasets we used a transfer learning approach: considering a convolutional neural network trained with a bottom-up approach, after pre-training with ImageNet we kept the weights of the bottom layers and re-trained only the last convolutional layers to achieve faster learning. In particular, as in [15], we fine-tuned the last 7 convolutional layers. As required by VGG16 input size, images were resized to 224×224 pixels. RmsProp optimizer with a starting learning rate of 1×10^{-5} was used to train the model for 200 epochs. Throughout the study, the metrics used to evaluate the classification performance of the models are accuracy, sensitivity, specificity and area under the Receiver operating characteristic (ROC) curve.

3 Results

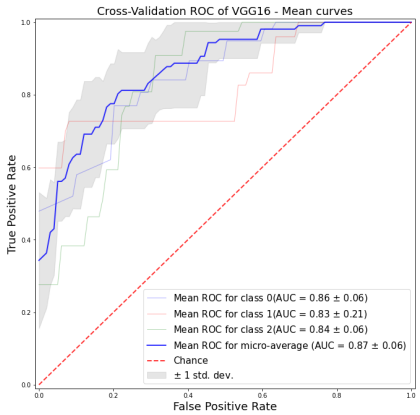
3.1 Classification of Controls, DR and NoDR patients

Evaluation metrics of the model (with and without augmentation) classifying Controls, DR and NoDR subjects on NHS Lothian are reported in Table 1.

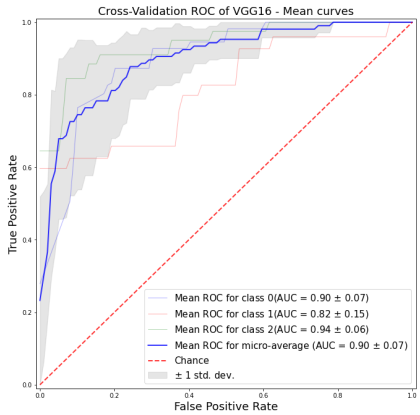
Table 1: Table of classification performances in the DR study with standard error

	Controls		DR		NoDR	
	Without augm	**With augm**	**Without augm**	**With augm**	**Without augm**	**With augm**
Acc %	75.75 ± 2.30	83.03 ± 4.45	88.74 ± 2.20	85.89 ± 3.59	77.58 ± 4.66	85.89 ± 3.04
Sen %	89.70 ± 5.46	92.73 ± 3.04	60.76 ± 14.27	59.62 ± 13.88	32.67 ± 3.00	58.50 ± 11.17
Spe %	60.73 ± 2.90	72.36 ± 6.67	95.39 ± 1.82	92.04 ± 3.59	91.37 ± 3.94	96.53 ± 2.02

The average ROC curves obtained with and without data augmentation are showed in Figure 2. Data augmentation helps in improving the average classification accuracy from 78.38% to 83.29%. The effectiveness of transfer learning is then verified by comparing the model with a new one with the same CNN architecture as VGG16, but with random initialisation of the weights. The latter achieves only 50.00% accuracy.



(a) Average ROC curve obtained without data augmentation



(b) Average ROC curve obtained with data augmentation

Fig. 2: Average ROC curves obtained when classifying Controls, DR and NoDR patients

3.2 Model validation on OCTAGON dataset

The generalisation ability of the model in classifying OCTA scans from different devices is tested using the OCTAGON dataset with the following two labels: diabetic and Control. For this reason, we classified diabetic patients, combining DR and NoDR in the same class. We repeated the previous analysis on NHS Lothian

dataset by combining DR and NoDR in the same class, achieving 87.33% and 84.24% classification accuracy, with and without data augmentation respectively. Using the model without data augmentation on the OCTAGON dataset, the accuracy drops to 83.64% (Table 2a). Moreover, we test the model pre-trained on ImageNet and fine-tuned on OCTAGON without data augmentation (Table 3a).

Table 2: Classification statistics (with standard error) obtained classifying Controls vs Diabetes on OCTAGON dataset, using VGG16 pre-trained on ImageNet and fine-tuned on NHS Lothian

	Diabetes
Overall Acc %	83.64 ± 4.74
Sen %	77.50 ± 7.75
Spe %	95.00 ± 4.47

(a) Model fine-tuned on NHS Lothian without data augmentation

	Diabetes
Overall Acc %	63.64 ± 5.75
Sen %	43.57 ± 9.50
Spe %	100.00 ± 0.00

(b) Model fine-tuned on NHS Lothian with data augmentation

Table 3: Classification statistics (with standard error) obtained classifying Controls vs Diabetes on OCTAGON dataset, using VGG16 pre-trained on ImageNet and fine-tuned on OCTAGON

	Diabetes
Overall Acc %	87.27 ± 4.15
Sen %	75.00 ± 14.14
Spe %	94.64 ± 2.95

(a) Model fine-tuned on OCTAGON without data augmentation

	Diabetes
Overall Acc %	90.91 ± 3.64
Sen %	80.00 ± 10.95
Spe %	97.5 ± 2.24

(b) Model fine-tuned on OCTAGON with data augmentation

The model trained on NHS Lothian with data augmentation achieves the classification statistics displayed in Table 2b when classifying Controls vs Diabetes on OCTAGON images. Interestingly, we can observe how in this case the accuracy drops significantly, from 87.33% achieved when classifying Controls vs Diabetes on NHS Lothian, to 63.64% when performing the same classification task on OCTAGON data. A possible explanation for this could be that data augmentation may push the images of NHS Lothian even further from the distribution of images in OCTAGON and thus worsen the classification accuracy on the latter dataset.

On the other hand, the model only pre-trained on ImageNet and fine-tuned on OCTAGON with data augmentation, achieves an accuracy of 90.91%($\pm$3.64%), as showed in Table 3b. In this case, data augmentation helps in improving the

performance since test images are from the same distribution of the training set, on which the model was fine-tuned.

4 Discussion and conclusions

In the current study, we investigated the non-invasive detection of DR in OCTA retinal scans using deep learning. In order to address the limited size of our datasets, we employed a transfer learning approach. We also verified how our model can be successfully applied to a different dataset - OCTAGON. In particular, we achieved 83.29%($\pm$4.31%) accuracy when classifying DR, NoDR and Controls with data augmentation and pre-training from ImageNet. On the other hand, if we start to train our model from scratch, giving a random initialisation to the weights, classification accuracy decreases significantly. For this reason, we can confidently assert that transfer learning plays a critical role in achieving a satisfactory classification performance.

Our novel contributions were the ability to independently classify diabetic eyes with and without retinopathy and the investigation of the consistency in the classification accuracy of our model on a dataset composed of OCTA images collected with a different imaging device. In particular, we verified how our model pre-trained on ImageNet and fine-tuned on NHS Lothian is able to achieve satisfactory performance in classifying OCTAGON images without re-training. Model robustness is a fundamental aspect when deploying AI screening tools to critical settings such as predictive healthcare, where it can essentially be life-critical [21].

In [15], the authors used transfer learning with VGG16 architecture to detect DR from OCTA images. They reported an accuracy of 87.28%, using a dataset of 131 OCTA images, thus achieving a slightly higher accuracy, but with a bigger dataset than in our work. Other authors have classified DR using fundus images. They achieved classification accuracy comparable to our study, but using considerably larger datasets. This imaging modality is usually not able to reveal subtle abnormalities correlated with early DR [15], as we were able when independently classifying diabetic eyes with and without retinopathy. In particular, Sayres et al. reported a 88.4% accuracy in DR classification on a dataset of 1,796 retinal fundus images from 1,612 diabetic patients using Inception-V4 architecture [23]. In [16], the authors achieved a high classification accuracy of 93.49% using Inception-V3 architecture. They had available 8,816 fundus images from 5,278 patients. Lin et al. used a CNN with 4 convolutional layers and obtained 86.10% accuracy with a datset of 21,123 fundus photographs [18]. In [17], the authors used a cross-disease attention network to grade both DR and diabetic macular edema (DME) by exploring the relationship between the two diseases. Their method, trained on the 1200 fundus images of Messidor [7] dataset, achieved 92.6% average accuracy over ten folds. Wang et al. [29] employed a network called Zoom-in-Net to generate attention maps highlighting suspicious regions and Detecting DR. They achieved 90.5 classifying referable/nonreferable DR on Messidor. Efforts in the automated classification of DR on the basis of researched

biomarkers have been suggested in [1], where the authors extracted six quantitative features from the images and used them to train a support vector machine (SVM) in order to detect DR. In general, statistical learning methods rely on manual image segmentation, which lacks consistency [11] and could lead to errors, to perform feature extraction. On the other hand, deep learning methods have the advantage that they can directly process the raw images as input.

From the satisfactory results obtained when applying our model on a different dataset, we can argue that a deep learning system for automatic detection of DR, applicable to images collected with any OCTA device, can be achieved. Limitations of our work are the use of a modest dataset size and the inclusion of both left and right eye from the same participant as independent samples. Future works will validate our procedure on larger cohorts and will account for possible correlations between eyes.

In summary, we were able to verify how transfer learning techniques are useful to tackle the issue related to the limited size of OCTA datasets and achieve satisfactory performance when detecting DR and NoDR, how geometric data augmentation helps in improving the performance further and how our approach maintains consistent performance across different OCTA devices.

5 Acknowledgements

RA and AF are supported by the United Kingdom Research and Innovation (grant EP/S02431X/1), UKRI Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics. YG is supported by the Medical Research Council (MRC). MOB is supported by grants from EPSRC (EP/R029598/1, EP/R021600/1, EP/T008806/1), Fondation Leducq (17 CVD 03), and the European Union's Horizon 2020 research and innovation programme under grant agreement No 801423.

References

1. Alam, M., Zhang, Y., Lim, J.I., Chan, R.V., Yang, M., Yao, X.: Quantitative optical coherence tomography angiography features for objective classification and staging of diabetic retinopathy. Retina **40**(2), 322–332 (2020)
2. Alam, M.N., Son, T., Toslak, D., Lim, J.I., Yao, X.: Quantitative artery-vein analysis in optical coherence tomography angiography of diabetic retinopathy. In: Ophthalmic Technologies XXIX. vol. 10858, p. 1085802. International Society for Optics and Photonics (2019)
3. Baker, M.L., Hand, P.J., Wang, J.J., Wong, T.Y.: Retinal signs and stroke: revisiting the link between the eye and brain. Stroke **39**(4), 1371–1379 (2008)
4. Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., Vardoulakis, L.M.: A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–12 (2020)
5. House of Commons Health Committee and others: Managing the care of people with long-term conditions. Second Rep Sess **1**, 1–89 (2014)

6. Cunha-Vaz, J.G., Koh, A.: Imaging techniques **10**, 52–64 (2018). https://doi.org/10.1159/000487412
7. Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., et al.: Feedback on a publicly distributed image database: the messidor database. Image Analysis & Stereology **33**(3), 231–234 (2014)
8. Díaz, M., Novo, J., Cutrín, P., Gómez-Ulla, F., Penedo, M.G., Ortega, M.: Automatic segmentation of the foveal avascular zone in ophthalmological OCT-A images. PloS one **14**(2) (2019)
9. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: International conference on machine learning. pp. 647–655 (2014)
10. Frost, S., Kanagasingam, Y., Sohrabi, H., Vignarajan, J., Bourgeat, P., Salvado, O., Villemagne, V., Rowe, C.C., Macaulay, S.L., Szoeke, C., et al.: Retinal vascular biomarkers for early detection and monitoring of Alzheimer's disease. Translational psychiatry **3**(2), e233–e233 (2013)
11. Giarratano, Y., Bianchi, E., Gray, C., Morris, A., MacGillivray, T., Dhillon, B., Bernabeu, M.O.: Automated and Network Structure Preserving Segmentation of Optical Coherence Tomography Angiograms. arXiv preprint arXiv:1912.09978 (2019)
12. Hong, J.T., Sung, K.R., Cho, J.W., Yun, S.C., Kang, S.Y., Kook, M.S.: Retinal nerve fiber layer measurement variability with spectral domain optical coherence tomography. Korean Journal of Ophthalmology **26**(1), 32–38 (2012)
13. Khadamy, J., Aghdam, K.A., Falavarjani, K.G.: An update on optical coherence tomography angiography in diabetic retinopathy. Journal of ophthalmic & vision research **13**(4), 487 (2018)
14. Le, D., Alam, M., Miao, B.A., Lim, J.I., Yao, X.: Fully automated geometric feature analysis in optical coherence tomography angiography for objective classification of diabetic retinopathy. Biomedical optics express **10**(5), 2493–2503 (2019)
15. Le, D., Alam, M.N., Lim, J.I., Chan, R., Yao, X.: Deep learning for objective OCTA detection of diabetic retinopathy. In: Ophthalmic Technologies XXX. vol. 11218, p. 112181P. International Society for Optics and Photonics (2020)
16. Li, F., Liu, Z., Chen, H., Jiang, M., Zhang, X., Wu, Z.: Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm. Translational vision science & technology **8**(6), 4–4 (2019)
17. Li, X., Hu, X., Yu, L., Zhu, L., Fu, C.W., Heng, P.A.: Canet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. IEEE transactions on medical imaging **39**(5), 1483–1493 (2019)
18. Lin, G.M., Chen, M.J., Yeh, C.H., Lin, Y.Y., Kuo, H.Y., Lin, M.H., Chen, M.C., Lin, S.D., Gao, Y., Ran, A., et al.: Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy. Journal of ophthalmology **2018** (2018)
19. MacGillivray, T., Trucco, E., Cameron, J., Dhillon, B., Houston, J., Van Beek, E.: Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions. The British journal of radiology **87**(1040), 20130832 (2014)
20. Mwanza, J.C., Gendy, M.G., Feuer, W.J., Shi, W., Budenz, D.L.: Effects of changing operators and instruments on time-domain and spectral-domain oct measurements of retinal nerve fiber layer thickness. Ophthalmic Surgery, Lasers and Imaging Retina **42**(4), 328–337 (2011)

21. Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A.: Secure and robust machine learning for healthcare: A survey. arXiv preprint arXiv:2001.08103 (2020)
22. Sasongko, M., Wong, T., Nguyen, T., Cheung, C., Shaw, J., Wang, J.: Retinal vascular tortuosity in persons with diabetes and diabetic retinopathy. Diabetologia **54**(9), 2409–2416 (2011)
23. Sayres, R., Taly, A., Rahimy, E., Blumer, K., Coz, D., Hammel, N., Krause, J., Narayanaswamy, A., Rastegar, Z., Wu, D., et al.: Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. Ophthalmology **126**(4), 552–564 (2019)
24. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 806–813 (2014)
25. Takase, N., Nozaki, M., Kato, A., Ozeki, H., Yoshida, M., Ogura, Y.: Enlargement of foveal avascular zone in diabetic eyes evaluated by en face optical coherence tomography angiography. Retina **35**(11), 2377–2383 (2015)
26. Tanner, M.A., Wong, W.H.: The calculation of posterior distributions by data augmentation. Journal of the American statistical Association **82**(398), 528–540 (1987)
27. Vadalà, M., Castellucci, M., Guarrasi, G., Terrasi, M., La Blasca, T., Mulè, G.: Retinal and choroidal vasculature changes associated with chronic kidney disease. Graefe's Archive for Clinical and Experimental Ophthalmology **257**(8), 1687–1698 (2019)
28. 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)
29. Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H., Wang, X.: Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 267–275. Springer (2017)
30. Yao, X., Alam, M.N., Le, D., Toslak, D.: Quantitative optical coherence tomography angiography: A review. Experimental Biology and Medicine **245**(4), 301–312 (2020). https://doi.org/10.1177/1535370219899893, `https://doi.org/10.1177/1535370219899893`, pMID: 31958986
31. Yoon, S.P., Grewal, D.S., Thompson, A.C., Polascik, B.W., Dunn, C., Burke, J.R., Fekrat, S.: Retinal microvascular and neurodegenerative changes in Alzheimer's disease and mild cognitive impairment compared with control participants. Ophthalmology Retina **3**(6), 489–499 (2019)
32. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in neural information processing systems. pp. 3320–3328 (2014)