CATARACT DETECTION AND GRADING BASED ON COMBINATION OF DEEP CONVOLUTIONAL NEURAL NETWORK AND RANDOM FORESTS

Jing Ran[1], Kai Niu[1], Zhiqiang He[1], Hongyan Zhang[2], Hongxin Song[2]

[1]Key Laboratory of Universal Wireless Communations, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China
{ranjing2012,niukai,hezq}@bupt.edu.cn, zhanghongyantr@163.com, songhongxin2012@163.com

Abstract: Cataract is one of the most common eye diseases which leads to visual impairment and is the main cause of blindness. Early intervention and timely treatment can largely avoid cataract blindness. Cataract grading based on fundus images by artificial intelligence algorithms is a feasible method to assistant doctors to diagnose cataracts more effectively. In this paper, a method that Deep Convolutional Neural Network (DCNN) combined with Random Forests (RF) is proposed for six-level cataract grading. In this method, DCNN consists of three modules for feature extraction at different levels on fundus images, while RF implements more elaborate six-level cataract grading based on the feature datasets generated by DCNN. The six-level grading allows doctors to understand the patient's condition more accurately than four-level grading. The accuracy of six-level grading achieved by the proposed method is up to 90.69% on average, with superiority in specificity and sensitivity indicators. Our experimental results also show that RF improves the grading accuracy and reduces the concussion of DCNN on small datasets.

Keywords: Cataract; Six-level grading; Random Forests; Deep Convolutional Neural Network; Combination

1 Introduction

Cataract is clouding of lens inside the eyes which leads to vision loss. Cataract is an age-related eye disease, and its incidence rate rises with age, but it occurs to newborns occasionally, or it may be the sequelae of eye injury and other eye diseases.

Cataract is the leading cause of blindness with high incidence rate. Recently, a report [1] forecasts that the number of people with moderate or severe vision impairment is 237.1 million and blindness is 38.5 million in 2020, of which 24% (57.1 million) and 35% (13.4 million) are caused by cataract. In China, the situation is grim, 35 million cataract patients are recorded, nearly 500 thousand new patients every year.

Until today, surgery is still the most effective treatment of cataract. Although the cataract surgery is convenient and valid, cataract diagnosis is a heavy task for ophthalmologists due to the long-term cultivation and the shortage of ophthalmologists and the increasing number of cataract patients. So, finding out a viable and fast way to help cataract diagnosis is significant for society. With the popularity of medical Artificial Intelligence (AI) today, it is feasible to use AI to help diagnose cataract [2] to alleviate the aforementioned conflicts.

Fundus image is the main processing object of ophthalmology AI assistant algorithms. Previous works are concerned with feature recognition and extraction [3] [4]. Then, researchers develop AI diagnosis systems for specific eye diseases like diabetic retinopathy [5], glaucoma [6] and cataract [2] [7-12]. Those cataract diagnosis systems or methods can be divided into two categories by the methods they use to extract features. One is manual extraction and the other kind is automatic extraction, the former dominates in quantity.

In the first type of previous works, Micheal et al [7] graded fundus images into normal and abnormal, which was the first try of machine learning in the fundus images grading. Meimei Yang et al [8] adopted Back Propagation Network to class fundus images into normal, mild, medium and severe grades. An ensemble learning based method was put forward in [9] to improve grading accuracy. Lately, Z Qiao et al [10] cascaded genetic algorithm and SVM, which used genetic algorithm to assign suitable weights to features before classification by SVM.

But in the other kind, Y Dong et al [11] used 5-layer DCNN to extract features and SVM to grade, which freed researchers from cataract-related feature design and extraction. L Zhang et al [12] picked green channel images from fundus images to be graded by 5-layer deep convolutional network. It reduced the interference of illumination and decreased calculations.

Compared to the first type of the previous works, the second type shows the potentiality and convenience of feature extraction by DCNN, however, the dilemma also appears that the data is insufficient for DCNN to make further improvement on grading accuracy.

In this paper, we propose the method that DCNN combined with RF (DCNN-RF). We employ DCNN to extract features. Considering the dimensions of fundus

978-1-5386-6067-6/18/$31.00 ©2018 IEEE

images and high similarity between them, residual units [13] are introduced to reduce computational complexity and extract rich detailed features. In addition, RF [14] is combined with the DCNN to achieve better performance on small datasets due to the high accuracy, relatively strong robustness and rapidity of RF.

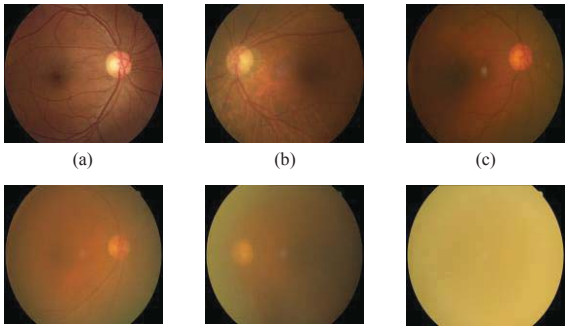


Figure 1 Images of cataract in six grades; (**a**) non-cataractous (**b**) slightly mild (**c**) mild (**d**) medium (**e**) slightly severe (**f**) severe

Moreover, the grading by RF in this paper is different from all previous works. The domestic demand of cataract surgery is mainly considered, so, more detailed six-level grading is made according to [15]. The cataract fundus images in six grades are shown in Figure 1. Results of experiments show that the proposed method achieves better performance than manual feature extraction method on four-level cataract grading and the method is well adapted to the six-level cataract grading.

This paper is organized as following: Section 2 represents the framework of our method and describes the details of the framework; Section 3 shows experiments and analysis; Section 4 is conclusion.

2 Methodology

In this section, the framework of proposed method is represented and the details of three main parts in it are described. Those three parts are preprocessing that consists of standardization and so on, feature extraction that is achieved by DCNN and RF grader which completes cataract grading on the extracted feature datasets.

2.1 Framework

The framework of our method for cataract grading is represented by a flow chart in Figure 2.

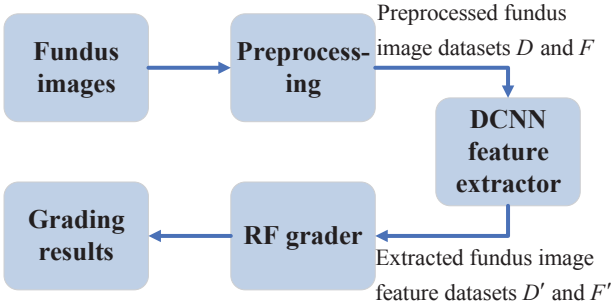


Figure 2 Flow chart of proposed method for cataract grading

The process of cataract grading by DCNN-RF [13] [14] contains three main parts, preprocessing, DCNN feature extractor and RF grader.

Preprocessing is based on fundus images, and it contains some common operations like desensitizing. The preprocessed images are divided into training dataset $D=\{(\boldsymbol{x}_i,y_i)\,|\,i=1,...,N\}$ and test dataset $F=\{(\boldsymbol{x}_j,y_j)\,|\,j=1,...,M\}$, $\boldsymbol{x}_i$ is the i-th image in D and y_i is the grading label of $\boldsymbol{x}_i$, and the same for $\boldsymbol{x}_j$ and y_j. N and M are the number of images in D and F. DCNN feature extractor refers to DCNN implements feature extraction on D and F, then gets $D'=\{(\boldsymbol{x}'_i,y'_i)\,|\,i=1,...,N\}$ and $F'=\{(\boldsymbol{x}'_j,y'_j)\,|\,j=1,...,M\}$, $\boldsymbol{x}'_i$ and $\boldsymbol{x}'_j$ are feature vectors of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The whole feature set of those extracted feature vectors is K, and the size of it is 512. Then D' and F' are sent to RF grader, and the RF algorithm achieves further six-level cataract grading on them.

2.2 Preprocessing

Fundus images are captured by specialized fundus cameras. This acquisition process introduces many divergences because of the differences between cameras, operators and surroundings. In order to standardize the fundus images, we take specifications standardization to get unified rectangular fundus images, desensitization to wipe off personal information of patients for privacy protection, and general nonlinearity brightness adjustment to reduce the interference of different illumination for further steps such as feature extraction.

2.3 Deep Convolutional Neural Network feature extractor

The DCNN feature extractor adopts a residual network with 17 convolutional layers to learn more abstract representations of fundus images. The whole DCNN contains shallow module, residual module and pooling module. The shallow and residual module extract the features on shallow, medium and deep level, and pooling module outputs the final feature vectors for RF grader.

Shallow module

The shallow module and some feature maps gained by layers in it are shown in Figure 3. In this module, one convolutional layer is used to extract detailed and basic fundus image features. The operation of a convolutional layer completed by convolutional kernels is represented by (1) and (2).

$$Y^C = X^G * H^{GC} \tag{1}$$

$$y^c_{i,j} = \sum_g \sum_m \sum_n x^g_{i+m,j+n} h^{gc}_{m,n} \tag{2}$$

Where X are input images and the superscript G means input channels of this convolutional layer. H is the convolutional kernel, and C is the number of the

output channel. $*$ means convolutional operation, and Y^C are result feature maps. Formula (2) is the detailed expansion of the convolutional operation. $x^g_{i+m,j+n}$, $h^{gc}_{m,n}$, $y^c_{i,j}$ are elements of X^G, H^{GC} and Y^C respectively.

BatchNorm layer normalizes batches of data to enable the model to use greater learning rate for faster convergence. ReLU is an activation function adopted to add non-linear information to feature maps. Pooling operation in (3) helps to make the fundus feature representation approximately unchanged, therefore, features that are similar but at different locations in fundus images of the same grade are comparable.

$$z_{i,j} = g(\boldsymbol{y}^c_{p_i,q_j}) \tag{3}$$

Where $g(\cdot)$ is pooling operation, which can be maximum, average and so on. $\boldsymbol{y}^c_{p_i,q_j}$ is the (i,j)-th block which is selected in order on $\boldsymbol{y}^c$ by the pooling window with size (p,q), $z_{i,j}$ is the result of pooling on this block.

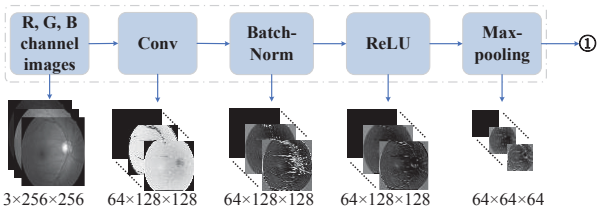
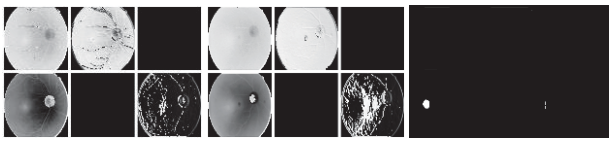


Figure 3 Composition of shallow module

The features extracted are fundamental concepts such as edge information, brightness information and background texture information. Besides, the shallow module differentiates various features like optic disc and choroidal texture into feature maps as shown in Figure 4.



Non-cataractous mild Slightly severe

Figure 4 Shallow level features learned by shallow module

Residual module

The residual module shown in Figure 5 is the main module for feature extraction and abstraction. This module is composed of four residual layers to deepen the network and learn the features better, and each residual layer contains two residual units. Figure 6 represents the framework of a residual unit.

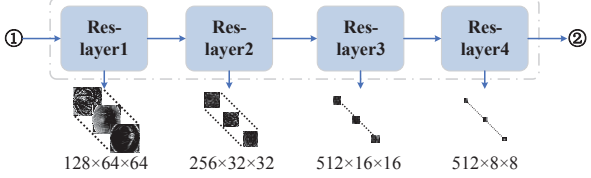


Figure 5 Composition of residual module

Residual unit is the key structure of the DCNN in proposed method. Conv1 and Conv2 are feature extraction layers, as a skeleton to achieve a shortcut connection. BatchNorm normalizes data distribution to avoid gradient disappearance. ReLU provides non-linear modeling capabilities for the network. The unit is expressed by (4) and (5).

$$\boldsymbol{y}_l = h(\boldsymbol{x}_l) + F(\boldsymbol{x}_l, W_l) \tag{4}$$

$$\boldsymbol{x}_{l+1} = f(\boldsymbol{y}_l) \tag{5}$$

Where $\boldsymbol{x}_l$ and $\boldsymbol{x}_{l+1}$ are input and output of the l-th residual unit, and $\boldsymbol{y}_l$ is the intermediate output. F is a residual function. W_l is a set of weights and biases of layers in this unit, which are Conv1 and Conv2. Generally, $h(\boldsymbol{x}_l)$ is an identity mapping, and to simplify, we set f an identity mapping in the residual unit, so, $\boldsymbol{x}_{l+1} = \boldsymbol{y}_l$.

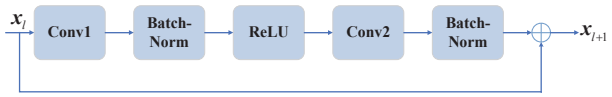
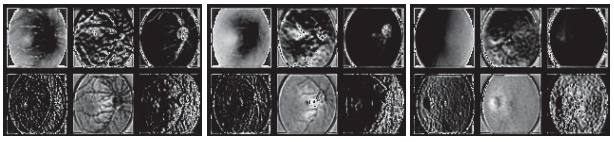


Figure 6 Framework of residual unit

The medium features like spatial distribution and illumination in Figure 7 are captured by residual module. Deeper features learned by the residual module as well are high-order concepts like chaos and overall texture distribution in fundus images as shown in Figure 8.



Non-cataractous mild Slightly severe

Figure 7 Medium level features learned by residual module

Non-cataractous mild Slightly severe

Figure 8 Deep level features learned by residual module

Pooling module

The pooling module in Figure 9 is a simple average pooling layer for data dimension reduction and over-fitting prevention. Instead of max pooling, average pooling is used to integrate and fully use the features extracted by the residual module. Then datasets D' and F' consist of feature vectors produced by the pooling module are sent to RF as in Figure 9.

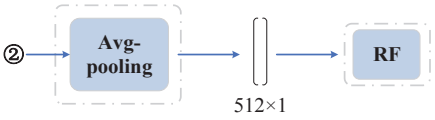


Figure 9 Pooling module and RF

2.4 Random Forests grader

Because the qualified fundus images are not easy to

obtain, the number of fundus images we use is only a few thousand. That is not enough for DCNN. But the conducted cross validation to reduce the disturb of contingency brings in data changes. Thus, considering the dependence of DCNN on data, RF is used to reduce DCNN fluctuations on our dataset.

RF is a combination of tree predictors. The t-th tree in RF is grown on a training set D_t' which is drawn randomly from D'. When the t-th tree is trained by D_t', it picks k features randomly from K, and k is set to empirical value $\log_2 K$. The tree estimates performance by Gini Coefficient to choose suitable features from k features for sample division. The process of RF training and testing is as follows.

Input
Training dataset D', test dataset F', feature set K
Training
For $t = 1, 2, ..., T$
pick n_t samples randomly from D' as D_t';
pick k features randomly from K to generate decision tree $G_t\left(\boldsymbol{x}_{i_t}'\right), i_t = 1, ..., n_t$ on D_t'.
Testing
Input F' into decision trees $G_t, t = 1, ..., T$.
Output average instead of voting on the prediction of $G_t\left(F'\right), t = 1, ..., T$ as final result.

Based on the detailed and effective features extracted by DCNN, RF achieves more detailed six-level cataract grading than previous four-level grading, and with good performance that is shown in the next section.

3 Experiments and Analysis

In this section, some details of dataset and evaluation criteria used in experiments are represented firstly, and then the experimental results and analysis are shown.

3.1 Dataset

We have 5408 preprocessed images as experimental dataset so far. The dataset consists of 1948 (1268, 496, 616, 540, 540) non-cataractous (slightly mild, mild, medium, slightly severe, severe). Each image in the dataset is labeled by two ophthalmologists and checked by three experienced graders.

3.2 Evaluation Criteria

For observation, some common evaluation criteria are used to measure the performance of the new method on cataract detection and grading. They are accuracy, sensitivity and specificity. The formulas of them are $S_a = \frac{T_P + T_n}{T_p + F_p + T_n + F_n}$, $S_e = \frac{T_p}{T_p + F_n}$ and $S_p = \frac{T_n}{T_n + F_p}$. S_a, S_e, S_p represent accuracy, sensitivity and specificity respectively. T_P, F_p, T_n, F_n are the number of true positive, false positive, true negative and false negative samples.

3.3 Experimental results and analysis

Cross validation is used in experiments to ensure the reliability. We conduct 20 groups of 5-fold cross validation for each experiment, and some of the results are shown in following tables and diagrams. M-SVM is used to represent methods that manual feature extraction combined with SVM classification for convenience.

Table I Cataract detection results of DCNN-RF

Results\Indicators	S_e (%)	S_p (%)	S_a (%)
Results	97.26	96.92	97.04

The presence or absence of cataract is our foremost concern, so we test the ability of cataract detection of the proposed method. In this experiment, cataract data is treated as positive data and non-cataractous data is negative. Average result of cross validation experiments in Table I is up to 97.04% accuracy, with high specificity and sensitivity at the same time.

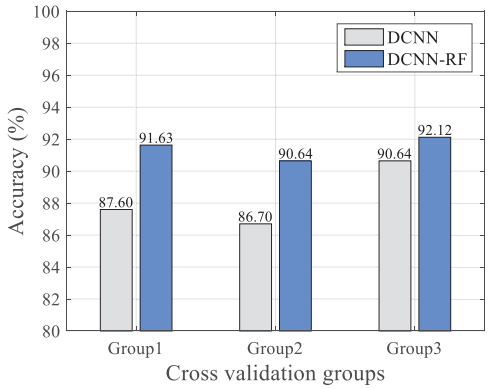


Figure 10 Results comparison between DCNN and DCNN-RF

In Figure 10, it is obvious that DCNN-RF achieves higher accuracy of six-level grading compared to single DCNN, and reduces the fluctuation range from 3.94% to 1.48%. This kind of trend is consistent in 20 groups of this experiment. All of those indicate that the RF helps to achieve higher grading accuracy and lower fluctuation of accuracy to some extent.

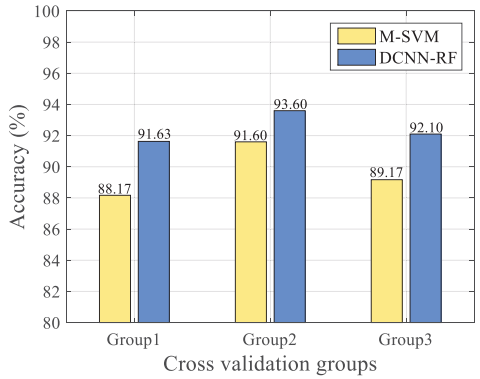


Figure 11 Results comparison between M-SVM and DCNN-RF

For further observation, we compare DCNN-RF with M-SVM on the four-level cataract grading. The manual extraction of features mostly references [9] to achieve. Results are in Figure 11. Compared with M-SVM, the proposed method has some improvements in performance on the four-level cataract grading.

In Table II, six-level grading results of proposed method represents that non-cataractous and severe grades have both high sensitivity and specificity, which means that our method has strong ability to identify and diagnose these levels' cataracts. On other levels, our method also achieves high specificity and not bad sensitivity. Those results and the 90.69% average grading accuracy show the method is well adapted to the six-level grading.

Table II Cataract grading on six-level criterion by DCNN-RF

Grades	Non-cataractous		Slightly mild		Mild		Medium		Slightly severe		Severe		S_a (%)
Indicators	S_e (%)	S_p (%)	S_e (%)	S_p (%)	S_e (%)	S_p (%)	S_e (%)	S_p (%)	S_e (%)	S_p (%)	S_e (%)	S_p (%)	
Results	95.89	98.15	88.96	96.13	76.31	98.15	85.22	98.05	87	98.69	99.5	99.34	90.69

4 Conclusions

The method DCNN combined with RF is proposed in this paper for six-level cataract grading. DCNN extracts features from fundus images automatically. RF completes cataract grading on the feature datasets produced by DCNN. The grading is six-level which is more difficult than four-level one. Experiments show that this method achieves 97.4% average detection accuracy, 90.64% grading accuracy and good performance on both sensitivity and specificity. The comparison experiments represent that the proposed method helps to achieve higher grading performance and lower fluctuation than single DCNN, and it is also more accurate than the method that manual extraction combined with SVM on the four-level grading. Although this method performs well, it is still sensitive to data changes, and the parameters in RF are empirical or need iteration to be found out. In the future, we will try to improve this method to make it more automatic and adaptable.

Acknowledgements

This work is supported by Major National Scientific &Technological Specific Project of China under grant number 2017ZX03001004 and the National Natural Science Foundation of China (No. 61671080).

References

[1] S R Flaxman, R Bourne, S Resnikoff, et al. Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. The Lancet Global Health, 2017, 5(12): e1221-e1234.

[2] V Harini, V Bhanumathi. Automatic cataract classification system. Communication and Signal Processing (ICCSP), 2016 International Conference on. IEEE, 2016: 0815-0819.

[3] S Aslani, H Sarnel. A new supervised retinal vessel segmentation method based on robust hybrid features. Biomedical Signal Processing and Control, 2016, 30: 1-12.

[4] S Bharkad. Automatic segmentation of optic disk in retinal images. Biomedical Signal Processing and Control, 2017, 31: 483-498.

[5] A Gulshan, L Peng, M Coram, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama, 2016, 316(22): 2402-2410.

[6] S Willikens, E Zitron, E Scholz, et al. Retinal Arterio-Venule-Ratio (AVR) in the cardiovascular risk management of hypertension. European Heart Journal, 2013, 34(suppl_1).

[7] PC Siddalingaswamy, GK Prabhu. Automated detection of anatomical structures in retinal images. Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on. IEEE, 2007, 3: 164-168.

[8] M Yang, JJ Yang, Q Zhang, et al. Classification of retinal image for automatic cataract detection. E-health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on. IEEE, 2013: 674-679.

[9] JJ Yang. J Li, R Shen, et al. Exploiting ensemble learning for automatic cataract detection and grading. Computer methods and programs in biomedicine, 2016, 124: 45-57.

[10] Z Qiao, Q Zhang, Y Dong, et al. Application of SVM based on genetic algorithm in classification of cataract fundus images. Imaging Systems and Techniques (IST), 2017 IEEE International Conference on. IEEE, 2017: 1-5.

[11] Y Dong, Q Zhang, Z Qiao, et al. Classification of cataract fundus image based on deep learning. Imaging Systems and Techniques (IST), 2017 IEEE International Conference on. IEEE, 2017: 1-5.

[12] L Zhang, J Li, H Han, et al. Automatic cataract detection and grading using Deep Convolutional Neural Network. Networking, Sensing and Control (ICNSC), 2017 IEEE 14th International Conference on. IEEE, 2017: 60-65.

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

[14] L Breiman. (2001) Random forests. Machine Learning, 45(1), 5–32.

[15] 李建军, 徐亮, 杨桦, 等. 需要手术治疗的白内障远程筛查眼底像评估标准 (征求意见稿). 眼科, 2015 (6): 367-368.