

Detecting abnormal fundus images by employing deep transfer learning

Yan Yu

Wannan Medical College <https://orcid.org/0000-0002-3149-0271>

Xiao Chen

Anhui Normal University

Xiang-Bing Zhu

Anhui Normal University

Peng-Fei Zhang

Wannan Medical College

Yin-Fen Hou

Wannan Medical College

Rong-Rong Zhang

Wannan Medical College

Chang-Fan Wu (✉ wucangfan@sina.com)

Research article

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

Posted Date: February 20th, 2020

DOI: <https://doi.org/10.21203/rs.2.24133/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: To develop and validate a deep transfer learning (DTL) algorithm for detecting abnormalities in fundus images from nonmydriatic fundus photography examinations. **Methods :** A total of 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, and cataracts. We tested our model using a subset of the publicly available Messidor dataset (using 366 images) and evaluated the testing performance of the DTL model for detecting abnormal fundus images. **Results :** In the internal validation dataset (n=273 images), the AUC, sensitivity, accuracy, and specificity of the DTL for correctly classified fundus images were 0.997, 97.41%, 97.07%, and 96.82%, respectively. For the test dataset (n=273 images), the AUC, sensitivity, accuracy, and specificity of the DTL for correctly classifying fundus 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. **Key words :** Fundus images; Deep transfer learning; Developing and validation; Artificial intelligence.

Background

Retinal disease is one of the main causes of blindness worldwide, 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 the aging population and the prevalence of high myopia and diabetes, visual disabilities will continue to increase [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 the key to avoiding vision loss. Community health care centers with population concentration, comprehensive monitoring, capabilities of analyzing and evaluating individual or group health have the 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 these 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 artificial intelligence-assisted systems based on fundus photographs to screen retinal diseases. However, many of these studies have been devoted to identifying DMR[4] and glaucoma[5, 6], and studies about retinal disease recognition aiming to establish a classification of normal and abnormality in multicategorical retinal diseases have been very limited.

Artificial intelligence (AI) using machine learning algorithms, such as support vector machines (SVMs), naive Bayes classifiers and convolutional neural networks (CNNs), has received extensive attention after demonstrating that it could perform at least as well as humans in image classification tasks[4, 7]. As the digital imaging modality rapidly develops, image processing, computer vision, and machine learning are being used to automatically detect retinal lesions based on color fundus photographs. This is of great significance for the implementation of computationally assisted retinal disease detection and the promotion of large-scale screenings[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, transfer learning was a highly effective technology, especially in domains where limited data[10] were available. The essential characteristics of DTL are compared with traditional image recognition methods, which do not need to rely on manual labeling and a large quantity of labeled training data and do not require much cost and time for data collection. The purpose of this study is to develop and validate an effective transfer learning algorithm for detecting abnormal fundus photographs and to provide an accurate and timely referral by employing a small multicategorical retinal disease image database. Additionally, new insights are generated for the screening program to efficiently build a detection model with a few labeled fundus photographs and some relation graph data.

Methods

Image dataset 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. The image is 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 the optic disc is only one or none, macular region vessels cannot be distinguished. After removing 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. Figure 1 shows the workflow of this study. The images were extracted from the ophthalmic clinics, inpatients and physical examination centers in our hospital. Three datasets were applied for DTL training (normal 254, abnormal 402), internal validation (normal 116, abnormal 157) and testing (normal 155, abnormality 251). The training dataset was used to adjust common parameters (weights, biases, etc.) in the network, and the test dataset was applied to evaluate 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 the digital retinography system Topcon and NIDEK. In this study, three licensed ophthalmologists were invited for image labeling. The normal images were labeled as 0, and the abnormal images were labeled as 1. Fund images were classified between November and December 2018. The images were randomly assigned to every ophthalmologist, each ophthalmologist classified between 100 and 300 fundus photographs, and each image was classified more than three times. The images that obtained two or more consistent labels were transferred into a

subgroup and made available for study. In this process, the labeling outcomes were blind. The senior ophthalmologist dealt with controversial image labeling. A total of 656 fundus images were randomly selected from 929 images as the training dataset, and the remaining images were considered as the internal validation dataset. To improve the accuracy of image recognition with only a small number of training datasets, several data preprocessing steps were implemented for normalization and standardization. To evaluate the model performance, an independent subset of the Messidor database was used for the test dataset. The 366 fundus images (normal 155, abnormal 251) were randomly selected from the 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 the JPG data format and cropped black borders since convolutional neural networks are sensitive to color when extracting features.

Data processing

Data preprocessing can detect trends, minimize noise, underline important relationships 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, and 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 image resolutions were $3,352 \times 3,364$ pixels.

To improve the accuracy of image recognition with a small database and avoid overfitting, data augmentation was introduced into the preprocessed data to expand the range of training data samples while keeping the prognostic features in the image. According to the characteristics of color photographs and convolutional neural networks, it is highly invariant in the form of rotation, mirroring, etc.[12]. Figure 2 shows the process of training dataset augmentation in Python. The parameter probability is the ratio of the images that perform the operation on the input images. Data augmentation was introduced into the original small dataset to increase the number of training data samples. After data augmentation, the training dataset was expanded to 7,000 images, including 3,500 normal and 3,500 abnormal fundus images.

Structure of DTL

Inception-ResNet-v2 is an open-source framework with prior training that employs various object images and has been widely used in many fields. Inception-ResNet-v2 is a costlier hybrid Inception version with significantly improved classification performance[22]. The Inception architecture has been shown to achieve very good performance at a relatively low computational cost. Residual connections have also been proven to be more accurate on the classification task and can learn faster[22, 23]. Inception-ResNet-v2 has deeper layers and adds connections between the Inception-ResNet modules (Inception-ResNet-A, Inception-ResNet-B and Inception-ResNet-C) and the reduction modules (Reduction-A, Reduction-B). More

details can be learned in the literature[22]. The classification accuracy of Inception-ResNet-v2 outperformed any other architecture on benchmark datasets at that time.

In this study, the Inception-ResNet-v2 architecture was applied 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 computational performance while maintaining effective classification results. To achieve the transfer, we remove the dense layer and the softmax layer of the pretrained network. We need to eliminate the last two layers because the dimensions of the dense layer and softmax layer must be equal to the number of classes in our task. Then, we add adaptation layers to construct the new architecture. On this basis, the source pretraining model on the large-scale dataset was transferred to the target small dataset, and the model weights and image features, except for the last two layers, are extracted as the input of the new dense layer and the softmax layer to finish our specific task. Then, we fine-tune the convolutional layers by unfreezing and updating the pretrained weights to classify medical images. In the target task, a modified softmax layer outputs two categories (Fig. 3). The exponential decay learning rate[13] can asymptotically reduce the learning rate to stabilize the model in the later stage of training. The 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 uses an Adam optimizer and exponentially decaying learning rate with an initial learning rate of 0.0001 and a decay rate of 0.7 to minimize the loss. The model is saved for evaluation when training continues at 100 epochs.

Statistical analysis

Our model was implemented on an Ubuntu 16.04 computer with one graphical processing unit (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), and the 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, and DTL training and validation were completed in January 2019. Figure 4 shows the training process performance of the model. The accuracy of the training increased rapidly and ran to a subsequent plateau after approximately 30,000 training steps. As the training continued, a learning rate lower than what we initially set was more favorable; therefore, it was beneficial that we used an exponential decay learning rate.

The internal validation performance of the model is presented in Fig. 5. The performance of the internal validation dataset (normal 116, abnormal 157) and the AUC, sensitivity, accuracy, and specificity of the DTL for correctly classifying fundus images were 0.997, 97.41%, 97.07%, and 96.82%, respectively. A total

of 273 images were randomly selected from the test dataset to validate the performance of the DTL. The performance of the 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 characteristics of misclassified photographs. The false-negative cases (n = 5) of the internal validation dataset included peripheral retinal microlesions (n = 2), micro maculopathy (n = 1), and high myopic fundus (n = 2). The false-positive cases of the internal validation dataset were 3. The false-negative cases (n = 24) of the testing dataset included high myopic fundus (n = 17), peripheral retinal microlesions (n = 2), microvascular 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 comparison with the image's true state are listed in Fig. 7.

Discussion

In this study, the DTL model achieved robust performance in abnormal fundus image detection, and 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 the test dataset.

AI-based automated detection of retinal diseases using deep learning and transfer learning systems has been reported in several studies. The initial focus was on deep learning technology. Ting et al. [14] validated their deep learning system (DLS) using 494,661 retinal images, demonstrating that 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, the AUC was 0.942; for AMD, the AUC was 0.931. Similarly, Li et al. [15] described the development and validation of an artificial intelligence-based method 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 dataset achieved an AUC, sensitivity, and specificity of 0.955, 92.5%, and 98.5%, respectively. Stevenson et al. [16] showed their proof-of-concept AI system performance with 4,435 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 the above studies, our independent testing performance, the AUC, sensitivity, accuracy, and specificity of the DTL were 0.926, 88.17%, 87.18%, and 86.67%, respectively, and the results were relatively low. This may be attributed to the outputs of our model being divided into normal groups and abnormal groups, the latter including a multitude of disease states; thus, some rare and microlesions failed to be detected by DTL. Previous studies have demonstrated that AI will become a tool to quickly and reliably detect and diagnose eye diseases based on medical imageology. AI-based DL could be used with high sensitivity and accuracy in the detection and identification of fundus diseases. The application of AI in ophthalmology may increase accessibility and achieve high efficiency in large-scale eye disease screening programs. Although some studies have shown outstanding research results, some limitations should be considered. First, most of the studies required a large manually labeled dataset to train and validate, which requires considerable 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 features and relevance. By comparison, our study is, to our knowledge, the first to develop a DTL to detect abnormal fundus images by employing a small dataset.

The deep transfer learning classification has been used for many years in disease screening research. Santin et al [17]. performed transfer learning to characterize the abnormal cartilage by using a pretrained 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%, and 64%, respectively. In an independent sample of 189 new thyroid images, the AUC was 0.70. Compared with this study, they all deployed a small dataset, but the performance of the Inception-ResNet-v2 architecture was significantly better than that of the VGG16. 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 that were similar to the results found in AO-SLO images, showing the utility of their method. Christopher et al. [19] demonstrated 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, which 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. These deep learning systems showed that the models can learn faster by employing transfer learning with fewer data.

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

This study presented an automated screening model that was trained with a relatively smaller number of fundus images. It can attain clinically acceptable performance in abnormal fundus image detection and will benefit medical institutions with no retinopathy screening program or a lack of experienced ophthalmologists. Additionally, 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 to utilize relation-labeled graph data 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 centers for screening retinal disease. The techniques described in this study, with great potential, apply in other medical field image classifications.

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 disease 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 diagnoses so that ophthalmologists 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 diseases with higher accuracy and robustness based on multidomain data. In 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 include improving this method and validating and evaluating its applicability in the community health care center.

Declarations

Acknowledgements

None.

Authors' contributions

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

Corresponding author

Correspondence to Chang-Fan Wu. Department of Ophthalmology Yijishan Hospital of Wannan Medical College 92 West Zheshan Rd, Wuhu, 241001 P. R. China; wucangfan@sina.com

Phone: 86-139-0963-2351

Funding

This study was supported in part by the Natural Science Foundation of China (Grant No.81700867), Natural Science Foundation of Anhui province, China (Grant No.1808085MH253). The funder 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 a reasonable request.

Ethics approval and consent to participate

This retrospective study was approved by the institutional review board of the Department of Ophthalmology, Yijishan Hospital of Wannan Medical College. It was conducted following 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

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

²College 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 ŞŞ, 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

Due to technical limitations, all table files are only available for download from the Supplementary Files section.

Figures

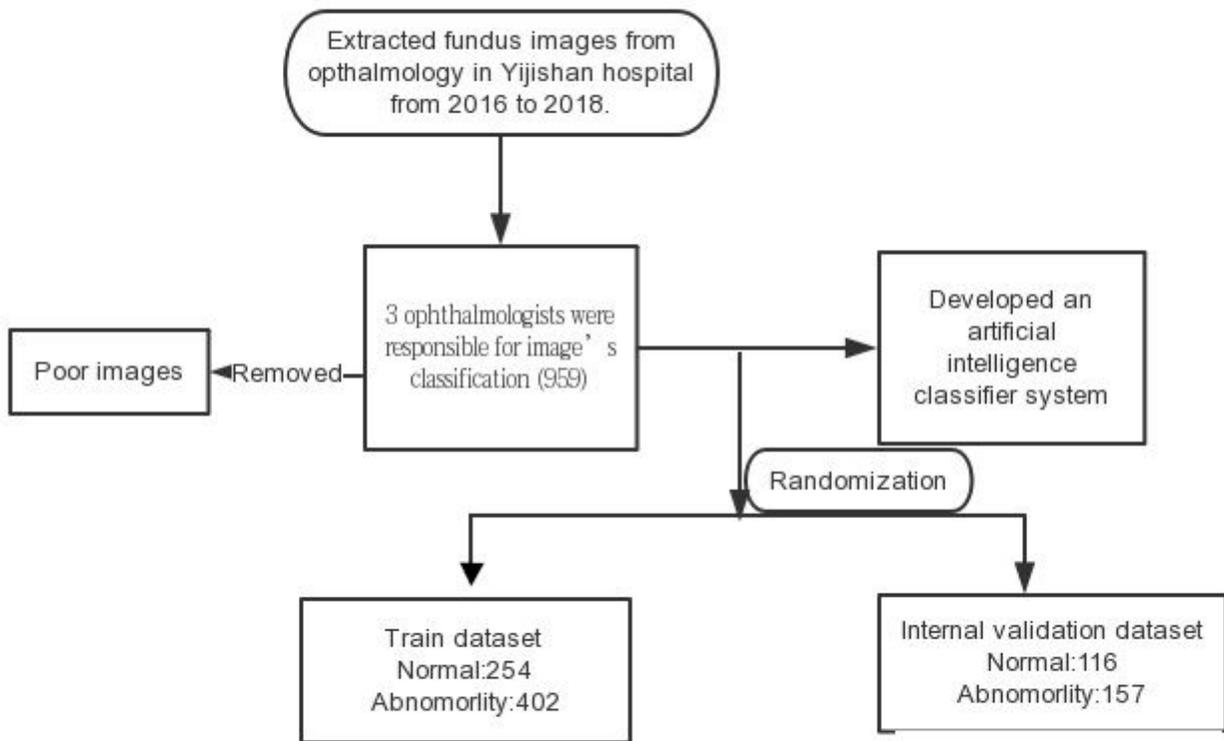


Figure 1

Development workflow for image labeling in this study.

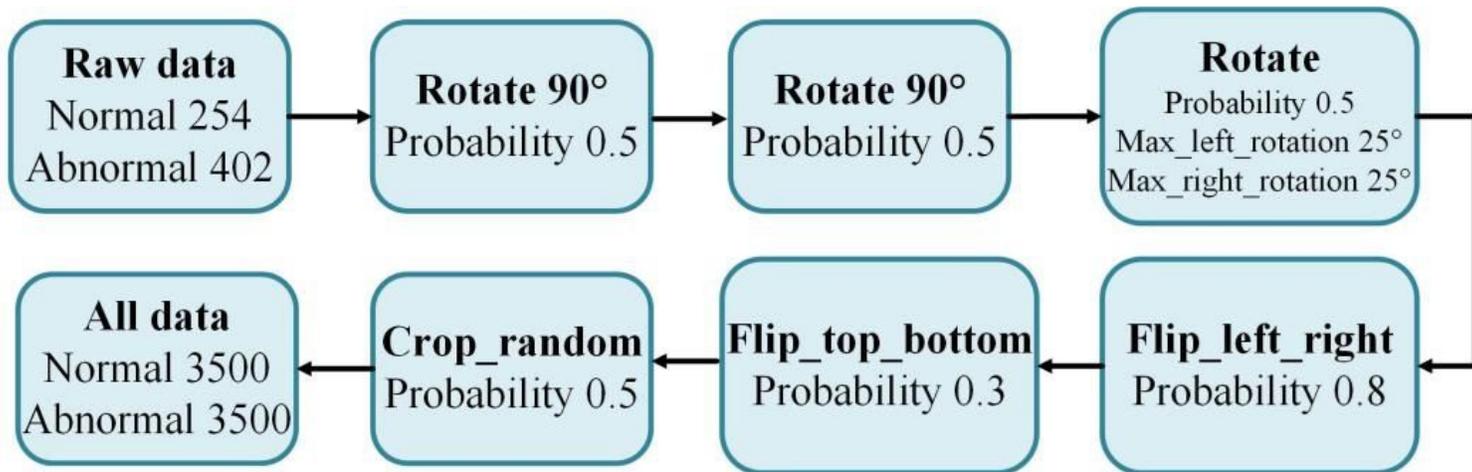


Figure 2

Original fundus images that were selected from Yijishan Hospital of Wannan Medical College were manually labeled as normal and abnormal. As shown in the figure, the steps were the process of training dataset aggregation.

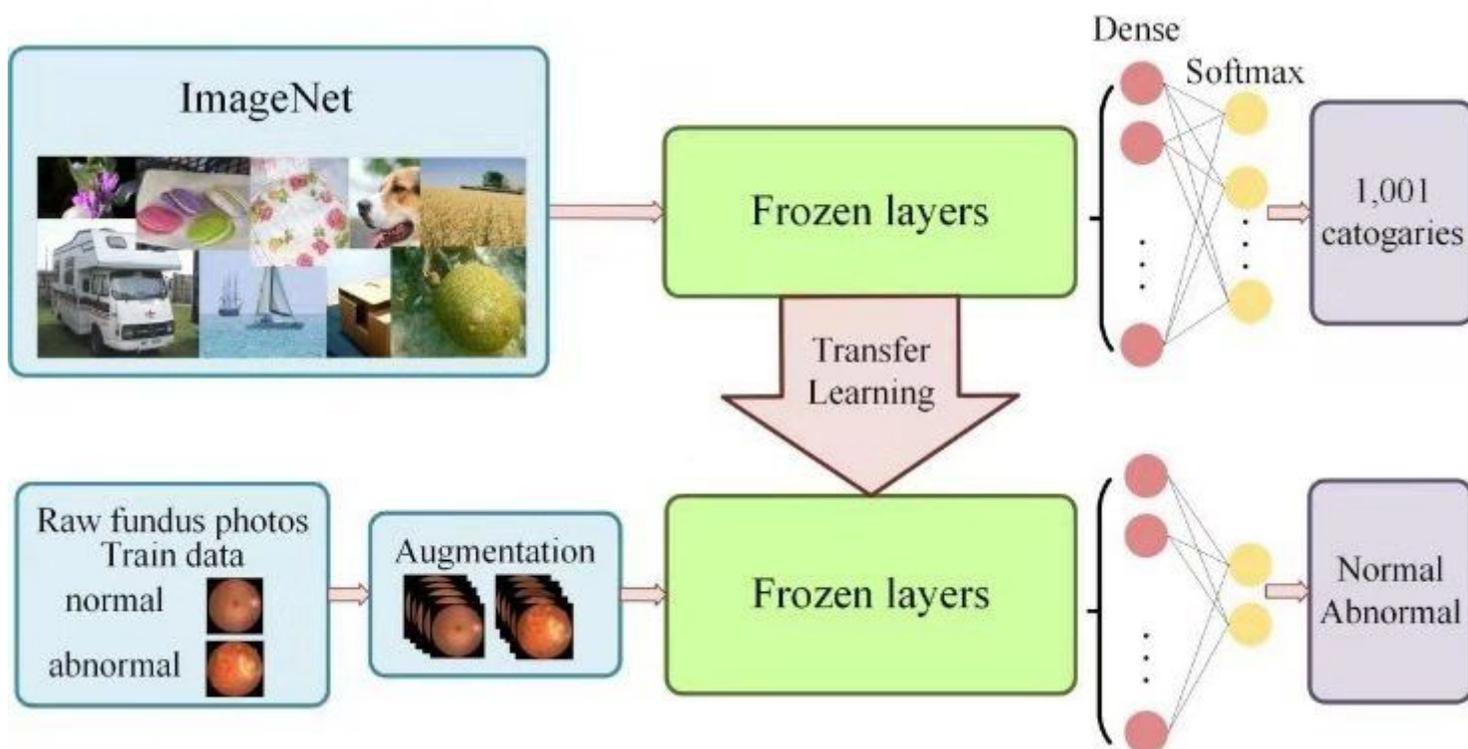


Figure 3

Illustration of the proposed procedure in this study.



Figure 4

The accuracy and the learning rate of the training process.

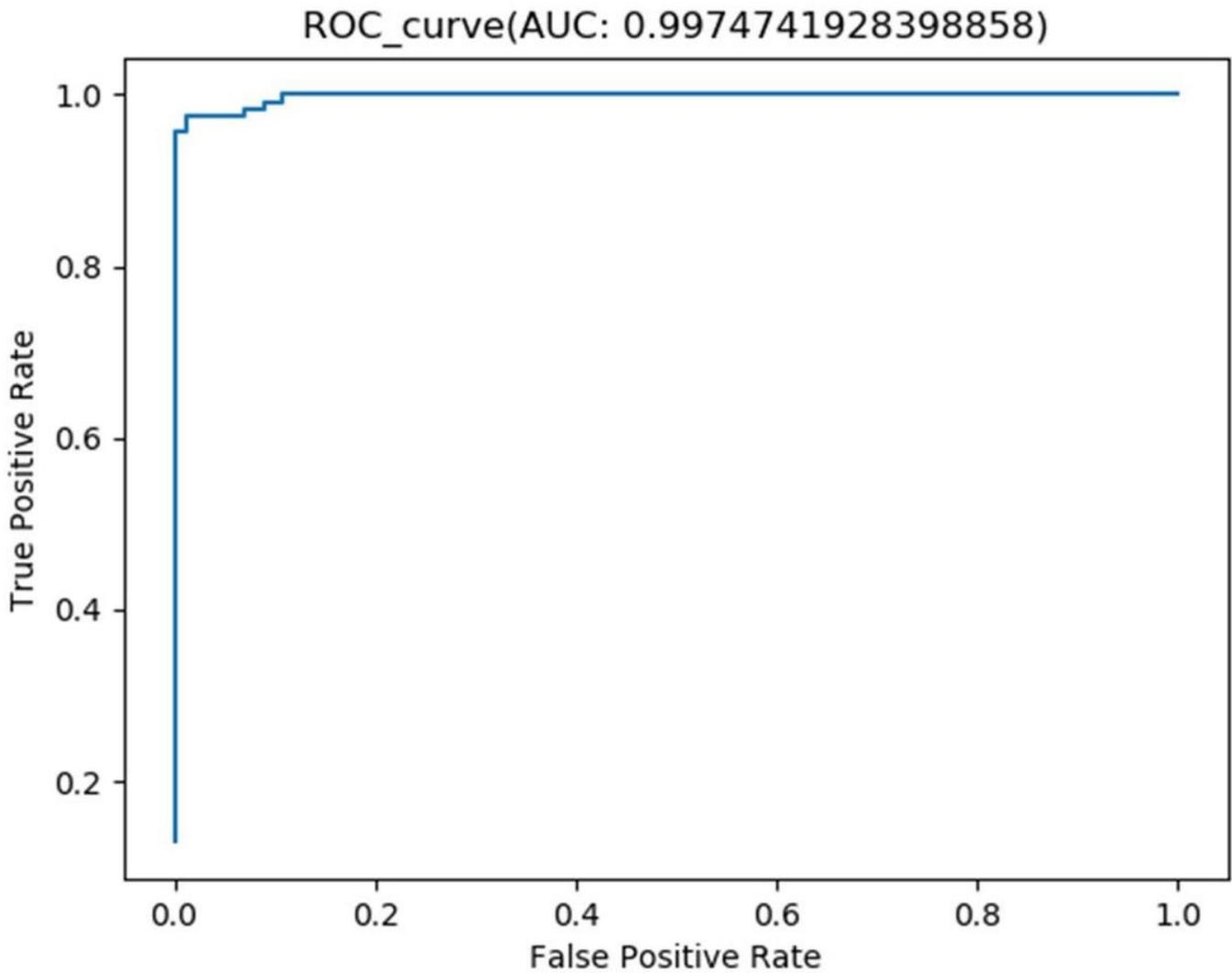


Figure 5

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

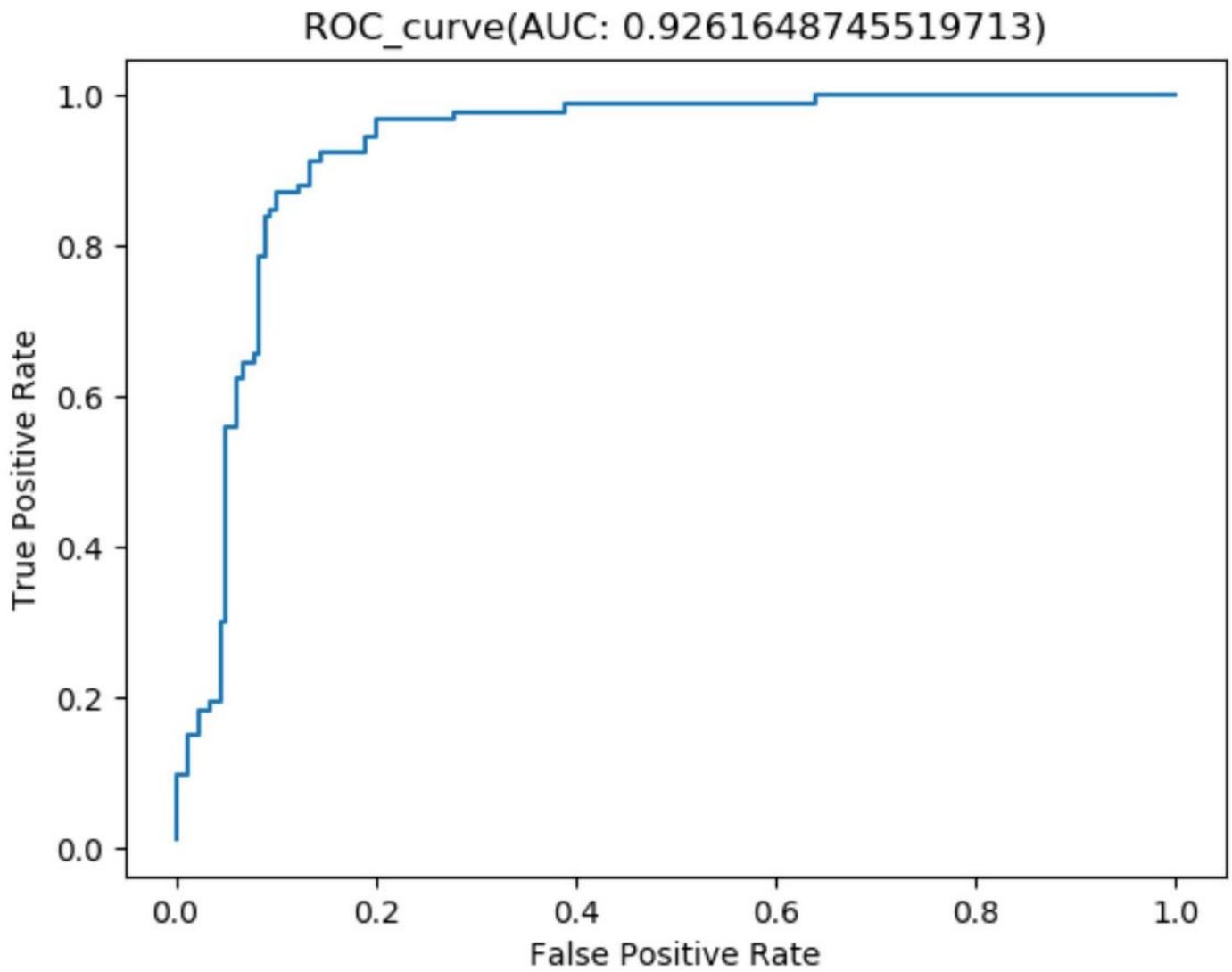


Figure 6

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

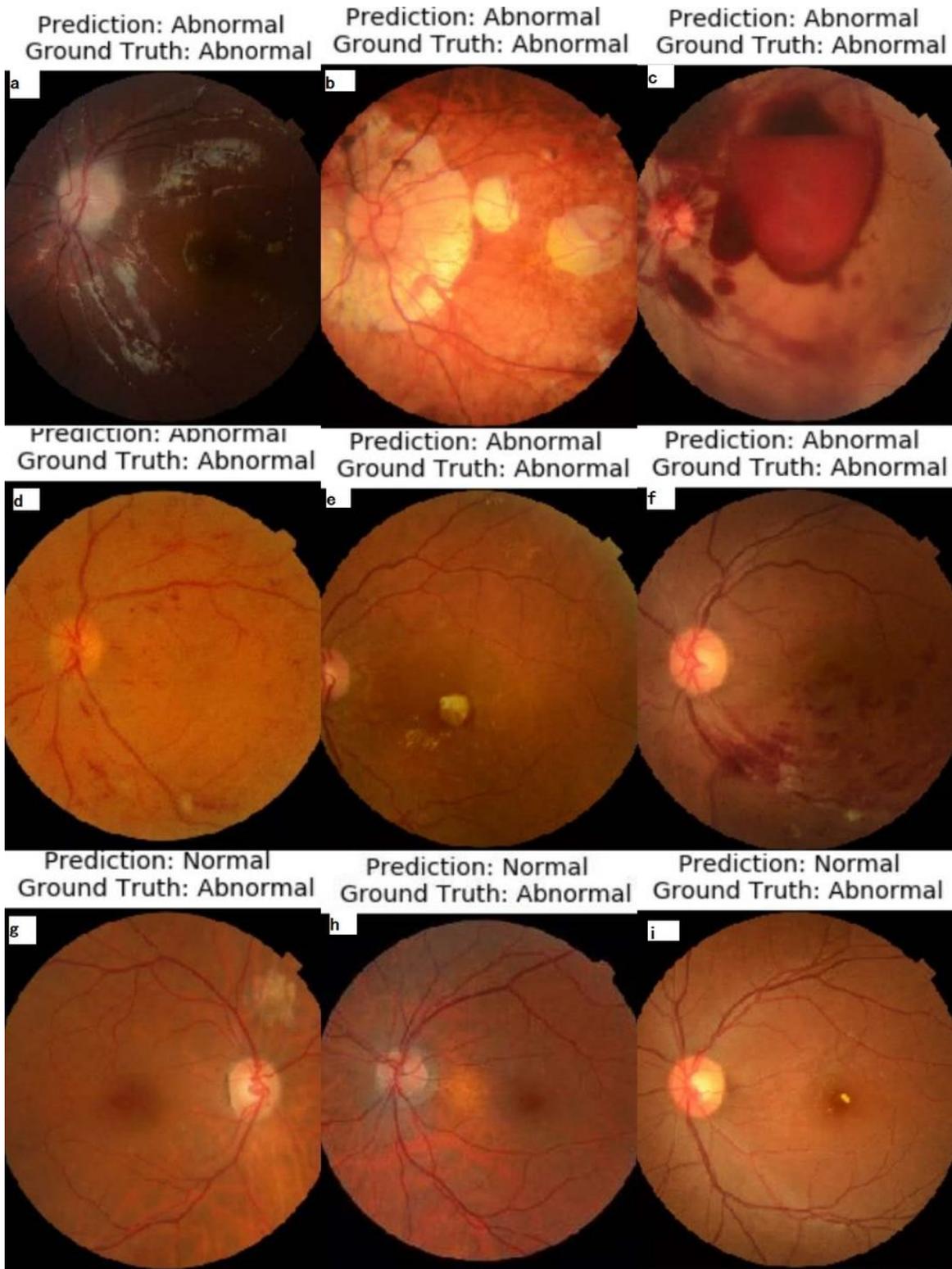


Figure 7

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

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [SupplementaryContent.docx](#)