A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification

Hongyang Jiang, Jie Xu*, Rongjie Shi, Kang Yang, Dongdong Zhang, Mengdi Gao, He Ma and Wei Qian

***Abstract*—The characteristics of diabetic retinopathy (DR) fundus images generally consist of multiple types of lesions which provided strong evidence for the ophthalmologists to make diagnosis. It is particularly significant to figure out an efficient method to not only accurately classify DR fundus images but also recognize all kinds of lesions on them. In this paper, a deep learning-based multi-label classification model with Gradient-weighted Class Activation Mapping (Grad-CAM) was proposed, which can both make DR classification and automatically locate the regions of different lesions. To reducing laborious annotation work and improve the efficiency of labeling, this paper innovatively considered different types of lesions as different labels for a fundus image so that this paper changed the task of lesion detection into that of image classification. A total of five labels were pre-defined and 3228 fundus images were collected for developing our model. The architecture of deep learning model was designed by ourselves based on ResNet. Through experiments on the test images, this method acquired a sensitive of 93.9% and a specificity of 94.4% on DR classification. Moreover, the corresponding regions of lesions were reasonably outlined on the DR fundus images.**

***Index Terms* — Diabetic Retinopathy, Deep Learning, Multi-label Classification, Grad-CAM.**

I. INTRODUCTION

With the development of economy and the improvement of living standard, people are increasingly concerned about their physical health, especially eye health. As is known to all, unhealthy diet and overuse of the eye may probably cause various eye diseases earlier than expected. Some retinal fundus diseases, such as diabetic retinopathy (DR), age-related macular degeneration (AMD), hypertensive retinopathy (HR), retinal vein occlusion and so on, were usually accompanied by special lesions on the retinal fundus image[1]. Among them, DR leads the rate of incidence and blind. It is rather significant to detect these lesions early and prevent deterioration.

Currently, other than eye hospitals many primary hospitals, community hospitals and physical examination centers have carried out fundus examination, especially DR detection[2]. However, the detection and diagnosis of the fundus need quite professional knowledge, which led to the scarcity of qualified ophthalmologists. Besides, traditional manual fundus inspection was very time-consuming and laborious. As the improvement of color fundus camera equipment and artificial intelligence technology, researchers proposed some novel methods to assist less experienced ophthalmologists to quickly make more accurate DR detection and diagnosis.

Previous research about DR screening mainly focused on two aspects. First, typical lesions of DR, such as bleeding points, microaneurysms, hemorrhages, exudates and cotton wool spots, were recognized separately based on object detection algorithms[3-5]. To develop thus lesion detection algorithms, sufficient annotated data of high quality were necessary. However, less available DR fundus images with lesions annotation information can be provided for the researcher, which limit the performance of algorithms. Second, most researchers have put forward image-based DR classification methods[6-8]. Labeling the whole fundus can be much easier and faster than pixel annotation, but lacking abundant interpretable information became the main shortcoming.

As the deep learning technology went step by step towards application of real scenes and acquired amazing effect. Deep learning-based algorithms have been recognized as one of the best algorithms for solving computer vision problems. When being faced with the small amount of the data during development of lesion detection algorithms, some researchers designed manifold data augmentation methods to improve the detection results, such as affine transformation, color conversion and morphological distortion[9]. In addition, other researchers tried their best to study the weakly or semi supervised theory to maximum the value of not too much annotated data[10,11]. Even though existed methods can help to alleviate the deficiencies of learning data, some innovative ideas need to be proposed urgently.

This paper proposed a novel approach to simultaneously complete DR classification and detection tasks based on multi-label[12] and Gradient-weighted Class Activation Mapping (Grad-CAM)[13]. Multi-label made a fundus image with more than one label, which have already been applied in non-medical image classification. In this study, different lesions of DR were considered as different labels of DR fundus images, so that the detail location of lesions on the fundus image need not to be provided by the annotation experts, which not only saved much annotation time but also decrease some incorrect or omitted labeling of lesions. To explain the classification results, Grad-CAM was furthermore utilized to exhibit the concrete location of lesions. The main contributions of this paper contained two aspects. First, lesions

Hongyang Jiang, Rongjie Shi, Kang Yang and Dongdong Zhang is with the Beijing Zhizhen Internet Technology Co., Ltd.

Jie Xu* is with Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Lab, Beijing, China (fionahsu920@foxmail.com).

Mengdi Gao is with the Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, China.

He Ma is with the college of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China and also with the Key Laboratory of Medical Image Computing, Ministry of Education.

Wei Qian is with the College of Engineering, University of Texas at El Paso, Texas 79968, USA.

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

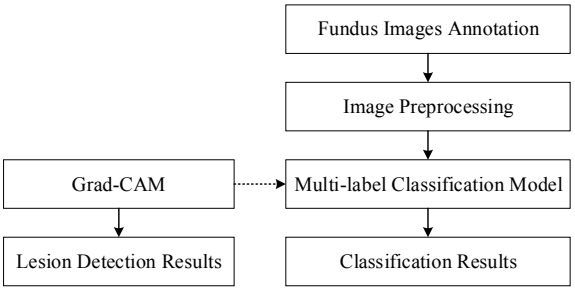

Figure 1. The algorithm framework of DR classification and lesions detection.

of DR were regarded as labels for collecting more learning data. Second, multi-label classification and Grad-CAM were combined together for achieving lesion detection. Keras, that is a high-level neural networks API written in Python and capable of running on top of Tensorflow, was employed for our experiments[14].

II. Method and Methodology

To complete DR lesions detection based on image classification model, this paper designed a multi-label classification model with Grad-CAM and introduced a new lesion labeling scheme for accelerating the collection speed of DR fundus images with multiple lesions. The whole algorithm framework is displayed in Fig 1, which contains three modules, fundus images annotation, image preprocessing and multi-label classification. The technical details are set forth below.

A. Fundus image annotation

The original DR fundus images used in this study were from public data (Messidor [15]) and private data provided by Beijing Tongren Eye Center. All the private data have been processed with sensitive information elimination. This study designed a novel annotation method, called element annotation. The element here was defined as one kind of lesions or one kind of structures on the fundus image. The process of element annotation can be described as follows.

- selection of elements to label.

  Target elements should be affirmed in advance and listed concisely. The definition of these elements should also be described clearly with diagram if necessary.

- annotation training for candidates.

  To produce standardized annotation data, the candidates for labeling data should be trained uniformly according to a piece of pre-established standard.

- multi-label generation.

  Each fundus image owned a label vector $l=\{e_1, e_2, \ldots, e_k\}$, where e_i ($i \in [1, k]$) represents one specified label and the value of e_i can only be 0 or 1.

B. Image pre-processing and augmentation

A series of image preprocessing operations were implemented to denoise the fundus images and enhance the imaging features of them. First of all, the meaningless contents of a normative fundus image, such as black background and

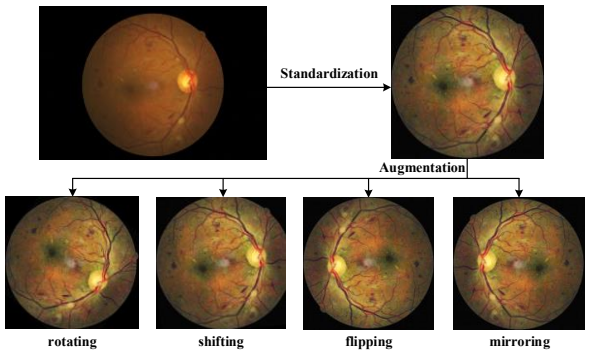

Figure 2. The algorithm framework of DR classification and lesions detection.

technical parameters on the image, were eliminated through traditional pixel value analysis methods. Then, the pixel values of each fundus image were normalized into fixed range from 0 to 1 and the sizes of all fundus images were resized to 512×512. Finally, all the training fundus images were enhanced by the contrast limited adaptive histogram equalization (CLAHE)[15].

Moreover, some traditional image augmentation methods including image shifting, rotating, flipping and mirroring was randomly applied during training phase. The results of image preprocessing and augmentation can be shown in Fig 2.

C. Multi-label classification model and training

This paper proposed a multi-label classification model based on the architecture of Resnet[16], which can be illustrated in Fig 3. The base model was a modified Resnet50 model without fully connected (FC) layers. On the top of base model, we added three convolutional layers instead of the original FC layer to construct a full convolution network. Besides, the number of output nodes in the output layer was equal to the number of categories we pre-set. It should be noted that the activation function of the output layer was no more a softmax function, but a sigmoid function. At the same time, the loss function was changed from the categorical cross entropy to the binary cross entropy.

In the training phase, we initialized the base model with pretrained weights on the ImageNet data[17] and randomly initialized the newly added layers. In addition, a two-stage training strategy was adapted. First stage, the weights of base model were frozen and just trained the other layers with a large learning rate (0.01). Second stage, all the weights of the model were trained again using a small learning rate (0.001). We also automatically allocated each optimum threshold to each

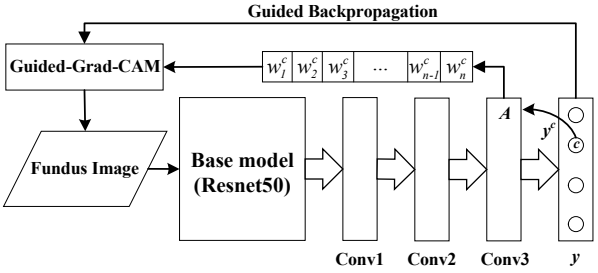

Figure 3. The architecture of multi-label classification model with Guided-Grad-CAM.

category when making real-time assessment on the validation data. Hence, it was convenient for us to analyze the learning effect of each category.

D. Gradient-weighted Class Activation Mapping

Interpretability for a deep learning model has become an essential characteristic. Researchers cared more about the inference process when they were developing their models, especially the classification model. Based on our proposed multi-label classification model, the Grad-CAM technology was employed to give inferential explanation (specific location) of each category on the original fundus image.

The Grad-CAM is the general form of CAM, which can be applied to any deep learning model with convolution structure. Commonly, the last convolution layer can be chosen to compute the Grad-CAM. We assumed that the output map of the last convolution layer was denoted as A^k, where k is the number of these output maps. The final Grad-CAM $I_{Grad\text{-}CAM}$ can be calculated as follows,

$$w_k^c = \frac{1}{Z}\sum_{i=1}^{W}\sum_{j=1}^{H}\frac{\partial y^c}{\partial A_{ij}^k} \tag{1}$$

$$I_{Grad-CAM}^c = ReLU\left(\sum_{k=1}^{K} w_k^c \cdot A^k\right) \tag{2}$$

where y^c denotes the score of class c before the softmax and the size of A^k is $W \times H$. Through differential operation of y^c with respect A^k, we obtained w_k^c as the weight of map A^k for class c and Z is the normalization factor. After performing a weighted summation of map A^k, an activation function of the rectified linear unit (ReLU) was implemented. Besides, the map of Guided Backpropagation of each predicted result $I_{Guided-Backprop}^c$ was computed. Then we can acquire a more fine-grained result (Guided-Grad-CAM) via point-wise multiplying the Grad-CAM and the Guided Backpropagation as follows.

$$I_{Guided-Grad-CAM}^c = I_{Guided-Backprop}^c \cdot I_{Grad-CAM}^c \tag{3}$$

To give a final integrated Guided-Grad-CAM of the multi-label classification results, we combined all the Guided-Crad-CAMs together through normalization,

$$I_{Guided-Grad-CAM} = \frac{1}{Z}\sum_{c=1}^{C} I_{Guided-Grad-CAM}^c \tag{4}$$

where Z is the normalization factor and C is the number of categories of our multi-label classification model.

III. Results and Discussion

Based on our pre-defined data annotation standard, totally 25 ophthalmologists participated in the DR fundus image annotation work. It was notable that fundus images of low quality including extremely bright and dark illumination, image globally blur and noise interference, should not be added in our dataset. In this experiments, four kinds of lesions were labeled, that was small bleeding point or microaneurysm (BP/MA), hemorrhage (HM), hard exudate (HEX) and cotton wool spot (CWS) respectively. The labels of each fundus image were validated by more than three ophthalmologists and the detail information can be shown in Fig 4. A total of 3228 fundus images have been accumulated, including 744, 712, 716, 736, 720 labels of BP/MA, HM, HEX, CWS, healthy, respectively. From this well-annotated multi-label dataset, 2878 fundus images were randomly selected as training set and the rest as test set.

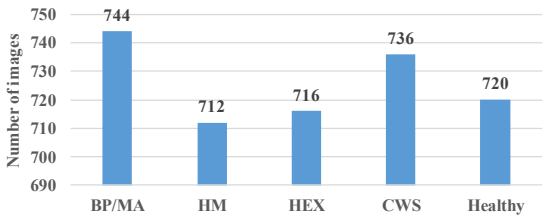

Figure 4. The distribution of fundus images of four lesions (BP/MA, HM, HEX, CWS and healthy) in our dataset.

TABLE I. Performance Indexs of Five Categories: BP/MA, HM, HEX, CWS and Healthy.

Lesion Type	Sensitivity	Specificity	Accuracy	AUC
BP/MA	85.5%	90.7%	89.4%	0.941
HM	100%	98.6%	98.9%	1.000
HEX	93.3%	92.7%	92.8%	0.978
CWS	94.6%	86.8%	88.6%	0.971
Healthy	93.9%	94.4%	94.2%	0.989

To efficiently complete the training of our multi-label classification model, all the experiments were implemented on a cloud server with an Ubuntu system of 16.04.5 LST version and x86_64 architecture. The hardware configuration contained twelve Intel Xeon CPUs of 2.40GHz cores, 100GB memory and one NVIDIA Tesla P40 GPU of 24GB memory.

In the experiments, to balance the training effect of every category, we allocated a weight that was inversely proportional to the number of each category, to the loss function of its corresponding category when computing the total loss. This paper drew the receiver operating characteristic (ROC) curve to make assessment on the classified result of each category. The ROC curves of each category can be seen in Fig 5 and the evaluation indexes including sensitivity, specificity and area under the curve (AUC) are illustrated in TABLE I.

Based on the well-trained multi-label classification model, the Guided-Grad-CAM was figured out and highlighted the target lesions on the fundus image, which are displayed in Fig

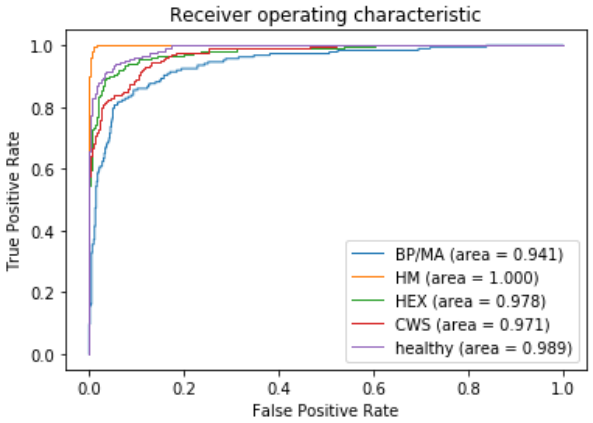

Figure 5. The ROC curves of five categories: BP/MA, HM, HEX, CWS and healthy.

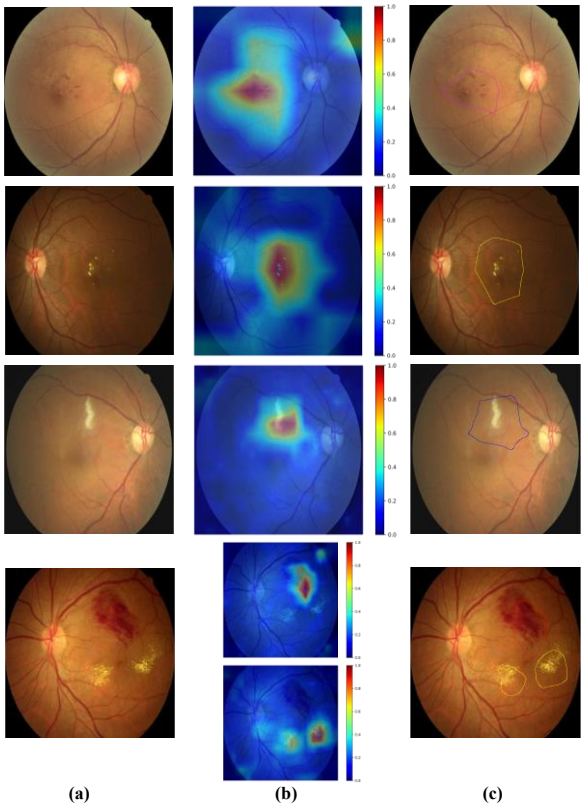

Figure 6. (a) Original fundus images; (b) Guided-Grad-CAMs of multi-label classification model; (c) The detection results of different lesions: pink, red, yellow, blue solid line denotes the BP/MA, HM, HEX, CWS, respectively.

6. Fig 6(a) shows the original fundus images, Fig 6(b) displays the Guided-Grad-CAMs of the classification results and Fig 6(c) illustrates the detection results of DR lesions. Through basic image processing methods, such as threshold segmentation and edge detection, we can point out the lesions using irregular solid line with different colors.

IV. Conclusion

This paper proposed a multi-label classification model with the interpretable Grad-CAM. With the limitation of ophthalmologist resources, simplifying data annotation work can greatly increase the quantity of valuable data. In the fundus image annotation phase, we formulated an element annotation standard focusing on lesions of DR fundus images for multi-label classification, which improved the efficiency of labeling work. Besides, to accomplish the lesion detection on a fundus image through a multi-classification model, this paper utilized Grad-CAM to automatically outline the specific region of each lesion. The experimental results demonstrated the effectiveness and accuracy of DR classification and lesion detection by our method. In addition, as the accumulation of DR fundus images, more abundant lesions or features of DR can be taken as separate categories adding to our multi-label classification model and more precise lesion location can be acquired from the Grad-CAM.

Acknowledgment

We would like to thank Beijing Zhizhen Internet Technology Co., Ltd. (the official website is www.zhenhealth.cn) for kindly offering financial support. Besides, we are particularly grateful to Beijing Tongren Eye Center for providing the research data and medical guidance. Finally, we also feel appreciated to the researchers in the Northeastern University for providing experimental guidance.

References

[1] Nirmala S R, Nath M K, Dandapat S., "Retinal image analysis: A review," *International Journal of Computer & Communication Technology (IJCCT)*, 2(VI), pp. 11-15, 2011.

[2] Pieczynski J, Grzybowski A., "Review of diabetic retinopathy screening methods and programmes adopted in different parts of the world," *Journal-Review of Diabetic Retinopathy Screening Methods and Programmes Adopted in Different Parts of the World*, 2015.

[3] Kar S S, Maity S P., "Automatic detection of retinal lesions for screening of diabetic retinopathy," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 608-618, 2017.

[4] Amin J, Sharif M, Yasmin M, et al., "A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions," *Journal of Computational Science*, vol. 19, pp. 153-164, 2017.

[5] Yu S, Xiao D, Kanagasingam Y., "Exudate detection for diabetic retinopathy with convolutional neural networks," *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1744-1747, 2017.

[6] Jiang H, Yang K, Gao M, et al., "An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2045-2048, 2019.

[7] Dutta S, Manideep B C, Basha S M, et al., "Classification of diabetic retinopathy images by using deep learning models," *International Journal of Grid and Distributed Computing*, vol. 11, no. 1, pp. 89-106, 2018.

[8] Mansour R F., "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy," *Biomedical engineering letters*, vol. 8, no. 1, pp. 41-57, 2018.

[9] Shorten, Connor, and Taghi M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data 6.1*, no. 60, 2019.

[10] Wang R, Chen B, Meng D, et al., "Weakly Supervised Lesion Detection From Fundus Images," *IEEE transactions on medical imaging*, vol. 38, no. 6, pp. 1501-1512, 2018.

[11] Costa P, Galdran A, Smailagic A, et al., "A weakly-supervised framework for interpretable diabetic retinopathy detection on retinal images," *IEEE Access*, vol.6, pp. 18747-18758, 2018.

[12] Wang J, Yang Y, Mao J, et al., "Cnn-rnn: A unified framework for multi-label image classification," *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2285-2294, 2016.

[13] Selvaraju R R, Cogswell M, Das A, et al., "Grad-cam: Visual explanations from deep networks via gradient-based localization," *Proceedings of the IEEE international conference on computer vision*, pp. 618-626, 2017.

[14] Gulli A, Pal S., "Deep learning with Keras," *Packt Publishing Ltd*, 2017.

[15] MESSIDOR: Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. [Accessed June 08, 2017]. Available from: http://www.adcis.net/en/Download-Third-Party/Messidor.html.

[16] Sahu S, Singh A K, Ghrera S P, et al., "An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE," *Optics & Laser Technology*, vol. 110, pp. 87-98, 2019.

[17] He K, Zhang X, Ren S, et al., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

[18] Russakovsky O, Deng J, Su H, et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211-252, 2015.