WEAKLY-SUPERVISED LOCALIZATION OF DIABETIC RETINOPATHY LESIONS IN RETINAL FUNDUS IMAGES

Waleed M. Gondal[⋆⋄∗†], *Jan M. Köhler*[∗⋆‡], *René Grzeszick* [⋄], *Gernot A. Fink* [⋄] *and Michael Hirsch*[§]

[⋆] Bosch Center for Artificial Intelligence, Robert Bosch GmbH, Stuttgart, Germany
[⋄] Department of Computer Science, TU Dortmund University, Germany
[§]Max Planck Institute for Intelligent Systems, Tübingen, Germany

ABSTRACT

Convolutional neural networks (CNNs) show impressive performance for image classification and detection, extending heavily to the medical image domain. Nevertheless, medical experts are skeptical in these predictions as the nonlinear multilayer structure resulting in a classification outcome is not directly graspable. Recently, approaches have been shown which help the user to understand the discriminative regions within an image which are decisive for the CNN to conclude to a certain class. Although these approaches could help to build trust in the CNNs predictions, they are only slightly shown to work with medical image data which often poses a challenge as the decision for a class relies on different lesion areas scattered around the entire image. Using the DiaretDB1 dataset, we show that on retina images different lesion areas fundamental for diabetic retinopathy are detected on an image level with high accuracy, comparable or exceeding supervised methods. On lesion level, we achieve few false positives with high sensitivity, though, the network is solely trained on image-level labels which do not include information about existing lesions. Classifying between diseased and healthy images, we achieve an AUC of 0.954 on the DiaretDB1.

Index Terms— deep learning, weakly-supervised object localization, lesion detection, diabetic retinopathy.

1. INTRODUCTION

The World Health Organization (WHO) estimates that in 2002 the reason for almost 5 million blind people was diabetic retinopathy (DR), accounting for about 5% of world blindness[1]. The estimated global prevalence of referable DR (RDR) among patients with diabetes is 35.4% [1]. At the same time the prevalence of diabetes among adults has increased from 4.7% in 1980 to 8.5% in 2014 accounting for 422 million people with diabetes [2]. RDR is considered to be the fifth most common cause of moderate to severe visual impairment [3]. Regular retinal screening for people with diabetes is recommended in order to be treated as early as possible before a moderate or severe DR has evolved leading to visual impairment. Lacking qualified personnel in developing countries [4] to assess retinal images, automated grading and detection algorithms have been developed.

While first approaches using neural networks to detect diabetic retinopathy on retinal images without additional feature extraction showed a low classification accuracy [5, 6], recent approaches based on deep neural networks [7, 8, 9] report good performance. For medical experts, these algorithms represent black box approaches as only a classification result but no information to why this conclusion is reached is provided. To overcome this obstacle and build trust in such automated healthcare monitoring systems, lesion areas can be detected and displayed as a basis to judge the rating.

A lot of research has been conducted to detect specific lesion areas, like blood vessel transformations, exudates, microaneurysms and hemorrhages [10, 11, 12, 13, 14]. Winder et al. [15] give an overview of literature from 1998-2008 using digital imaging techniques for DR. These approaches have used automatic image-processing techniques, partly combined with machine learning algorithms. Recently, lesion areas responsible for DR are detected using CNNs [16, 17].

All these approaches have in common that specific lesion categories are detected which lead to DR but they cannot directly be connected to the prediction outcome of a deep learning algorithm. We present a method to localize areas of images which are responsible for a CNN to conclude the DR status. Though, not trained explicitly, it is shown that these areas map with the lesion areas.

2. RELATED WORK

Class-specific saliency detection in CNNs has recently received a lot of attention as it can be useful in numerous deep learning applications, e.g., in autonomous driving, where detecting a person in the scene is as important as determining its exact location in the scene. Methods for saliency map predic-

[∗]Both authors contributed equally to this work.

[†]The work of the paper was performed while the author was an intern at Bosch Center for Artificial Inelligence.

[‡]Corresponding author: jan.koehler@de.bosch.com

[1]www.who.int/blindness/causes/priority

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

tion identify regions which are visibly distinctive. Though, these regions may not necessarily map to areas that are decisive for image classification.

In contrast, weakly-supervised object localization corresponds to highlighting the class-specific discriminative regions which influence certain predictions. Even though CNNs can recognize the class of an object in the image, it is not easy for them to localize the object in the image. Recently proposed approaches [18, 19] visualize the internal representations learned by the inner layers of CNNs in order to understand their properties. In [18], deconvolutional networks are used to visualize the patterns activated by each unit. [19] shows that while being trained to recognize scenes, CNNs learn object detectors. It demonstrates that the same network can perform both scene recognition and object localization in a single forward-pass. In [20], class-specific maps are constructed by identifying the pixels that are most useful to predict the classification score and then back-projecting the corresponding information. Another approach mentioned in [21] tries to identify the regions causing maximal activations while masking different portion of the images.

In [22], the last fully connected layers are treated as convolutions and a max pooling is applied to localize the object. The localization is limited to a point lying in the boundary of the object. Based on the similar approach, [23] proposes class activation maps (CAMs) claiming to identify the complete extent of the object instead of one point. They use global average pooling (GAP) to leverage the linear relation between the softmax predictions and the final convolutional layer, which results in highlighting the most discriminative image regions relevant to the predicted result. A recent comparison of three localization methods is given in [24].

Object localization on retina images poses a challenge as the lesion areas - among others small red dots, microaneurysms, hemorrhages - responsible for diabetic retinopathy are scattered around the image and are often not localized within a few image regions. To the best of our knowledge, [25] is the only approach proposed so far to detect the lesion areas within retina images which is trained in weakly-supervised fashion using only image-level labels to conclude the lesion areas. Using a generalization of backpropagation, an ensemble of CNNs is learned in which each CNN excels in the detection of a certain lesion type.

3. METHOD

This section describes our proposed deep learning approach for localizing discriminative features in DR.

The aim is to learn a representation that enables localization of discriminative features in a retina image while at the same time achieving good classification accuracy on the same. Our proposed CNN architecture is able to highlight decisive features in a single forward pass which facilitates medical diagnosis through visual inspection. Since good class-specific features and high classification accuracy are key, we adopted the award-wining CNN architecture *o_O solution* by Antony and Brüggemann [26].

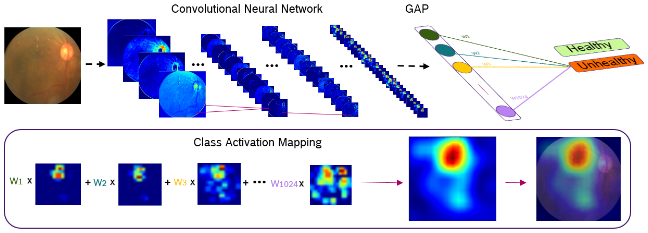

Fig. 1: CNN setup for generating CAMs.

3.1. Localization with Class Activation Maps

The architecture has been designed to achieve good image level classification accuracy in DR. To make it capable of weakly-supervised localization, we modify it to compute CAMs introduced by [23]. The final dense layers are removed from the proposed CNN architecture in order to retain spatial information and replaced by a GAP layer instead. The GAP layer performs average pooling on K feature maps of the last convolutional layer, $A^{\mathrm{k}} \in R^{\mathrm{u\ x\ v}}$ having width u and height v. The resultant spatially pooled values are then fully connected to output classification scores y^{c} via ω_k^c, where c corresponds to the classes.

$$y^c = \sum_k \omega_k^c \sum_x \sum_y A_{xy}^k \tag{1}$$

The weights ω_k learned in the last layer encode the importance of each feature map A^k with respect to the class c. The final localization map L_{CAM} is produced by computing the weighted linear sum of these feature maps

$$L_{CAM} = \sum_k \omega_k^c A^k. \tag{2}$$

The localization map is then upsampled to the size of original input image, highlighting the class-specific image regions. The generation of class activation maps is depicted in Fig. 1.

Fine-tuning of CAM: Most DR lesions are of extremely small size on typical retina images. CAMs perform well in detecting those regions, but the upsampled localization map tends to produce coarse heatmaps rendering a fine-grained resolution impossible. To refine the localization map, as hinted by [23], the spatial resolution of the feature maps A^k from the final convolutional layer is increased. In our network, we remove strides from the first and third convolutional layers, resulting in the feature maps A^k of resolution 32×32 pixels. Moreover, a new convolutional layer of dimension 3×3 pixels and stride one with 1024 kernels is added to the network. These modifications improve the overall localiza-

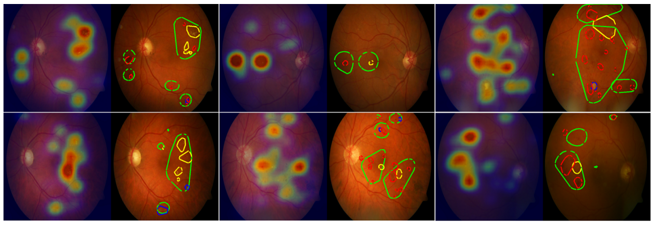

Fig. 2: The weakly-supervised localization results on DiaretDB1 images. In each pair of images, the left image shows the input image overlaid with a corresponding localization heatmap, highlighting RDR affected regions. The right image compares our detection boundary (in green) with the ground truth: yellow, blue and red marked regions represent exudates, red lesions and hemorrhages respectively. Please note that this figure is best viewed on screen rather than on print.

tion ability of the network.

Improving Classification Accuracy: The removal of dense layers from the network leads to a decrease in the overall classification accuracy of the network. We also observed that increasing the spatial resolution of feature maps A^k slows down the training process, significantly. At the same time, the introduction of batch normalization [27] in each convolutional layer during the training process enabled us to achieve faster training convergence with higher learning rates. We also employ regularization within our network to avoid over-fitting and for making our model more generic for RDR recognition and lesion localization. This is helpful since the dataset for the localization task, DiaretDB1, and the dataset used for training, were taken with different appliances [28, 29] (see section 4.1).

3.2. Generation of Region Proposals

As shown in Fig. 2, CAMs generate heatmaps highlighting class-specific discriminative regions. Heatmaps are good for qualitative analysis of the approach. However, for the evaluation of the localization results, well defined region proposals are required. To achieve this, heatmaps are first normalized between 0 and 1, assigning each pixel a value according to its intensity. The high intensity regions are then selected using binary segmentation, giving us the predicted regions P_i for RDR lesion areas, where i enumerates the predicted regions. We empirically found that a threshold value of 0.65 yields good regions. For each P_i obtained, max-pooling is performed to get one score S_i which serves as the confidence measure of prediction for each P_i.

4. EXPERIMENTS

This section describes the datasets, experiments and their detailed comparison with other methods.

4.1. Datasets

Two publicly available datasets, *Kaggle Diabetic Retinopathy Dataset* and *DiaretDB1*, were used for this study. We use the Kaggle dataset for training and evaluate our lesion localization approach on the dataset *DiaretDB1* [28].

Kaggle Dataset: The dataset [30] provided by EyePACS [29] contains 88,702 color fundus images of which 80% were used for training and 20% for validation. For classification, the first two classes of the five DR levels were grouped into non-referable DR (NRDR) and the remaining three classes into RDR. In our experiments, we were more concerned with improving the network's lesion level detections performance than improving the classification accuracy, where people have already achieved remarkable results.

DiaretDB1 Dataset: This dataset is used to validate our lesion level detections. The dataset contains 89 color fundus images, hand-labeled by four experts for four different DR lesion types [28]. As suggested in [28], we only consider those pixels as ground truth whose confidence level of labeling exceeds an average of 75% between experts.

4.2. Implementation

In the kaggle dataset, retina sphere is surrounded by black margins containing no information. These black regions were cropped and the images were resized to 512×512 pixels. All training images were individually standardized by subtracting mean and dividing by standard deviation which were computed over all the pixels in an image. In addition to random brightness and contrast enhancements, the images were randomly rotated, flipped horizontally and vertically in data augmentation performed during training. The network was implemented using Tensorflow and trained on Tesla K80 GPU for 150 epochs. Gradient descent optimizer was used with the momentum of 0.8. L2 regularization was performed on weights with weight decay factor of 0.0005. The initial learning rate was 0.01 which was decayed by 1% after each epoch.

4.3. Evaluation on DiaretDB1 Lesion Detection

We assess performance at both image and lesion level.

4.3.1. Performance at Image Level

Most of the studies done on RDR lesion detection at image level have not mentioned their criteria for selecting true positives (TP) and false negatives. Therefore, for the sake of clarity we evaluated our approach for two scenarios. In the first scenario, an image is considered to be TP for a lesion, if there is a minimum overlap of 50% between P_i and the corresponding lesion's ground truth G_j, where j is the number of ground truth annotations. In the second scenario an overlap of one pixel, whose confidence level is 0.75 or more, between P_i

Table 1: Lesion level performance comparison with different methods.

Method	Hemorrhages		Hard Exudates		Soft Exudates		RSD	
	SE%	**FPs/I**	**SE%**	**FPs/I**	**SE%**	**FPs/I**	**SE%**	**FPs/I**
Quellec *et al.* [25]	71	10	**80**	10	**90**	10	**61**	10
Dai *et al.* [31]	-	-	-	-	-	-	29	20.30
Ours (50% Overlap)	**72**	**2.25**	47	**1.9**	71	**1.45**	21	**2.0**
Ours (OnePixel Overlap)	91	1.5	87	1.5	89	1.5	52	1.5

Table 2: Image level sensitivity in %.

Method	H*	HE*	SE*	RSD*
Zhou *et al.*[32]	94.4	-	-	
Liu *et al.*[33]	-	83.0	**83.0**	-
Haloi *et al.*[34]		**96.5**	-	-
Mane *et al.*[35]	-	-	-	**96.4**
Ours (50% Overlap)	**97.2**	93.3	81.8	50
Ours (OnePixel Overlap)	97.2	100	90.9	50

* H, HE, SE, RSD: Hemorrhages, Hard Exudates, Soft-Exudates and Red Small Dots.

and G_j is considered to be a TP. Although the first criteria is more strict than the second, our method performs similar in both scenarios, ascertaining the precision of our approach.

Our CNN based model is trained to perform binary classification on RDR, achieving 93.6% sensitivity and 97.6% specificity on DiaretDB1 dataset with area under the Receiver Operating Characteristics (ROC) curves of 95.4%. For lesion level detection at image level we only report sensitivity as the pixel-wise comparison of P_i with G_j is possible to confirm the presence of certain RDR finding. However, if the model wrongly classifies a healthy image to be unhealthy, which is clearly a false positive (FP) at image level, it is not possible to relate this FP to any specific RDR lesion type. Thus, we only report the specificity over all lesion types which is 97.6%.

Given that our model is trained in weakly-supervised fashion for classifying RDR, it is remarkable that it performs comparable or even outperforms fully supervised methods for image level lesion detections which are trained specifically for detecting one or two types of lesions. The comparison of sensitivities is given in Tab. 2.

4.3.2. *Performance at Lesion Level*

Free-response Receiver Operating Characteristic (FROC) curves [36] are commonly used for lesions localization evaluation in medical imaging. In our evaluations only those regions G_j which have an overlap of at least 50% with a P_i are considered TP. Sometimes, P_i are way bigger than G_j, possibly covering one or more G_j, therefore, in order to penalize this, mean Intersection over union (mIOU) for each P_i with covered G_j is computed. The P_i is considered FP if its mIOU value is less than 0.5.

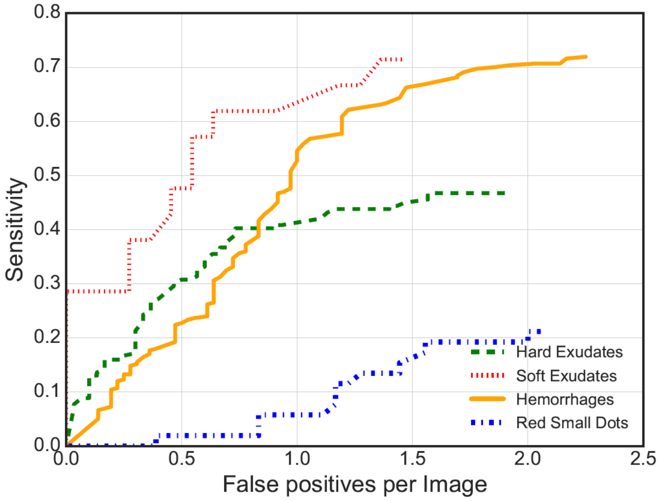

Fig. 3: FROC curves for all four types of DR lesions.

Our network does not perform well in detecting red small dots which are often one or two pixels wide on a 512 × 512 image. We suspect that this could be due to the architecture of CNNs where the information is compressed down the stream for inference, resulting in the loss of information for very small regions. Moreover the resolution of heatmaps is too coarse to highlight these small regions precisely. We compare our localization results with the method from [25] which employed CNN based weakly-supervised localization scheme in detecting RDR lesions. The comparison provided in Tbl. 1 shows that we have fewer FPs than other state of the art methods while achieving comparable results on sensitivity (SE). FROC plots are shown in Fig. 3.

5. CONCLUSION

We presented a deep learning approach that highlights regions on retinal images that are indicative for diabetic retinopathy to assist medical diagnosis. Our architecture is inspired by a recent top-performing supervised CNN architecture for diabetic retinopathy classification but modified to enable weakly supervised object localization. We demonstrate accurate localization with good sensitivity while maintaining high classification accuracy. Along with fast inference we hope that our approach will facilitate diagnostic inspection and be a useful tool for medical professionals.

6. REFERENCES

[1] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, et al., "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes care*, vol. 35, no. 3, pp. 556–564, 2012.

[2] World Health Organization, "Global report on diabetes," 2016.

[3] R. R. Bourne, G. A. Stevens, R. A. White, J. L. Smith, S. R. Flaxman, H. Price, J. B. Jonas, J. Keeffe, J. Leasher, K. Naidoo, et al., "Causes of vision loss worldwide, 1990–2010: a systematic analysis," *The Lancet Global Health*, vol. 1, no. 6, pp. e339–e349, 2013.

[4] S. Resnikoff, W. Felch, T.-M. Gauthier, and B. Spivey, "The number of ophthalmologists in practice and training worldwide: a growing gap despite more than 200 000 practitioners," *British Journal of Ophthalmology*, vol. 96, no. 6, pp. 783–787, 2012.

[5] G. Gardner, D. Keating, T. Williamson, and A. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool.," *British Journal of Ophthalmology*, vol. 80, no. 11, pp. 940–944, 1996.

[6] D. Usher, M. Dumskyj, M. Himaga, T. Williamson, S. Nussey, and J. Boyce, "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening," *Diabetic Medicine*, vol. 21, no. 1, pp. 84–90, 2004.

[7] E. Colas, A. Besse, A. Orgogozo, B. Schmauch, N. Meric, and E. Besse, "Deep learning approach for diabetic retinopathy screening," *Acta Ophthalmologica*, vol. 94, no. S256, 2016.

[8] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Journal of the American Medical Association (JAMA)*, vol. 316, no. 22, pp. 2402–2410, 2016.

[9] R. Arunkumar and P. Karthigaikumar, "Multi-retinal disease classification by reduced deep learning features," *Neural Computing and Applications*, pp. 1–6, 2016.

[10] Y. Hatanaka, T. Nakagawa, Y. Hayashi, Y. Mizukusa, A. Fujita, M. Kakogawa, K. Kawase, T. Hara, and H. Fujita, "CAD scheme for detection of hemorrhages and exudates in ocular fundus images," in *Proceedings SPIE of Medical Imaging*.

[11] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russell, M. Abràmoff, and P. Soliz, "Multiscale AM-FM methods for diabetic retinopathy lesion detection," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 502–512, 2010.

[12] S. Ravishankar, A. Jain, and A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 210–217.

[13] A. Osareh, B. Shadgar, and R. Markham, "A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 535–545, 2009.

[14] Ş. Ţălu, D. M. Călugăru, and C. A. Lupaşcu, "Characterisation of human non-proliferative diabetic retinopathy using the fractal analysis," *International journal of ophthalmology*, vol. 8, no. 4, pp. 770, 2015.

[15] R. J. Winder, P. J. Morrow, I. N. McRitchie, J. Bailie, and P. M. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Computerized Medical Imaging and Graphics*, vol. 33, no. 8, pp. 608–622, 2009.

[16] M. Haloi, "Improved microaneurysm detection using deep neural networks," *preprint arXiv:1505.04424*, 2015.

[17] P. Prentašić and S. Lončarić, "Detection of exudates in fundus photographs using convolutional neural networks," in *Int'l Symposium on Image and Signal Processing and Analysis (ISPA)*. IEEE, 2015, pp. 188–192.

[18] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014, pp. 818–833.

[19] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Object detectors emerge in deep scene cnns," *preprint arXiv:1412.6856*, 2014.

[20] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *preprint arXiv:1312.6034*, 2013.

[21] L. Bazzani, A. Bergamo, D. Anguelov, and L. Torresani, "Self-taught object localization with deep networks," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–9.

[22] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free? weakly-supervised learning with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 685–694.

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2921–2929.

[24] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE Transactions on Neural Networks and Learning Systems*, 2016.

[25] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, "Deep image mining for diabetic retinopathy screening," *Medical Image Analysis*, 2017.

[26] "https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15617," assessed on 2017-01-16.

[27] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *preprint arXiv:1502.03167*, 2015.

[28] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "The diaretdb1 diabetic retinopathy database and evaluation protocol.," in *British Machine Vision Conference (BMVC)*, 2007, pp. 1–10.

[29] J. Cuadros and G. Bresnick, "Eyepacs: an adaptable telemedicine system for diabetic retinopathy screening," *Journal of Diabetes Science and Technology*, vol. 3, no. 3, pp. 509–516, 2009.

[30] "https://www.kaggle.com/c/diabetic-retinopathy-detection," assessed on 2017-01-16.

[31] B. Dai, X. Wu, and W. Bu, "Retinal microaneurysms detection using gradient vector analysis and class imbalance classification," *PloS one*, vol. 11, no. 8, pp. e0161556, 2016.

[32] L. Zhou, P. Li, Q. Yu, Y. Qiao, and J. Yang, "Automatic hemorrhage detection in color fundus images based on gradual removal of vascular branches," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 399–403.

[33] Q. Liu, B. Zou, J. Chen, W. Ke, K. Yue, Z. Chen, and G. Zhao, "A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 78–86, 2017.

[34] M. Haloi, S. Dandapat, and R. Sinha, "A gaussian scale space approach for exudates detection, classification and severity prediction," *preprint arXiv:1505.00737*, 2015.

[35] V. M. Mane, R. B. Kawadiwale, and D. Jadhav, "Detection of red lesions in diabetic retinopathy affected fundus images," in *IEEE International Advance Computing Conference (IACC)*, 2015, pp. 56–60.

[36] D. P. Chakraborty, "Maximum likelihood analysis of free-response receiver operating characteristic (froc) data," *Medical physics*, vol. 16, no. 4, pp. 561–568, 1989.