Detecting abnormal fundus images by employing deep transfer learning

yan yu
Wannan Medical College https://orcid.org/0000-0002-3149-0271
Changfan Wu (wucangfan@sina.com)
First Affiliated Hospital of Wannan Medical University
Xiao Chen
Anhui Normal University
Xiangbing Zhu
Anhui Normal University
Yinfen Hou
First Affiliated Hospital of Wannan Medical University
Pengfei Zhang
First Affiliated Hospital of Wannan Medical University
Rongrong Zhang
First Affiliated Hospital of Wannan Medical University

Research article

Keywords: Fundus images, Deep transfer learning, Developing and validation, Artificial intelligence

Posted Date: August 31st, 2019

DOI: https://doi.org/10.21203/rs.2.13807/v1

Abstract

Background

To develop and validate a deep transfer learning (DTL) algorithm in detecting abnormalities of fundus images from non-mydriatic fundus photography examination.

Methods

1,295 fundus images from January 2017 to December 2018 at Yijishan Hospital of Wannan Medical College were collected for developing and validating the deep transfer learning algorithm in detecting abnormal fundus images. The DTL model was developed by using 929(normal 254, abnormal 402) fundus images, including normal fundus images and abnormal fundus images, the latter including, maculopathy, optic neuropathy, vascular lesion, choroidal lesions, vitreous disease, cataract and the others. We tested our model using a subset of the publically available MESSIDOR dataset (using 366 images) and evaluate the testing performance of the DTL model for detecting abnormal fundus images.

Results

In the internal validation data set (n=273 images), the AUC, sensitivity, accuracy and specificity of the DTL for correctly classified funds images were 0.997, 97.41%, 97.07% and 96.82%, respectively. For test data set (n=273 images), the AUC, sensitivity, accuracy and specificity of the DTL for correctly classification funds images were 0.926, 88.17%, 87.18% and 86.67%, respectively.

Conclusion

In the evaluation, the DTL presented high sensitivity and specificity for detecting abnormal fundus-related diseases. Further research is necessary to improve this method and evaluate the applicability of the DTL in the community health care center.

Background

Retinal disease is one of the main causes of blindness across the world, and the most common types of retinal conditions are dysfunctional retinal pigment epithelium and degenerating photoreceptors. Aging, diabetes, trauma, retinal vessel occlusion, hypertensive retinopathy, retinitis, and family history can result in retinal disease. With the increase in aging population and the prevalence of high myopia and diabetes, the visual disabilities will keep growing[1]. At present, the diagnosis of retinal diseases mainly relies on manual examination with the help of eye experts on retinal vessels, optic discs, the fovea and lesions. As the prevalence of vision disabilities increases[2], early detection and effective treatment are key to avoid vision loss. The community health care center with population concentration, comprehensive monitoring, capabilities of analyzing and evaluating individual or group health, have a prospect of providing large-scale screening and early diagnosis. However, one of the main barriers to implementing widespread screening is the deficit of medical resources, particularly in low- and middle-income countries[3]. Given the concerns, developing a safe and effective screening program for early intervention to prevent currently incurable blinding conditions is essential. Retinal fundus images have become one of the main references for screening and diagnosing retinal diseases. Recently, several research teams have investigated in artificial intelligence-assisted systems based on fundus photographs to screen retinal diseases. However, plenty of these studies devoted to identifying DMR[4] and glacuma[5, 6], studies about retinal diseases recognition aiming to establish a classification of normal and abnormality in multi-categorical retinal diseases have been very limited.

Artificial intelligence(AI) using machine learning algorithms, such as support vector machine (SVM), Naive Bayes classifier and convolutional neural networks (CNN), has received extensive attention after demonstrating that it could perform at least as well as humans in image classification tasks[4, 7]{J, 2018 #30;J, 2018 #30}. As the digital imaging modality rapidly develops, image processing, computer vision and machine learning are used in automatically detecting the retinal lesions based on color fundus photographs. This is of great significance for the implementation of computational assisted retinal diseases detection and promotion of large-scale screening[8]. Deep transfer learning is a new machine learning method that leverages existing knowledge to solve different but related domain problems[9]. Confirming past studies, the transfer learning was a high effective technology, especially in domain where with limited data[10]. The essential characteristic of DTL compare with traditional image recognition methods which does not need to rely on manually labeling and a large amount of labeled training data, and not require a lot of cost and time for data collection. The purpose of this study was to develop and validate an effective transfer learning algorithm for detecting abnormal fundus photographs and providing an accurate and timely referral by employing a small multi-categorical retinal disease image database. In the meantime, generating new insights into screening program, to efficiently build a detection model with a few labeled fundus photographs and some relation graph datas.

Methods

Image datasets characteristics

A total of 1,295 fundus images were selected from the Yijishan Hospital of Wannan Medical College from January 2017 to December 2018 in this retrospective study. These images included normal and abnormal fundus photographs, the latter including maculopathy, optic neuropathy, vascular lesion, choroidal lesions, vitreous disease, cataract and low-quality photographs. Image will be labeled as poor quality and removed from the training and validation dataset in the following situations, blurred areas accounted for 50% or more, macula lutea and optic disc is only one or none, macular region’s vessels can not be distinguished. After removed 366 poor images, the deep transfer learning (DTL) model was developed using 929 retinal fundus images (normal 370, abnormality 559) from January 2017 to December 2018. Fig. 1 shows the workflow of this study. The images were extracted from the ophthalmic clinics, inpatients and physical examination center in our hospital. Three data sets were applied for DTL training (normal 254, abnormal 402), intervalidation (normal 116, abnormal 157) and testing (normal 155, abnormality 251). The training dataset was used for adjusting common parameters (weights, biases, etc.) in the network, and the test dataset was applied for evaluating the performance of the DTL after training with some important metrics such as accuracy, specificity, and sensitivity. Images were captured through the use of common conventional desktop retinal cameras and digital retinography system Topcon and NIDEK. In this study, three licensed ophthalmologists were invited responsible for image labeling, the normal images were labeled as 0 and the abnormal images were labeled as 1. Funds images were classified between November and December 2018. The images were randomly assigned to every ophthalmologist, each ophthalmologist classified between 100 and 300 funds photographs, each image was classified more than three times. The image got two or more consistent labels will be transferred into a subgroup and made available for study, in this process, the labels outcomes were blind to each other. The senior ophthalmologist dealt with controversial image labeling. 656 fundus images were randomly selected from 929 images as the training dataset, and the remaining were considered as the intervalidation dataset. In order to improve the accuracy of image recognition with only a small number of training datasets, several data pre-processing steps were implemented for normalization and standardization. For evaluating the model performance, an independent subset of Messidor was used for the test dataset. The 366 fundus images (normal 155, abnormality 251) were randomly selected from Messidor dataset. To provide a standardized image format of the dataset for the succedent deep learning and final automated testing, all images were anonymized and saved as JPG data format and cropped black borders since convolutional neural networks are sensitive to color when extracting feature.

Data processing

Data Preprocessing can detect trends, minimize noise, underline important relationship and flatten the variable distribution in a time series[11]. In this study, several steps for data preprocessing were performed to normalize the images for variation including removing meaningless photographs where important retinal information was lost due to shooting angles, light, media opacities, etc., cropping the black edges but preserving the crucial regions, adjusting the brightness to balance the color of images, reducing noise and enhancing contrast. All dataset images resolution were 3352×3364 pixels.

For improving the accuracy of image recognition with a small database and avoiding over-fitting, data augmentation was introduced into the preprocessed data to expend the range of train data samples while keeping the prognostic features in the image. According to the characteristics of color photographs and the convolution neural networks, it is highly invariant in the form of rotation, mirroring, etc.[12]. Figure 2 shows the process of train dataset augmentation in Python. Parameter probability is the ratio of the images that perform the operation to the input images. Data augmentation was introduced into the original small dataset to increase training data samples. After data augmentation, the train dataset were expended to 7,000 images, including 3,500 normal and 3,500 abnormal fundus images, respectively.

Structure of DTL

Inception-ResNet-v2 is an open-source framework with prior training by employing various object images, which has been widely used in many fields. In this study, our deep learning model used Inception-ResNet-v2 architecture to achieve transfer learning. It can help to overcome the difficulties of obtaining large manually labeled datasets and reduce the computational costs. Our model demands relatively low computation performance while maintaining effective classification results. On this basis, the source pre-training model on the large-scale data set was transferred to the target small data set, and the model weights and image features except the last two layers were extracted as the input of the new dense layer and softmax layer to finish our specific task. Then we fine-tuned the convolutional layers by unfreezing and updating the pre-trained weights to classify medical images. In the target task, a modified softmax layer output two categories (Fig. 3). The exponentially decay learning rate[13] can asymptotically reduce the learning rate, to stable the model in the later stage of training. Adam optimizer is an adaptive learning rate optimization algorithm that is specifically designed for training deep neural networks. In this study, the transferred Inception-ResNet-v2 used an Adam optimizer and exponentially decaying learning rate. The model was saved for evaluating when training was continued at 100 epochs.

Statistical analysis

Our model was implemented on an Ubuntu 16.04 computer with one graphical processing units (NVidia GeForce GTX 1080 ti). The deep transfer learning model was implemented by TensorFlow1.12 and Python 3.6. The performance of the model was evaluated based on standard classification measures: accuracy (ACC), sensitivity (TPR), specificity (TNR), receiver operating characteristic curve (ROC) which used the probability values obtained for each sample predicted by the model and the area under the curve (AUC).

Results

The manual classification of retinal fundus images was completed in November and December 2018, the DTL training and validation were completed in January 2019. Fig. 4 shows the training process performance of the model. The accuracy of the training increased fast and ran to a subsequent plateau after around 30,000 training steps. As the training continues, a learning rate lower than what we initially set is more favorable, therefore, it is beneficial that we used an exponential decay learning rate.

Internal validation performance of the model is presented in Fig. 5. The performance of internal validation dataset (Normal 116, Abnormal 157), the AUC, sensitivity, accuracy, and specificity of the DTL for correctly classifying funds images were 0.997, 97.41%, 97.07% and 96.82%, respectively. 273 images were randomly selected from the test dataset to validate the performance of the DTL. The performance of DTL correctly classified the test dataset. The AUC, sensitivity, accuracy and specificity of the DTL were 0.926, 88.17%, 87.18% and 86.67%, respectively(Fig. 6).

Table 1 shows the characterizes of misclassified photographs. The false negative cases(n = 5) of internal validation dataset include peripheral retinal micro lesions(n = 2), micro maculopathy(n = 1), high myopic fundus(n = 2). The false positive cases of internal validation dataset are 3. The false negative cases(n = 24) of testing dataset including high myopic fundus(n = 17), peripheral retinal micro lesions(n = 2), micro vascular lesions(n = 2), optic neuritis(n = 2), and congenital optic neuropathy(n = 1). The partial prediction results of the deep transfer learning model in detecting abnormal fundus images by comparing with the images true state were listed in Fig.7.

Discussion

In this study, the DTL model has achieved robust performance in abnormal funds images detection, the AUC, sensitivity, accuracy and specificity of the DTL were 0.926, 88.17%, 87.18% and 86.67%, respectively, in an independent subset of test dataset.

AI-based automated detection of retinal diseases using deep learning and transfer learning systems has been reported by several studies. The initial focus was on deep learning technology. Ting et al. [14] validated their DLS using 494,661 retinal images, demonstrating the DLS had high sensitivity and specificity for identifying diabetic retinopathy and related eye diseases, for the detection of any DR (AUC = 0.94–0.96); for possible glaucoma, AUC was 0.942; for AMD, AUC was 0.931. Similarly, Li et al[15] describe the development and validation of an artificial intelligence-based in 71,043 retinal images acquired from a web-based, deep learning algorithm for the detection of referable diabetic retinopathy. Testing against the independent multiethnic data set achieved an AUC, sensitivity, and specificity of 0.955, 92.5%, and 98.5%. Stevenson et al. [16]showed their proof-of-concept AI system performance with 4435 images. The classifiers were for AMD and vascular occlusion, both with accuracies of 99.1%, sensitivities over 99%, and specificities of 88.9 %. In contrast to these studies, our independent testing performance, the AUC, sensitivity, accuracy and specificity of the DTL were 0.926, 88.17%, 87.18% and 86.67%, the results are relatively low. It may be attributed to the outputs of our model were just divided into normal groups and abnormal groups, the latter including a multitude of disease states, thus some rare and micro-lesions were failed to detect by DTL. Previous studies demonstrated that AI will become a tool to quickly and reliably detect and diagnose eye diseases based on medical imageology. The AI-based DL could be used with high sensitivity and accuracy in detection and identifying fundus diseases. The application of AI in ophthalmology may increase the accessibility and achieve high efficiency in large scale eye diseases screening programs. Even though some studies have shown outstanding research results, some limitations should be considered. First, most of the studies required a large manually labeled data set to train and validate. It takes a lot of time, manpower, and material resources. The diagnosis varies depending on the region. Second, more thorough research of false negative values should be performed to recognize feature and relevancy. By comparison, our study is, to our knowledge, the first study to develop a DTL to detecting abnormal fundus images.

The deep transfer learning classification has been used for many years in disease screening researches. Santin et al[17]. performed transfer learning to characterize the abnormal cartilage by using a pre-trained neural network VGG16 and adapted the final layers to a binary classification problem. The AUC, sensitivity and specificity of their study were 0.72,83%,64%, respectively. In an independent sample of 189 new thyroid images resulted in an AUC of 0.70. Similarly, Heisler M, et al. [18]demonstrated three different transfer learning methods to identify the cones in a small set of AO-OCT images using a base network trained on AO-SLO images which all obtained results similar to that of a manual rater. Using the results from the Fine-Tuning (Layer 5) method, they calculated four different cone mosaic parameters which were similar to the results found in AO-SLO images showing the utility of their method. Christopher et al [19] have testified that deep learning methodologies have high diagnostic accuracy for identifying fundus photographs with glaucomatous damage to the ONH in a racially and ethnically diverse population. The best performing model was the transfer learning ResNet architecture and achieved an AUC of 0.91 in identifying glaucomatous optic neuropathy (GON) from fundus photographs, outperforming previously published accuracies of automated systems for identifying GON in fundus images.

In this study, the reasons of false negative cases of testing dataset were analyzed. High myopic fundus approximately accounted more than half of all false negative cases. These results could contribute to that our experts labeled mild myopic fundus as normal. Therefore, the model confused with the mild myopic fundus images and pathologic myopic images. In the same way, the false positive cases include mild myopic fundus. Other reasons of false negative included peripheral retinal micro-lesions, vascular micro-lesions, optic neuritis and congenital optic neuropathy.

This study presented an automated screening model which was trained with a relatively smaller number of fundus images, it can attain clinically acceptable performance in abnormal fundus images detection, its will benefit the medical institutions with no retinopathy screening program or a lack of experienced ophthalmologists. Whatsmore, the study shows our proposed model with high accuracy and reproducibility in detecting abnormal fundus images, even though it trained with a limited dataset. The DTL will permit users utilizing amounts of relation labeled graph datas to construct a detection model for the target image data. In this study, the transfer learning algorithm shows a well-applied prospect in community health care center for screening retinal disease. The techniques described in this study, with a great potential, apply in other medical fields images classification.

DTL is surprisingly effective in image classification. However, our study in its current state has several limitations. First, due to a training set in which our experts labeled mild myopic fundus as normal, the DTL trained on this set accessed a higher than normal prior probability for eye diseases detection, which may cause a high false-negative rate. Second, our study dataset is not large and includes only patients from a local clinical setting. At present, the algorithm cannot be independent or matched with professional evaluation, but it can provide abnormal fundus images with obvious diagnosis so that opthalmologists can focus on more difficult cases.

Conclusions

In conclusion, the current project demonstrated that deep transfer learning presented a promising future in the diagnosis of various disease with higher accuracy and robustness based on multi-domain data, but current studies generally take data from a single auxiliary domain. In the future work, we will be dedicated to adding more auxiliary domain information to our model and explore a screening algorithm for classifying retinal pathologic lesions and providing treatment recommendations. Further steps are improving this method and validating and evaluating its applicability in the community health care center.

Declarations

Acknowledgements

None.

Authors' contributions

All authors participated in design of the study. YY, XC and CFW analyzed and interpreted the data. YY and XC were major contributors in writing manuscript.CFW,PFZ,XBZ and YFH supervised the manuscript. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Chang-Fan Wu.

Funding

This study was supported in part by Natural Science Foundation of China (Grant No.81700867), Natural Science Foundation of Anhui province, China (Grant No.1808085MH253).The funders is Peng-Fei Zhang.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author (wucangfan@sina.com) on reasonable request.

Ethics approval and consent to participate

This retrospective study was approved by the institutional review board of Department of Ophthalmology, Yijishan Hospital of Wannan Medical College. It was conducted in accordance with all relevant requirements of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Ophthalmology, Yijishan Hospital of Wannan Medical College, Wuhu 241000, China.

[2]College of Physics and Electronic Information, Anhui Normal University, 241000, China.

References

1.Song E, Qian DJ, Wang S, Xu C, Pan CW, Refractive error in Chinese with type 2 diabetes and its association with glycaemic control. Clinical & experimental optometry, 2018. 101(2): p. 213–219.

2.Flaxman SR, Bourne RRA, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, Das A, Jonas JB, Keeffe J, Kempen JH, Leasher J, Limburg H, Naidoo K, Pesudovs K, Silvester A, Stevens GA, Tahhan N, Wong TY, Taylor HR, Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. The Lancet. Global health, 2017. 5(12): p. e1221-e1234.

3.Subburaman GB, Hariharan L, Ravilla TD, Ravilla RD, Kempen JH, Demand for Tertiary Eye Care Services in Developing Countries. American journal of ophthalmology, 2015. 160(4): p. 619–27.e1.

4.Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, Lee A, Louw V, Anderson J, Liew G, Bolter L, Srinivas S, Nittala M, Sadda S, Taylor P, Rudnicka AR, Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. Ophthalmology, 2017. 124(3): p. 343–351.

5.Ahn JM, Kim S, Ahn KS, Cho SH, Lee KB, Kim US, A deep learning model for the detection of both advanced and early glaucoma using fundus photography. PloS one, 2018. 13(11): p. e0207982.

6.Kucur ŞS, Holló G, Sznitman R, A deep learning approach to automatic detection of early glaucoma from visual fields. PloS one, 2018. 13(11): p. e0206081.

7.Burlina, P.M., Joshi, N., Pekala, M., Pacheco, K. D., Freund, D. E., & Bressler, N. M.. (2017). Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. JAMA Ophthalmology.

8.Wang J, Ju R, Chen Y, Zhang L, Hu J, Wu Y, Dong W, Zhong J, Yi Z, Automated retinopathy of prematurity screening using deep neural networks. EBioMedicine, 2018. 35(undefined): p. 361–368.

9.Samala RK, Chan HP, Hadjiiski LM, Helvie MA, Cha KH, Richter CD, Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Physics in medicine and biology, 2017. 62(23): p. 8894–8908.

10.Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson. How transferable are features in deep neural networks? in International Conference on Neural Information Processing Systems. 2014.

11.<Nawi N M, Atomi W H, Rehman M Z. The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks[J]. Procedia Technology, 2013, 11_32–39..pdf>.

12.De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, van den Driessche G, Lakshminarayanan B, Meyer C, Mackinder F, Bouton S, Ayoub K, Chopra R, King D, Karthikesalingam A, Hughes CO, Raine R, Hughes J, Sim DA, Egan C, Tufail A, Montgomery H, Hassabis D, Rees G, Back T, Khaw PT, Suleyman M, Cornebise J, Keane PA, Ronneberger O, Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine, 2018. 24(9): p. 1342–1350.

13.Zeiler M D, ∗. ADADELTA: AN ADAPTIVE LEARNING RATE METHOD[J]. Computer Science, 2012.

14.Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng CY, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY, Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA, 2017. 318(22): p. 2211–2223.

15.Li Z, Keel S, Liu C, He Y, Meng W, Scheetz J, Lee PY, Shaw J, Ting D, Wong TY, Taylor H, Chang R, He M, An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs. Diabetes care, 2018. 41(12): p. 2509–2516.

16.Stevenson CH, Hong SC, Ogbuehi KC, Development of an artificial intelligence system to classify pathology and clinical features on retinal fundus images. Clinical & experimental ophthalmology, 2018. undefined(undefined): p. undefined.

17.M. Santin, C. Brama, H. Thero, E. Ketheeswaran, I. El-Karoui, F. Bidault, R. Gillet, P. Gondim Teixeira, A. Blum, Detecting abnormal thyroid cartilages on CT using deep learning. Diagn Interv Imaging, 2019. 100(4): p. 251–257.

18.Heisler M, Ju MJ, Bhalla M, Schuck N, Athwal A, Navajas EV, Beg MF, Sarunic MV, Automated identification of cone photoreceptors in adaptive optics optical coherence tomography images using transfer learning. Biomedical optics express, 2018. 9(11): p. 5353–5367.

19.Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, Girkin CA, Liebmann JM, Zangwill LM, Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs. Scientific reports, 2018. 8(1): p. 16685.

Tables

Reasons	Number	Proportion (%)
Internal Validation Data		
False Nagetive		
Peripheral retinal micro lesions	2	40
Micro maculopathy	1	20
High myopic fundus	2	40
Total	5	100
False Positive		
mild myopic fundus	1	33.3
normal	2	66.7
Total	3	100
Testing Dataset		
False Nagetive		
High myopic fundus	17	70.83
Peripheral retinal micro lesions	2	8.83
Micro vascular lesions	2	8.83
Optic neuritis	2	8.83
Congenital optic neuropathy	1	4.17
Total	24	100
False Positive		
mild myopic fundus	4	36.4
normal	7	63.6
Total	11	100

Table 1.False negative and false positive images of the internal validation dataset and testing dataset.

Figures

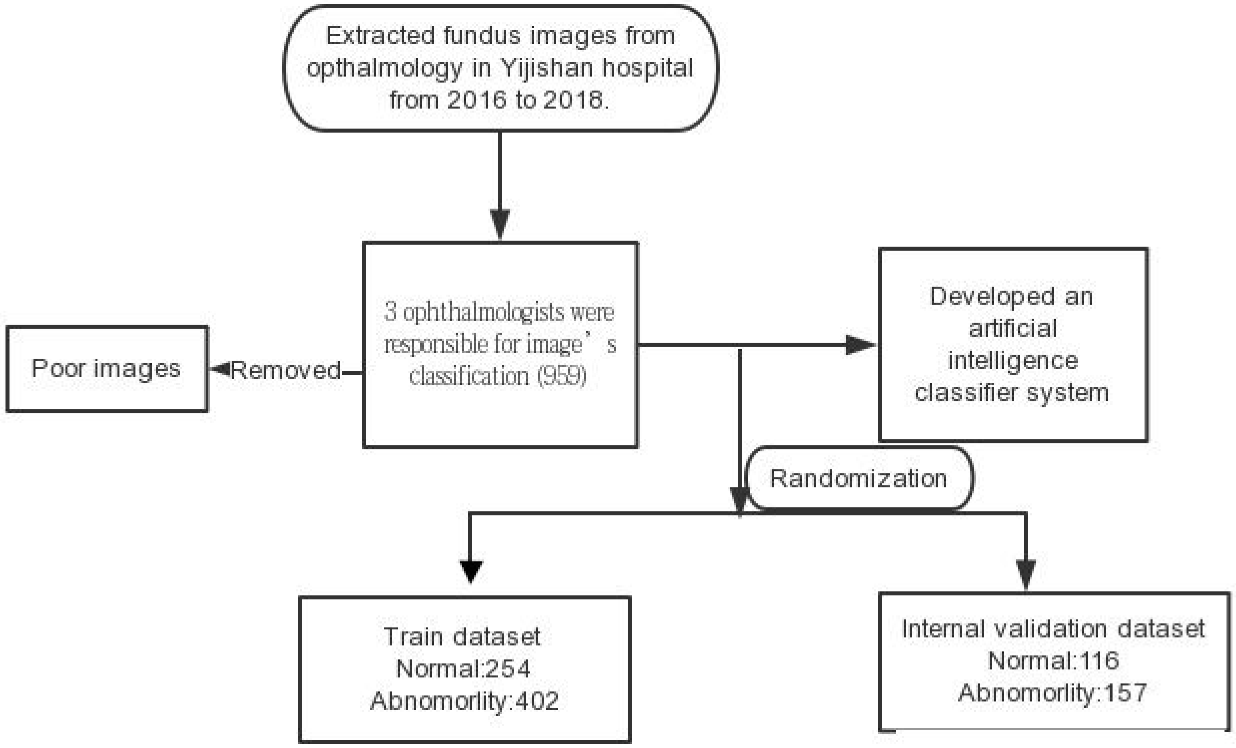

Figure 1

Development workflflow for image labeling of this study.

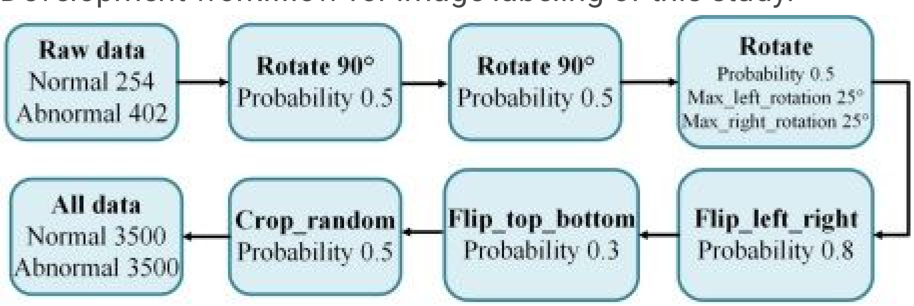

Figure 2

Original fundus images which were selected from Yijishan Hospital of Wannan Medical College were manually labled as normal and abnormality. As shown in the figure, the steps were the process of train dataset augmentation.

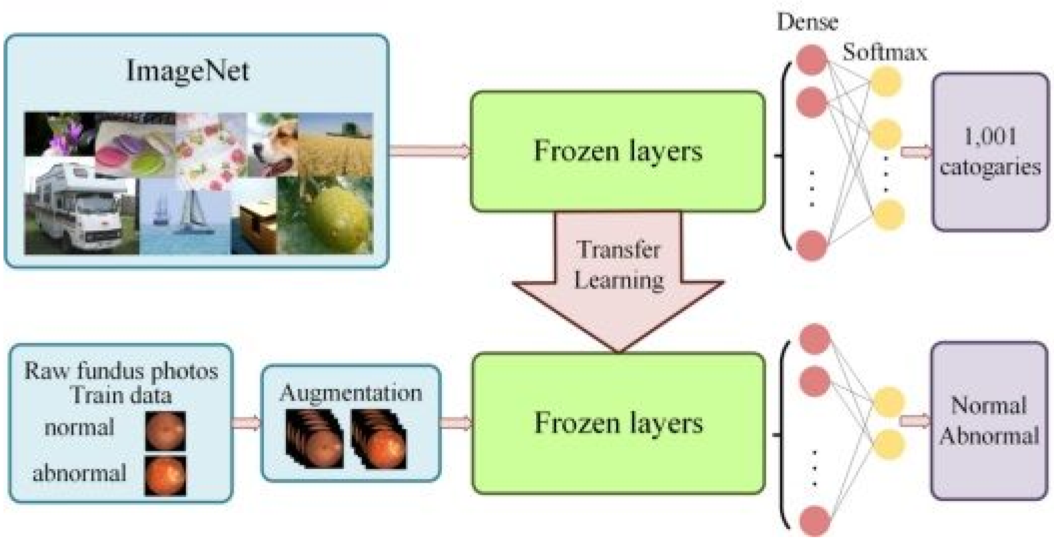

Figure 3

Illustration of the proposed procedure in this study.

Figure 4

The accuracy and the learning rate of train process.

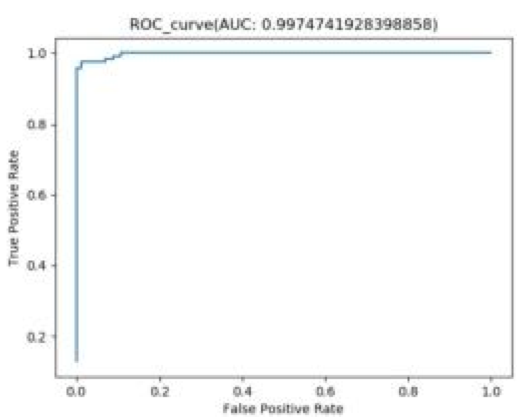

Figure 5

Receiver operating characteristic (ROC) curves of deep transfer learning in internal validation dataset.

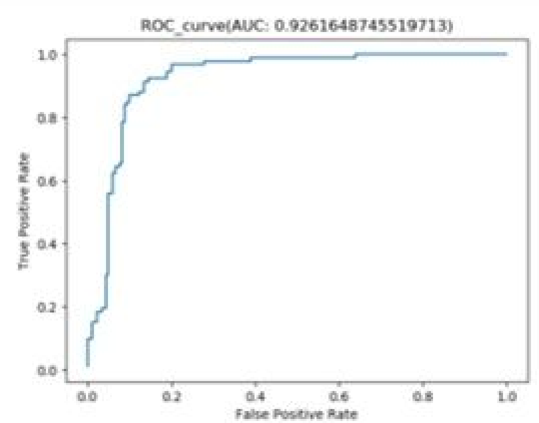

Figure 6

Receiver operating characteristic (ROC) curves of deep transfer learning in testing dataset.

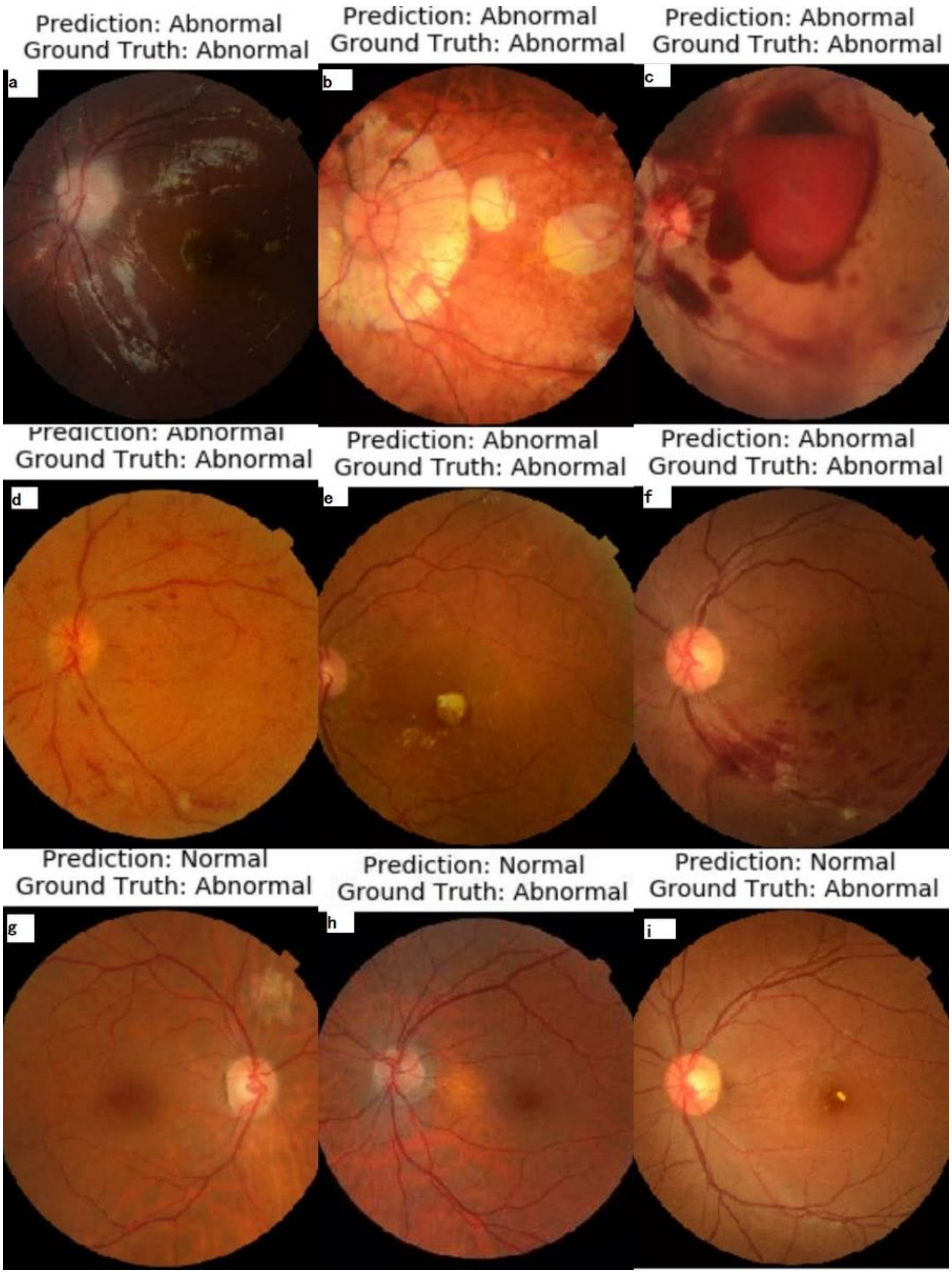

Figure 7

Examples of Fundus Images show the possibilities for the DTL: a,b,c,d,e,f,g: abnormal fundus images predicted as abnormal (true-positive); h,i,j :abnormal fundus images predicted as normal (false-negative).