Performance of Deep Transfer Learning for Detecting Abnormal Fundus Images

Yan Yu[1], Xiao Chen[2], XiangBing Zhu[2], PengFei Zhang[1], YinFen Hou[1], RongRong Zhang[1], ChangFan Wu[1]

[1]Department of Ophthalmology, Yijishan Hospital of Wannan Medical College, Wuhu, China, [2]Optoelectronic Technology Research Center, Anhui Normal University, Wuhu, China

Abstract

Purpose: To develop and validate a deep transfer learning (DTL) algorithm for detecting abnormalities in fundus images from non-mydriatic fundus photography examinations.

Methods: A total of 1295 fundus images were collected to develop and validate a DTL algorithm for detecting abnormal fundus images. After removing 366 poor images, the DTL model was developed using 929 (370 normal and 559 abnormal) fundus images. Data preprocessing was performed to normalize the images. The inception-ResNet-v2 architecture was applied to achieve transfer learning. 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 area under the curve (AUC), sensitivity, accuracy, and specificity of 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: DTL showed high sensitivity and specificity for detecting abnormal fundus-related diseases. Further research is necessary to improve this method and evaluate the applicability of DTL in community health-care centers.

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

INTRODUCTION

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 clinical examination with the help of eye experts in retinal vessels, optic discs, the fovea, and lesions. As the prevalence of vision disabilities increases,[2] early detection and effective treatment are the keys to avoiding vision loss. A community health-care center with population concentration, comprehensive monitoring, and capabilities of analyzing and evaluating individual or group health can provide 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

Access this article online

Quick Response Code:

Website:
www.jcurrophthalmol.org

DOI:
10.4103/JOCO.JOCO_123_20

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 (AI)-assisted systems, machine learning, and deep learning, based on fundus photographs, to screen retinal diseases. However, many of these studies have been devoted to identifying diabetic macular edema, age-related macular degeneration (AMD), and glaucoma,[4-6] and studies of retinal disease recognition to establish a classification of normal and abnormality in multicategorical retinal diseases have been very limited.[7,8]

AI using machine-learning algorithms, such as support vector machines, naive Bayes classifiers, and convolutional neural networks (CNNs), have received extensive attention after demonstrating that it can perform at least as well as humans in image classification tasks.[4,9] 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.[10] Deep transfer learning (DTL) is a new machine learning method that leverages existing knowledge to solve different but related domain problems.[11] Based on past studies, transfer learning is a highly effective technology, especially in domains where limited data are available.[12] Compared to the traditional image recognition methods, DTL does not need to rely on manual labeling and a large quantity of labeled training data, and does 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. In addition, new insights are generated for the screening program to efficiently build a detection model with a few labeled fundus photographs and some related graph data.

METHODS

Image dataset characteristics

A total of 1295 fundus images were selected in 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 images were labeled as poor quality and removed from the training and validation dataset in the following situations: blurred areas accounting for 50% or more of the image area, fovea or optic disc or both were not visible, and indistinguishable macular vessels. After removing 366 poor images, the DTL model was developed using 929 retinal fundus images (370 normal, 559 abnormal) 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 (254 normal, 402 abnormal), internal validation (116 normal, 157 abnormal), and testing (155 normal, 251 abnormal). 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 using common conventional desktop retinal cameras and the Digital Retinography System TRC-NW8F plus (Topcon Inc., Tokyo, Japan) and AFC-330 (NIDEK CO., Gamagori, Japan). In this study, three experienced ophthalmologists were invited for image labeling. The normal images were labeled as 0, and the abnormal images were labeled as 1. Fundus images were classified between November and December 2018. The images were randomly assigned to every ophthalmologist, and each image was labeled by three experts. 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 handled the cases of 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 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 testing dataset.[13] In this study, 366 fundus images (115 normal, 251 abnormal) 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 in JPG format with cropped black borders because CNNs 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.[14] 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; cropping the black edges but preserving the crucial regions; adjusting the brightness to balance the color of images; reducing noise; and enhancing contrast. Image monochromatic and contrast-limited adaptive histogram equalization algorithms were used to enhance contrast and reduce noise. The image resolution of the data was 3352×3364 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. The characteristics of color photographs and CNNs are highly invariant in terms of rotation, mirroring,

etc.[15] Two methods were applied for data augmentation: rotating the images and performing mirror image or vertical image processing. Figure 2 shows the process of training dataset augmentation in Python. The probability parameter is the ratio of the images that operate 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 7000 images, including 3500 normal and 3500 abnormal fundus images.

Structure of deep transfer learning

Inception-ResNet-v2 is an open-source framework with prior training from ImageNet pretrained networks trained for classifying 1.2 million natural color images into 1000 classes as part of the ImageNet Large Scale Visual Object Recognition Challenge and has been widely used in many fields.[16] Inception-ResNet-v2 is a costlier hybrid inception version with significantly improved classification performance. The inception architecture has been shown to achieve very good performance at a relatively low computational cost.[17] Residual connections have also been proven to be more accurate on the classification task and can learn faster.[17,18] 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 found in the literature. The classification accuracy of Inception-ResNet-v2 outperforms any other architecture on benchmark datasets.

In this study, the Inception-ResNet-v2 architecture was applied to achieve transfer learning. It can help 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 removed the dense layer and the softmax layer of the pretrained network. We eliminated 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. We then added adaptation layers to construct the new architecture. On this basis, the source pretrained 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, were extracted as the input of the new dense layer and the softmax layer to finish our specific task. We fine-tuned the convolutional layers by unfreezing and updating the pretrained weights to classify the medical images. In the target task, a modified softmax layer output two categories [Figure 3]. The exponential decay
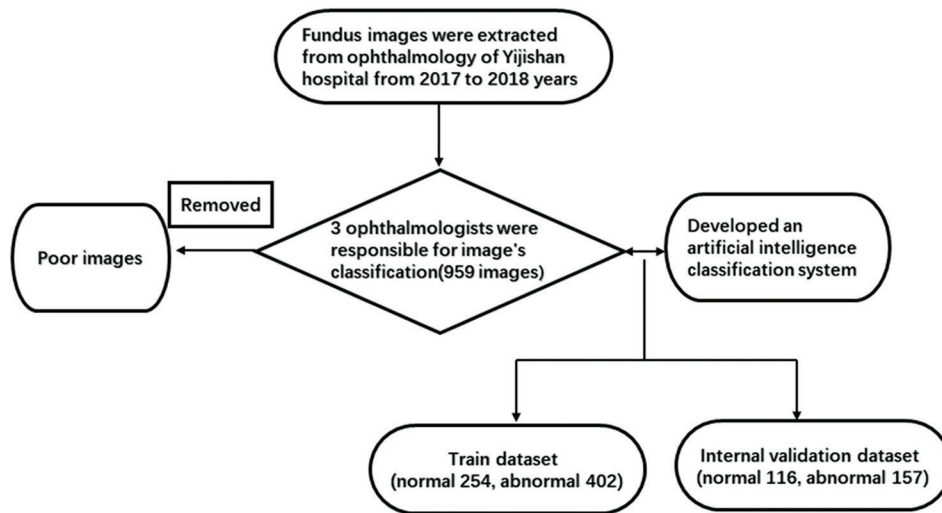

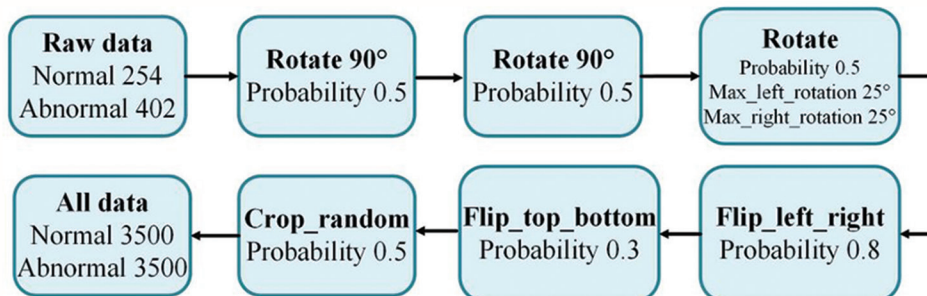**Figure 1:** Development workflow for image labeling in this study

Figure 2: The steps were the process of training dataset aggregation

learning rate can asymptotically reduce the learning rate to stabilize the model in the later stage of training.[19] An 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 was saved for evaluation when the training ran for 100 epochs. After repeated testing, the batch size was set as 256.

Statistical analysis

Our model was implemented on a computer running Ubuntu 16.04 with one graphical processing unit (NVidia GeForce GTX 1080 ti). The DTL model was implemented using TensorFlow1.12 and Python 3.6. The performance of the model was evaluated based on standard classification measures: accuracy (classification accuracy), sensitivity (true-positive rate), specificity (true-negative rate), and the receiver operating characteristic curve, 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 Figure 5. The performance of the internal validation dataset (116 normal, 157 abnormal) 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 testing dataset to validate the performance of the DTL. The DTL correctly classified the test dataset, with the AUC, sensitivity, accuracy, and specificity of the DTL being 0.926%, 88.17%, 87.18%, and 86.67%, respectively [Figure 6]. The results for some methods and tests of our fundus images are shown in Table 1. The Inception-ResNet-v2 classification performance is higher than Inception-v3.

Table 2 shows the characteristics of misclassified photographs. The false-negative cases and the false-positive cases of the internal validation dataset numbered 5 and 3, respectively. The false-negative cases of the testing dataset numbered 24. Table 3 shows the false-negative rate and false-positive rate of the testing dataset. The partial prediction results of the DTL model in detecting abnormal fundus images by comparison with the image's true state are summarized in Figure 7.

DISCUSSION

In this study, the DTL model achieved sufficient 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. This study presented an automated screening model that was trained with a 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. In addition, the study shows our proposed model has high accuracy and reproducibility in detecting abnormal fundus images.

AI-based automated detection of retinal diseases using deep learning, VGGNet-s, AlexNet, and supervised learning systems has been reported in several studies.[20-22] The initial focus
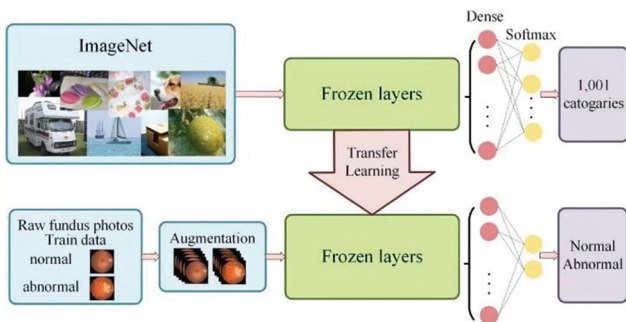

**Figure 3:** Illustration of the proposed procedure in this study

**Figure 4:** The accuracy and the learning rate of the training process

Table 1: The results for some methods and tests of our fundus images

Approaches	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
DA + inception-ResNet-v2 + TL	42.49	100	0	0.75
Inception-ResNet-v2 + FT	86.81	89.91	84.76	0.905
DA + inception-v3 (from scratch)	86.45	78.33	92.81	0.912
DA + inception-v3 + TL + FT	91.79	89.52	93.59	0.938
DA + inception-ResNet-v2 + TL + FT	97.07	97.41	96.82	0.997

AUC: Area under the curve, DA: Data augmentation, TL: Transfer learning, FT: Fine-tuning
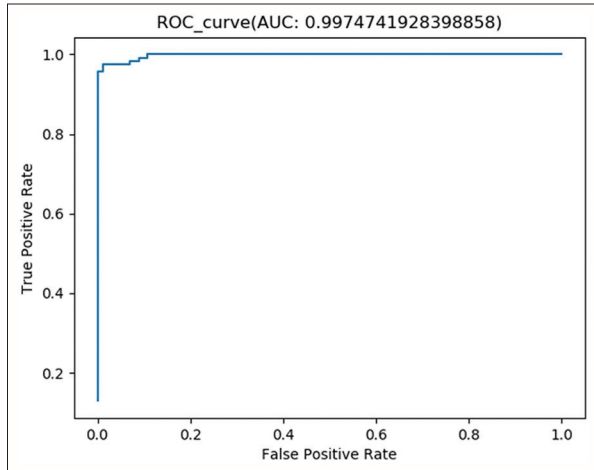
Figure 5: Receiver operating characteristic curves of deep transfer learning in the internal validation dataset
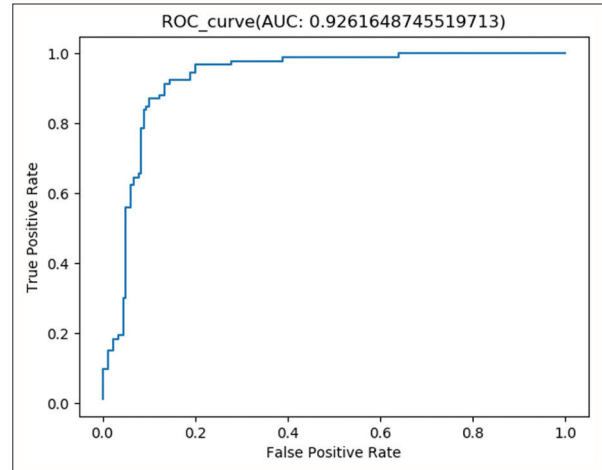
Figure 6: Receiver operating characteristic curves of deep transfer learning in the testing dataset

was on deep learning technology. Ting *et al*. validated their deep learning system (DLS) using 494,661 retinal images, demonstrating that DLS had high sensitivity and specificity for identifying diabetic retinopathy and related eye diseases for the detection of any diabetic retinopathy (AUC = 0.94–0.96); for possible glaucoma, the AUC was 0.942; for AMD, the AUC was 0.931.[23] Similarly, Li *et al*. described the development and validation of an AI-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.[24] Stevenson *et al*. 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%.[25] In contrast to the above studies, our independent testing performance in terms of 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. The comparison was among the DTL and VGGNet-s, AlexNet, and supervised learning, which are the latest deep CNNs for color fundus image classification tasks. The DTL has excellent performance in terms of classification accuracy (97.07%) and sensitivity (97.41%) under a small

number of training data. Although previous 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 the best of our knowledge, the first to develop a DTL to detect abnormal fundus images by employing a small dataset.

DTL classification has been used for many years in disease screening research. Santin *et al*. performed transfer learning to characterize abnormal cartilage 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.[26] Compared with this study, previous studies deployed a small dataset, but the performance of the Inception-ResNet-v2 architecture was significantly better than that of VGG16. Similarly, Heisler, *et al*. demonstrated three different transfer learning methods to identify the cones in a small set of adaptive optics optical coherence tomography (AO-OCT) images using a base network trained on adaptive optics scanning laser ophthalmoscopy 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

Table 2: False-negative and false-positive images of the internal validation dataset and testing dataset

Reasons	*n*	Proportion (%)
Internal validation dataset		
False-negative		
Peripheral retinal micro lesions	2	40
Micro maculopathy	1	20
High myopic fundus	2	40
Total	5	100
False-positive		
Mild myopic fundus	1	25
Normal	2	12.5
Total	3	100
Testing dataset		
False-negative		
High myopic fundus	17	70.83
Peripheral retinal micro lesions	2	8.33
Microvascular lesions	2	8.33
Optic neuritis	2	8.33
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 3: The calculation of false-negative rate and false-positive rate of testing data set

DTL	Standard	
	Positive	**Negative**
Positive	a (154)	b (11)
Negative	c (24)	d (82)
Total	a+c (178)	b+d (93)

False-negative rate=c/c+a=24/180≈13.3%, False-positive rate=
b/b+d=11/93≈11.83%. DTL: Deep transfer learning

Figure 7: Examples of fundus images show the possibilities for the deep transfer learning: (a-f) Abnormal fundus images predicted as abnormal (true-positive); (g-i) Abnormal fundus images predicted as normal (false-negative)

different cone mosaic parameters that were similar to the results found in AO-SLO images, showing the utility of their method.[27] Christopher *et al*. demonstrated that deep learning methodologies have high diagnostic accuracy for identifying fundus photographs with glaucomatous damage to the optic nerve head 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.[28] These transfer learning systems showed that the models can learn faster by employing transfer learning with fewer data. DTL will permit users to utilize relation-labeled graph data to construct a detection model for the target image data.

In this study, the reasons for false-negative cases of the testing datasets were analyzed. High myopic fundus accounted for 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 pathological myopic images. In the same way, false-positive cases included mild myopic fundus. Other reasons for false-negatives included peripheral retinal microlesions, vascular microlesions, optic neuritis, and congenital optic neuropathy.

DTL is surprisingly effective in image classification. However, in its current state, our study has several limitations. First, due to a training set in which our experts labeled mild myopic fundus as normal, the DTL model trained on this set achieved 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. In this study, a wide variety of ocular disease images may affect the performance of the algorithm. The algorithm's output divides the photographs into normal or abnormal groups but cannot reach a diagnosis for a specific disease.

In conclusion, the current project demonstrated that DTL presents a promising future in the diagnosis of various diseases with higher accuracy and efficiency based on color fundus image data. In future work, we will be dedicated to adding more auxiliary domain information to our model and exploring a screening algorithm for classifying retinal pathological lesions and providing treatment recommendations. Further steps include improving this method and validating and evaluating its applicability in the community health-care center.

Conflicts of interest

There are no conflicts of interest.

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. Clin Exp Optom 2018;101:213-9.
2. Flaxman SR, Bourne RR, Resnikoff S, Ackland P, Braithwaite T, Cicinelli MV, *et al*. Global causes of blindness and distance vision impairment 1990-2020: A systematic review and meta-analysis. Lancet Glob Health 2017;5:e1221-34.
3. Subburaman GB, Hariharan L, Ravilla TD, Ravilla RD, Kempen JH. Demand for Tertiary Eye Care Services in Developing Countries. Am J Ophthalmol 2015;160:619-270.
4. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, *et al*. Automated Diabetic Retinopathy Image Assessment Software: Diagnostic accuracy and cost-effectiveness compared with human graders. Ophthalmology 2017;124:343-51.
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: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:e0206081.
7. Son J, Shin JY, Kim HD, Jung KH, Park KH, Park SJ. Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. Ophthalmology 2020;127:85-94.
8. Choi JY, Yoo TK, Seo JG, Kwak J, Um TT, Rim TH. Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database. PLoS One 2017;12:e0187336.
9. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. JAMA Ophthalmol 2017;135:1170-6.
10. Wang J, Ju R, Chen Y, Zhang L, Hu J, Wu Y, *et al*. Automated retinopathy of prematurity screening using deep neural networks. EBioMedicine 2018;35:361-8.
11. 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. Phys Med Biol 2017;62:8894-908.
12. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Eprint Arxiv 2014;27:3320-8.
13. Decencière E, Zhang X, Cazuguel G, Laÿ B, Cochener B, Trone C, *et al*. Feedback on a publicly distributed image database: The Messidor database. Image Anal Stereol 2014;33:231-4.
14. Nawi NM, Atomi WH, Rehman MZ. The effect of data pre-processing on optimized training of artificial neural networks. Procedia Technol 2013;11:32-9.
15. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, *et al*. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24:1342-50.
16. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, *et al*. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 2015;115:211-52.
17. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. AAAI Conference on Artificial Intelligence; 2016.
18. He KM, Zhang XY, Ren SQ, Jian S. Deep residual learning for image recognition. IEEE Conf Comp Vis Pattern Recog 2016;90:777-8.
19. Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, *et al*. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med 2019;2:25.
20. Wan SH, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. Comput Electr Eng 2018;72:274-82.
21. Shanthi T, Sabeenian RS, Modified Alexnet architecture for classification of diabetic retinopathy images. Comput Electr Eng 2019;76:56-64.
22. Gegundez-Arias ME, Marin D, Ponte B, Alvarez F, Garrido J, Ortega C, *et al*. A tool for automated diabetic retinopathy pre-screening based on retinal image computer analysis. Comput Biol Med 2017;88:100-9.
23. Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, *et al*. 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;318:2211-23.
24. Li Z, Keel S, Liu C, He Y, Meng W, Scheetz J, *et al*. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. Diabetes Care 2018;41:2509-16.
25. Stevenson CH, Hong SC, Ogbuehi KC. Development of an artificial intelligence system to classify pathology and clinical features on retinal fundus images. Clin Exp Ophthalmol 2019;47:484-9.
26. Santin M, Brama C, Théro H, Ketheeswaran E, El-Karoui I, Bidault F, *et al*. Detecting abnormal thyroid cartilages on CT using deep learning. Diagn Interv Imaging 2019;100:251-7.
27. Heisler M, Ju MJ, Bhalla M, Schuck N, Athwal A, Navajas EV, *et al*. Automated identification of cone photoreceptors in adaptive optics optical coherence tomography images using transfer learning. Biomed Opt Express 2018;9:5353-67.
28. Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, *et al*. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep 2018;8:16685.