



Artificial Intelligence–Assisted Early Detection of Retinitis Pigmentosa — the Most Common Inherited Retinal Degeneration

Ta-Ching Chen^{1,2} · Wee Shin Lim³ · Victoria Y. Wang⁴ · Mei-Lan Ko⁵ · Shu-I Chiu³ · Yu-Shu Huang¹ · Feipei Lai⁶ · Chung-May Yang¹ · Fung-Rong Hu¹ · Jyh-Shing Roger Jang³ · Chang-Hao Yang¹ 

Received: 11 August 2020 / Revised: 2 June 2021 / Accepted: 21 June 2021 / Published online: 9 July 2021
© Society for Imaging Informatics in Medicine 2021

Abstract

The purpose of this study was to detect the presence of retinitis pigmentosa (RP) based on color fundus photographs using a deep learning model. A total of 1670 color fundus photographs from the Taiwan inherited retinal degeneration project and National Taiwan University Hospital were acquired and preprocessed. The fundus photographs were labeled RP or normal and divided into training and validation datasets ($n = 1284$) and a test dataset ($n = 386$). Three transfer learning models based on pre-trained Inception V3, Inception Resnet V2, and Xception deep learning architectures, respectively, were developed to classify the presence of RP on fundus images. The model sensitivity, specificity, and area under the receiver operating characteristic (AUROC) curve were compared. The results from the best transfer learning model were compared with the reading results of two general ophthalmologists, one retinal specialist, and one specialist in retina and inherited retinal degenerations. A total of 935 RP and 324 normal images were used to train the models. The test dataset consisted of 193 RP and 193 normal images. Among the three transfer learning models evaluated, the Xception model had the best performance, achieving an AUROC of 96.74%. Gradient-weighted class activation mapping indicated that the contrast between the periphery and the macula on fundus photographs was an important feature in detecting RP. False-positive results were mostly obtained in cases of high myopia with highly tessellated retina, and false-negative results were mostly obtained in cases of unclear media, such as cataract, that led to a decrease in the contrast between the peripheral retina and the macula. Our model demonstrated the highest accuracy of 96.00%, which was comparable with the average results of 81.50%, of the other four ophthalmologists. Moreover, the accuracy was obtained at the same level of sensitivity (95.71%), as compared to an inherited retinal disease specialist. RP is an important disease, but its early and precise diagnosis is challenging. We developed and evaluated a transfer-learning-based model to detect RP from color fundus photographs. The results of this study validate the utility of deep learning in automating the identification of RP from fundus photographs.

Keywords Retinitis pigmentosa · Image analysis · Deep learning · Fundus photograph · Artificial intelligence

Ta-Ching Chen and Wee Shin Lim have the same contribution

✉ Jyh-Shing Roger Jang
jang@mirlab.org

✉ Chang-Hao Yang
chyangoph@ntu.edu.tw

¹ Department of Ophthalmology, College of Medicine, National Taiwan University, Taipei, Taiwan

² Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

³ Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan

⁴ Case Western Reserve University School of Medicine, Cleveland, OH, USA

⁵ Department of Ophthalmology, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan

⁶ Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

Background

Retinitis pigmentosa (RP) is the most common type of inherited retinal disease (IRD), with a worldwide prevalence of approximately 1 in 3000–5000 [1]. Although RP is considered a rare disorder, it still affects a significant number of patients. According to our survey in the Taiwan inherited retinal degeneration project (TIP), approximately 60% of the IRD patients in Taiwan exhibit the RP phenotype, and similarly, totally more than 1 million people worldwide are affected by RP [1]. RP patients typically suffer from progressive vision loss and eventual blindness, imposing significant psychosocial and socioeconomic burdens on both patients and society. With the development of novel treatment approaches such as gene and cell therapies, early detection of patients with RP has become crucial [2, 3]. Accurately identifying these patients allows for timely treatment and management. Owing to the RP relative rareness compared to other eye conditions such as age-related macular degeneration (AMD) or glaucoma, first-contact health care providers may not be familiar with its symptoms. Incorrect first impressions or misinterpretation of fundusoscopic findings can lead to diagnostic errors, necessitating the development of better screening tools.

In recent years, artificial intelligence (AI) based on deep learning has been widely adopted for image recognition in ophthalmology [4]. Deep learning has been applied to fundus photographs and optical coherence tomography (OCT) images to detect retinal diseases such as AMD, diabetic retinopathy (DR), and macular edema [5–7]. Several studies have also applied deep learning to OCT images in the classification of IRDs [8–10]. However, deep learning has not yet been applied to color fundus photographs in the detection of RP. This is an unmet need as fundus images are easier and less expensive to acquire than OCT images. A deep learning model capable of detecting RP on fundus photographs would facilitate timely diagnosis and aid in clinical management. Furthermore, in conjunction with portable fundus cameras, such a model would also have applications in telemedicine and personalized health care [11].

In this study, we applied deep learning to create a model for automated detection of RP in fundus photographs. RP images used in this study were acquired from the database of our TIP project, and the corresponding normal controls were also obtained from National Taiwan University Hospital. We developed a transfer-learning-based model to classify the presence of RP in these fundus images. To our knowledge, this is the first study on the application of deep learning in the detection of RP from color fundus images. The model achieved high sensitivity and specificity with an AUROC of 96.74%, thus demonstrating the utility of deep learning in identifying RP. We further conducted error analysis and

gradient-weighted class activation mapping to identify regions of input that are important for RP classification, which provided insight for future research.

Methods

Subjects and Extracted Images

In this study, 1153 fundus images of RP were downloaded from the picture archiving and communication system. Acquisition of the RP images was approved by the Ethics Committee of the participating institution, National Taiwan University Hospital (Reference number: 201908089RIND). This study is of the retrospective nature, and all used images are fully anonymized. Normal fundus images were obtained from both the National Taiwan University Hospital and the Hsin-Chu Branch of National Taiwan University Hospital. Approval from the institutional review board of the Hsin-Chu Branch at National Taiwan University Hospital was acquired (Reference number: 108–025-E). Signed informed consents were obtained from all subjects. Four independent ophthalmologists, including two general ophthalmologists, one young retinal specialist, and one retinal and IRD specialist, were recruited to read the images for comparison.

Image Preprocessing

First, all images were de-identified to mask off the patient's personal information. As the images were captured by different machines with different resolutions, each image was cropped automatically to have the same height–width ratio of 1:1.75. Deep learning models were used to learn RP features and the corresponding classification from the entire input image. Unimportant regions, such as regions outside of the round boundary of the main object, were cropped out automatically. These unimportant regions could result from artifacts generated during image acquisition. After removing the irrelevant areas, the images were resized to 300×375 pixels, with pixel values from 0 to 1. Reasonable augmentations, such as image rotation and horizontal and vertical flips, were applied to increase the number of training images. The augmented dataset was used only for training the AI models.

Transfer Learning Classification

In this study, we applied three different models, Inception V3, Inception Resnet V2, and Xception. The inception network is a deep network designed to reach performance comparable to that of other deep networks, such as VGG16, using fewer parameters [12].

Transfer learning was applied to these models with initial model weights obtained from the training on the ImageNet dataset. The last output layer was replaced with two dense layers for classification. The first dense layer has 1024 units, and the second has one (for binary classification), all with the sigmoid activation.

Furthermore, we performed various tasks with different numbers of adaptable convolution layers. In task 1, we fixed all convolution layers and fine-tuned the last two layers only. In task 2, we fine-tuned top 20, 40, 60, 80 convolution layers, etc., together with the last two layers. The loss function is binary cross-entropy. The optimization was performed using the Adam optimizer with a learning rate of 0.0001 and a decay of 0.001. The models were trained for 80 epochs in batches of 30 images per step. Fivefold cross-validation was performed for model evaluation. We monitored the validation accuracy during the model training, and the model with the best validation accuracy was saved and used for prediction on test data. All models were trained on a computer with Ubuntu (16.04.6), Intel(R) i7-7740X CPU, two GeForce GTX 1080 Ti 11 Gb GPU, and 62GiB system memory.

Performance Metrics

A total of 935 RP and 324 normal images were collected and used for training the models. An independent test set of 193 RP images and 193 normal images was used to evaluate the performance of the trained models. True-positive rate (sensitivity)–false-positive rate (1 — specificity) receiver operating characteristic (ROC) curves were plotted. By using a ROC curve, we can easily observe the tradeoffs between sensitivity and specificity. We can use the area under a ROC (AUROC) curve to determine the model performance. Accuracy and F beta score were also used for model evaluation. Accuracy is the ratio of correctly labeled images against the total number of images. F beta score allows weighting when computing the harmonic mean between precision and sensitivity, where $\beta = 1$ considers equal weighting between precision and sensitivity and $\beta = 3$ considers more weighting toward sensitivity. The formula for F beta score is as follows:

$$F \text{ beta score} = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{sensitivity}}{(\beta^2 \cdot \text{precision}) + \text{sensitivity}}$$

Accuracy, F-measure, and confusion matrix were computed based on the optimal cutoff point on the ROC curve with Youden's J statistic [13]:

$$\text{Youden's } J \text{ statistic} = \text{sensitivity} + \text{specificity} - 1$$

Model Visualization and Interpretation

To visualize important features learned by the models during training, we adopted the gradient class activation map (grad-CAM) method. In this method, gradient information flowing from input layers to the last convolution layer of a convolutional neural network (CNN) is used, and coarse heat maps of important regions in the input images are generated [14]. For better visualization of grad-CAM, we retrained the models with one-hot encoding to have a two-class classification problem, with the output layer activation of SoftMax. In fact, there is no performance difference between the original binary classification and this two-class classification with our RP dataset.

Results

AI Program Design and Image Collection Flow

In this study, we aim to design an AI program for diagnosing RP using color fundus images, as shown in Fig. 1. Figure 1 A shows that the AI program loads color fundus images and trains a CNN using the images. The CNN extracts the image features in the intermediate hidden layers to predict the probability of the two classes of RP and normal. If the probability of RP is high, we can suggest the patient to consult an IRD specialist for further examination. This program has the potential to be developed into a decision support system for diagnosing RP in rural areas with limited medical resources, as shown in Fig. 1B.

To train our models, we followed the procedure outlined in Fig. 2. The collected images (1670) consisting of RP and normal images were separated into a training set (960 RP and 324 normal) and a test set (193 RP and 193 normal). For training, we used Xception as the pretrained model and performed fivefold cross-validation for model evaluation. On the training of each fold, the best model was obtained when the validation accuracy was maximum. The final model was the ensemble of the best models obtained from each fold. The final output was obtained by weighted average of these five models. Finally, we also visualized the important regions of the images learned by our model using grad-CAM.

Model Selection and Fine-Tuning

To search for the best model, we trained several CNNs (Xception, Inception V3, and Inception Resnet V2) and applied the best model for further testing. We found that the Xception model had the best performance in our test. The Xception model is the most recent version in the inception

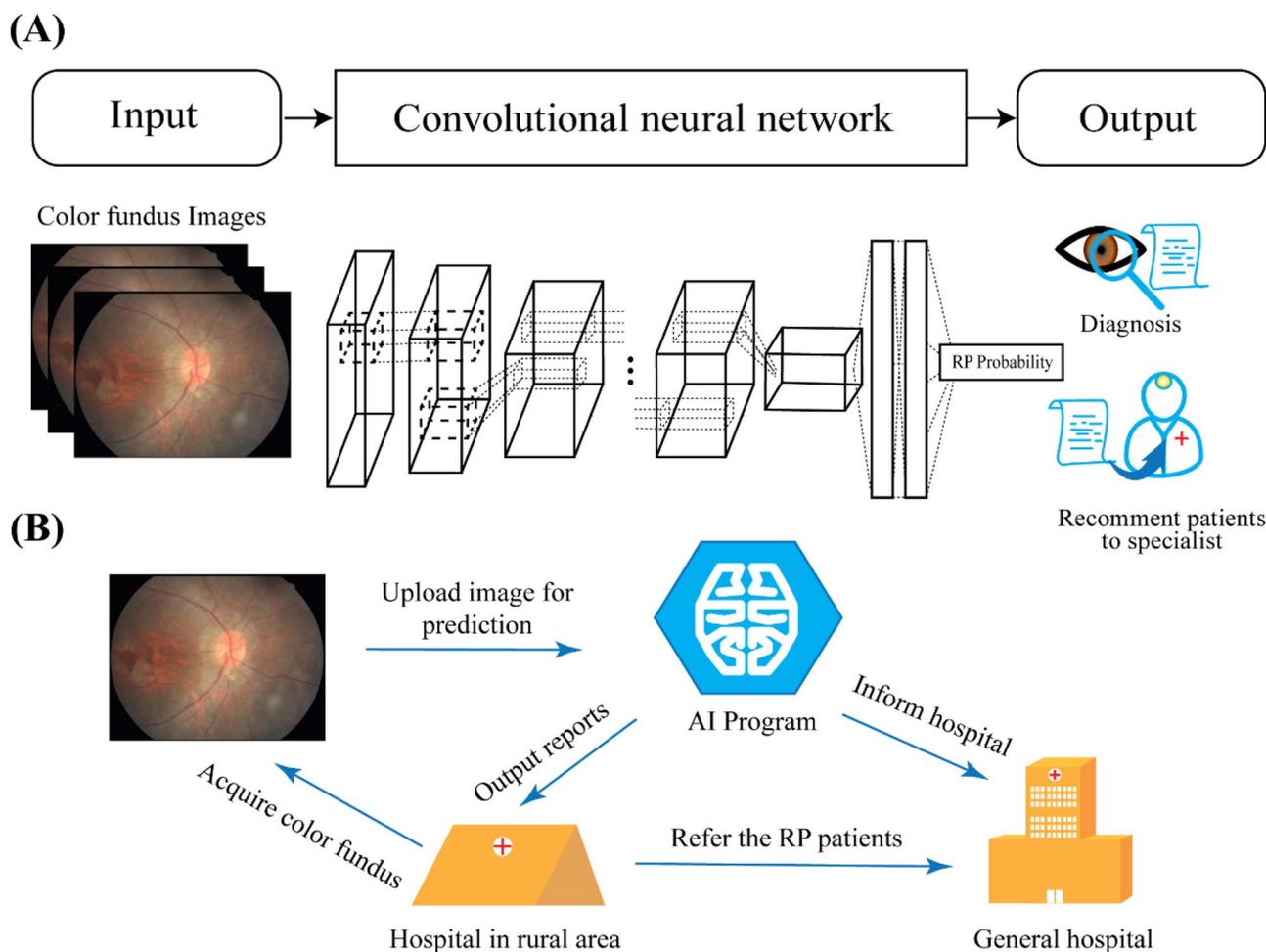


Fig. 1 Program design and usage. **A** Layout of our program design. **B** Outline of our program usage. The proposed artificial intelligence program can be developed into a decision support system, which provides aid in rural areas with limited medical resources

network series. Figure 3A shows the model architecture, where the input first goes through the entry and middle flows, which is repeated eight times, and finally through the exit flow. Convolution, separable convolution, and pooling layers are applied across the whole model in different blocks. The convolution and separable convolution layers in the model act as an image feature extractor. A portion of the feature map on each block is shown in gray scale in Fig. 3B. Grad-CAM was computed using the model weights and the feature map to visualize the importance area in the images. The redder the region, the more significant the AI model considers the area to be correlated with the RP traits, as shown in Fig. 3B.

Among the three CNN models we used, Xception demonstrated the best sensitivity and specificity for early detection of RP with an AUROC of 80%, better than those of Inception V3 and Inception Resnet V2 models, as shown in Fig. 4A, B. First, we trained the model with frozen weights in the convolution layers based on the pretrained CNN

models. The training dataset could reach 100% accuracy, but the validation dataset only reached about 80% accuracy, as shown in Fig. 4C. The discrepancy between the accuracies of the training and validation sets may indicate that overfitting could occur in the constructed models in the fivefold cross-validation. Hence, we changed the number of adaptable CNN layers from 20 layers up to all (126) layers and performed another experiment using the Xception model [15, 16]. After fine-tuning all CNN layers, the AUROC could reach as high as 99% accuracy in both the training and validation datasets, as shown in Fig. 4D.

Test Dataset Performance

Using the Xception model with transfer learning, we can achieve an AUROC of 99.46% for the validation dataset. An independent test dataset was used to evaluate the model objectively. The accuracy, F3, and AUROC for the training, validation, and test datasets are shown in Fig. 5. Figure 5A

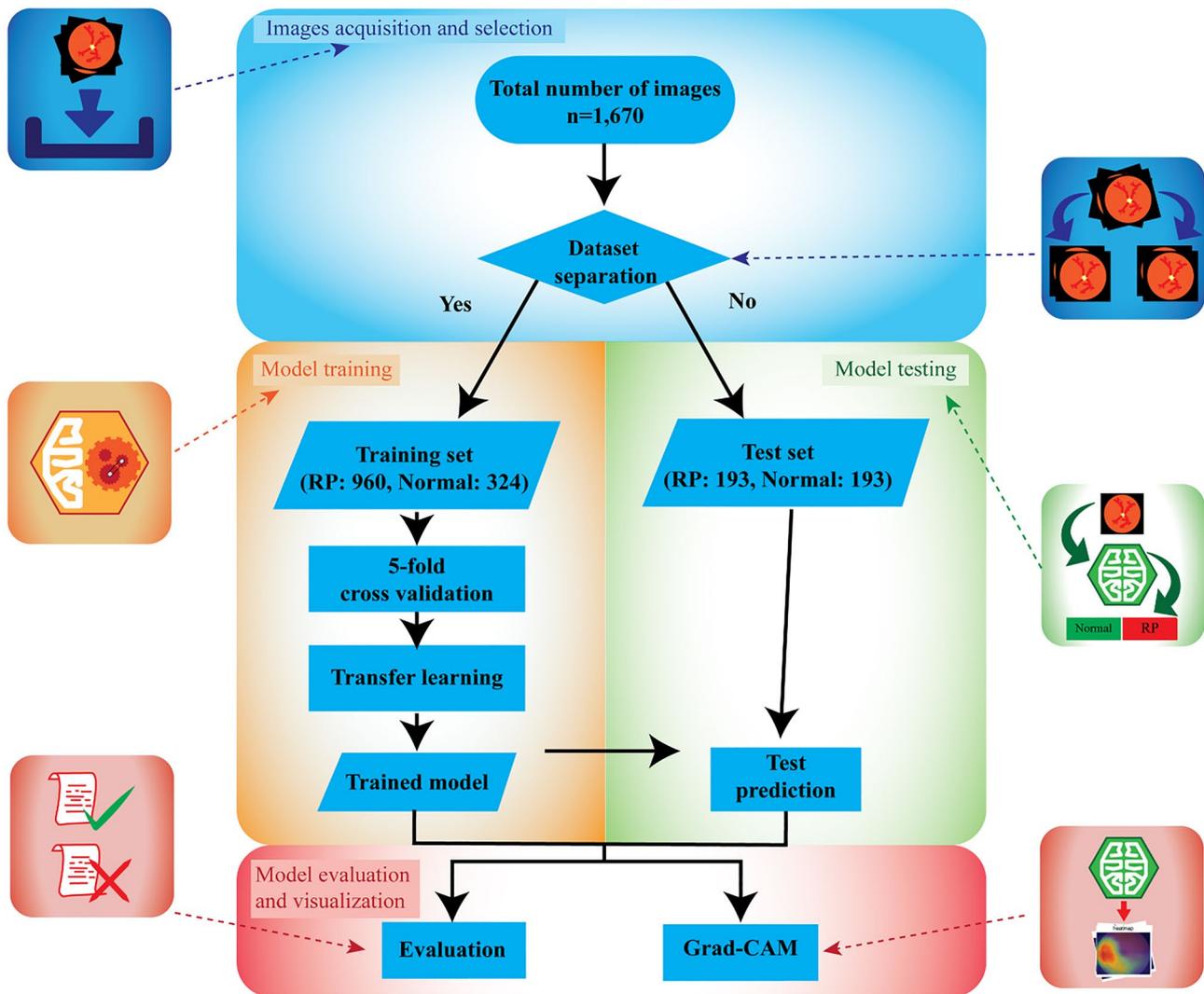


Fig. 2 Program flowchart, which shows the acquisition of color fundus images, training and testing of the model, and model visualization and evaluation

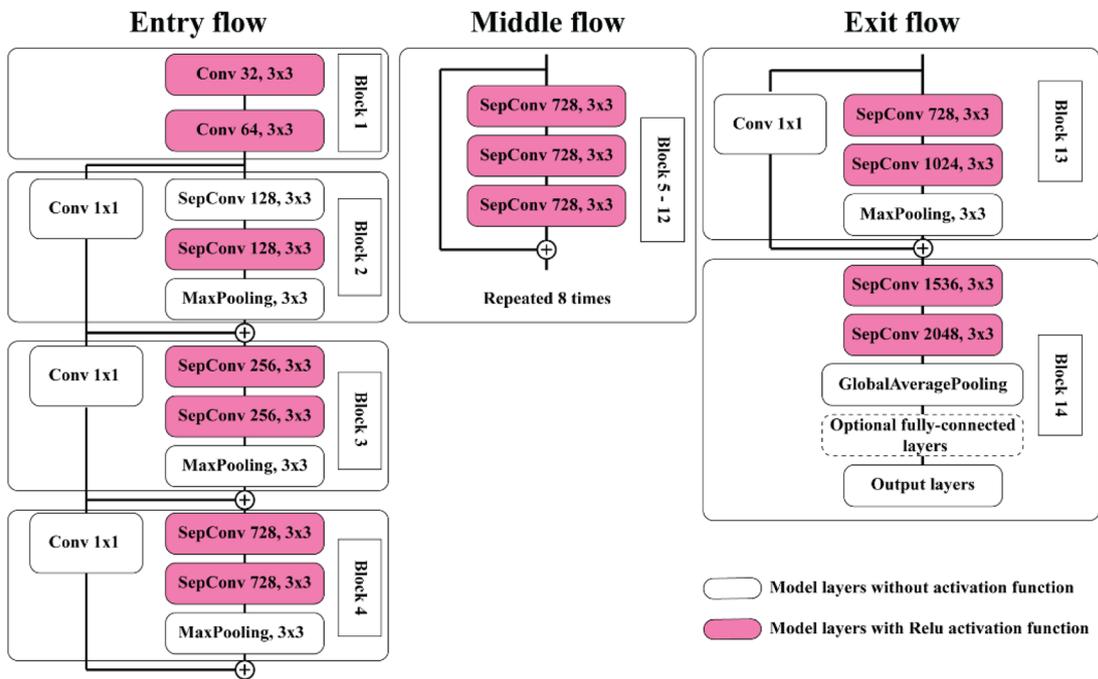
shows that the AUROC on our test dataset is 96.89%, which is satisfactory for common medical applications. From the ROC curve, an optimal cutoff threshold was selected by Youden's J statistic. Based on this threshold, we can measure the accuracy, F-score, and confusion matrices, as shown in Fig. 5B, C. The test dataset can archive an accuracy of 91.45% and F3 of 91.66%, which are only slightly lower than those of the validation dataset, with an accuracy of 96.65% and F3 of 96.88%. As the performance indices between the validation and test datasets are close, we can safely assume that the model is not likely to run into the risk of overfitting herein.

Model Interpretation and Error Analysis

Although the Xception network demonstrated the best performance in classifying RP color fundus images, the model is

similar to a black box as it lacks explanation on which parts of the images are important for the classification. In this study, we used grad-CAM to locate the important areas for the classification [14]. According to the results of heat maps shown in Fig. 6A, we could observe that the "hot area" of the normal group is more extensive, usually covered from the optic disc, vessel arcades, to the periphery. However, the hot area of the RP group focused mainly on the macular area. This could mean that the contrast between the periphery and the macula is an important feature for the identification of RP. The idea seems to be confirmed in the false-positive and false-negative cases shown in Fig. 6B. In our model, false-positive results were mostly obtained in cases of high myopia with highly tessellated retina, especially in the periphery, and false-negative results were mostly obtained in cases of unclear media such as cataract that led to a decrease in the contrast between the peripheral retina and the macula.

(A) Xception Model Architecture



(B) Xception model feature extraction

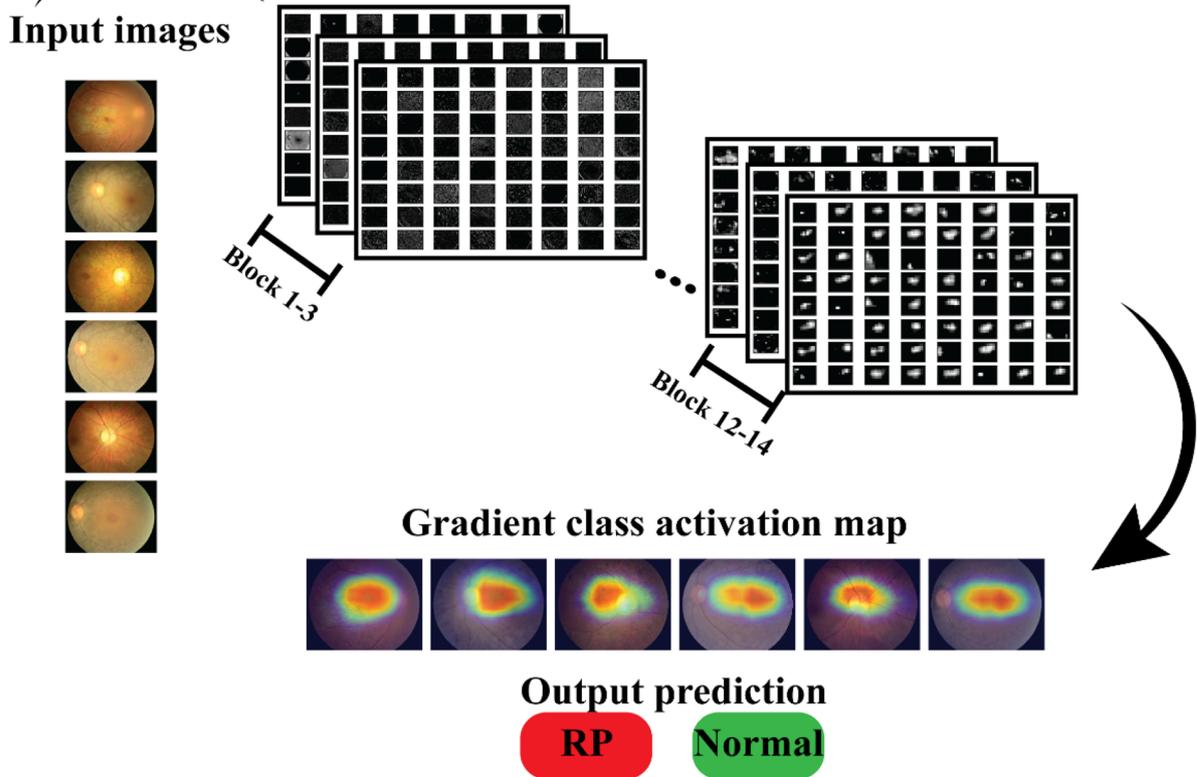


Fig. 3 Xception model architecture visualization. **A** Xception model architecture. **B** Feature map in hidden layers visualized in gray scale and visualization of the gradient CAM heat map

Performance Comparison Between the AI Model and Ophthalmologists

To test our AI model further, we compared the performance between our AI model and that of ophthalmologists in classifying the RP cases. We randomly picked 100 (70 RP and 30 normal) images as a new test set from our datasets. These randomly selected images were graded separately by four domain experts, including two general ophthalmologists,

one young retinal specialist, and one retinal and IRD expert. We retrained our model without all other images (excluding these 100 images) and compared the model performance with the grading results of the four experts. The model could reach an AUROC of 96.14% when predicting these randomly selected images, as shown in Fig. 7A. Moreover, our model had the highest values of accuracy (96.00%), precision (98.53%), sensitivity (95.71%), and F3 (95.99%) when compared with the results of the four experts, as shown in the Table in Fig. 7B. In conclusion, our model produced consistent and stable results in the validation and test sets. It also compares favorably with the domain experts based on another test set of 100 images.

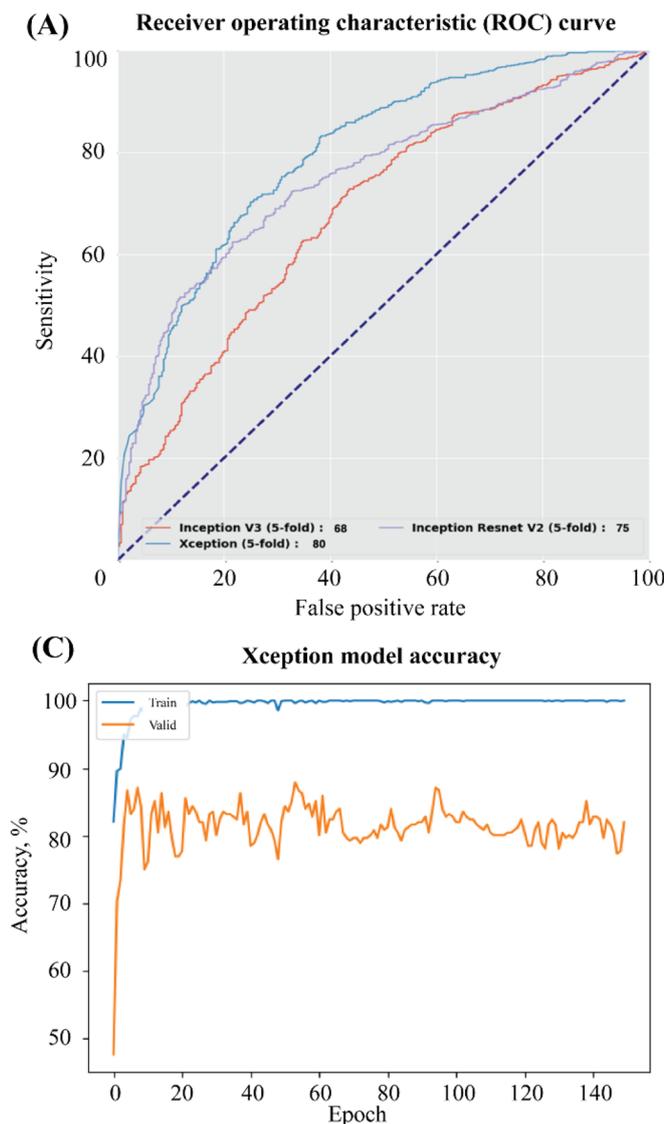


Fig. 4 Model evaluation and hyperparameter tuning. **A** Receiver operating characteristic (ROC) curve of the training dataset on three different convolutional neural network models. **B** Xception model has the highest area under ROC curve. **C** Xception model accuracy. **D**

Change in the area under the receiver operating characteristic when fine-tuning a different number of convolution layers of the Xception model

Discussion

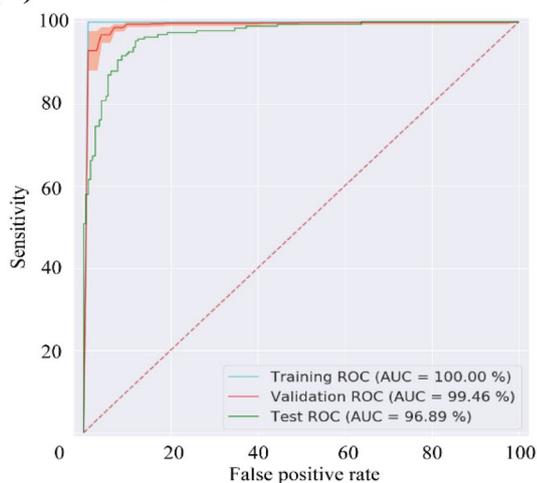
RP is a retinal degenerative disease characterized by progressive degeneration of rod photoreceptors followed by secondary cone loss. In the middle stages of the disease, fundus examination reveals characteristic bone spicule-like pigment deposits. In the later stages of the disease, fundus examination reveals widespread atrophy and degeneration of the retina. Although classic RP fundus photographs have distinct characteristics that are easily identifiable by ophthalmologists, very early or late stages of the disease may resemble other pathologies. Furthermore, RP is genetically heterogeneous, with over 100 genes implicated. The corresponding wide range of presentations complicates diagnosis [1]. This study aims to standardize and automate the diagnosis of RP using deep learning-based image recognition.

Deep learning has been widely applied to ocular imaging, and numerous models have demonstrated robust performance in detecting various retinal diseases, such as DR and AMD, from fundus photographs [5, 7]. However, to

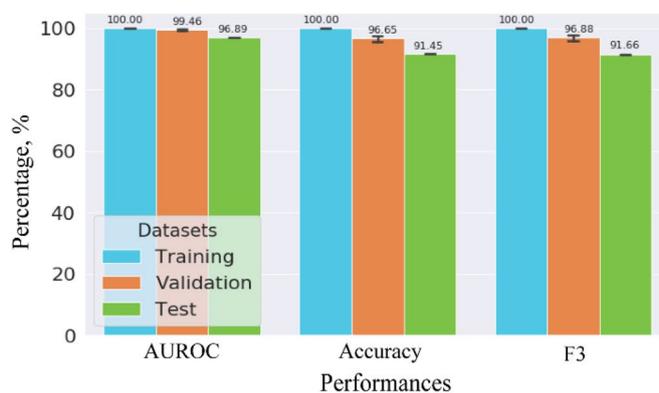
date, the value of deep learning in the detection of RP has not been explored. Our study applies deep learning to the identification of RP in color fundus images, and the results demonstrate that our deep learning-based algorithm can differentiate RP from healthy fundi with high sensitivity and specificity. We conducted error analysis to identify the types of fundus images that were misclassified by our algorithm. False-positive cases occurred predominantly in highly myopic eyes with tessellated fundi, which can be mistaken for early RP. False-negative cases occurred in low-quality fundus images, such as eyes with cataracts or corneal scarring.

The Xception model demonstrated the best performance, as compared to Inception V3 and Inception Resnet V2, in our study. Next, we discuss why the Xception model performed better than the other two models. Both Xception and Inception-ResNet V2 are the modifications of the Inception V3 model. The Inception-ResNet V2 model combines a residue connection and a revised version of the Inception architecture, which increases the model depth while retaining its computational efficiency

(A) Receiver operating characteristic (ROC) curve



(B)



(C)

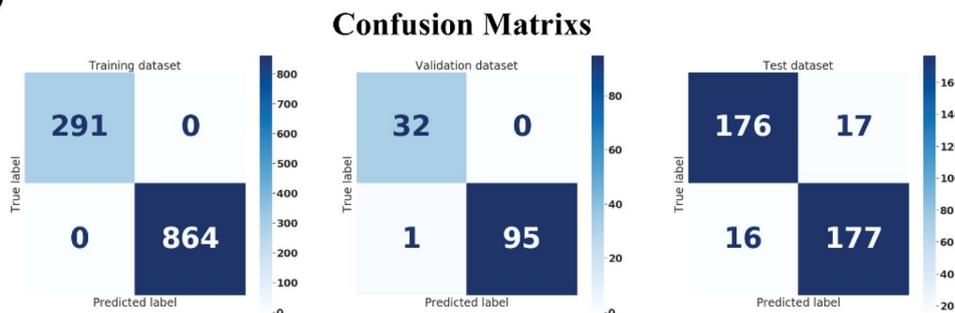
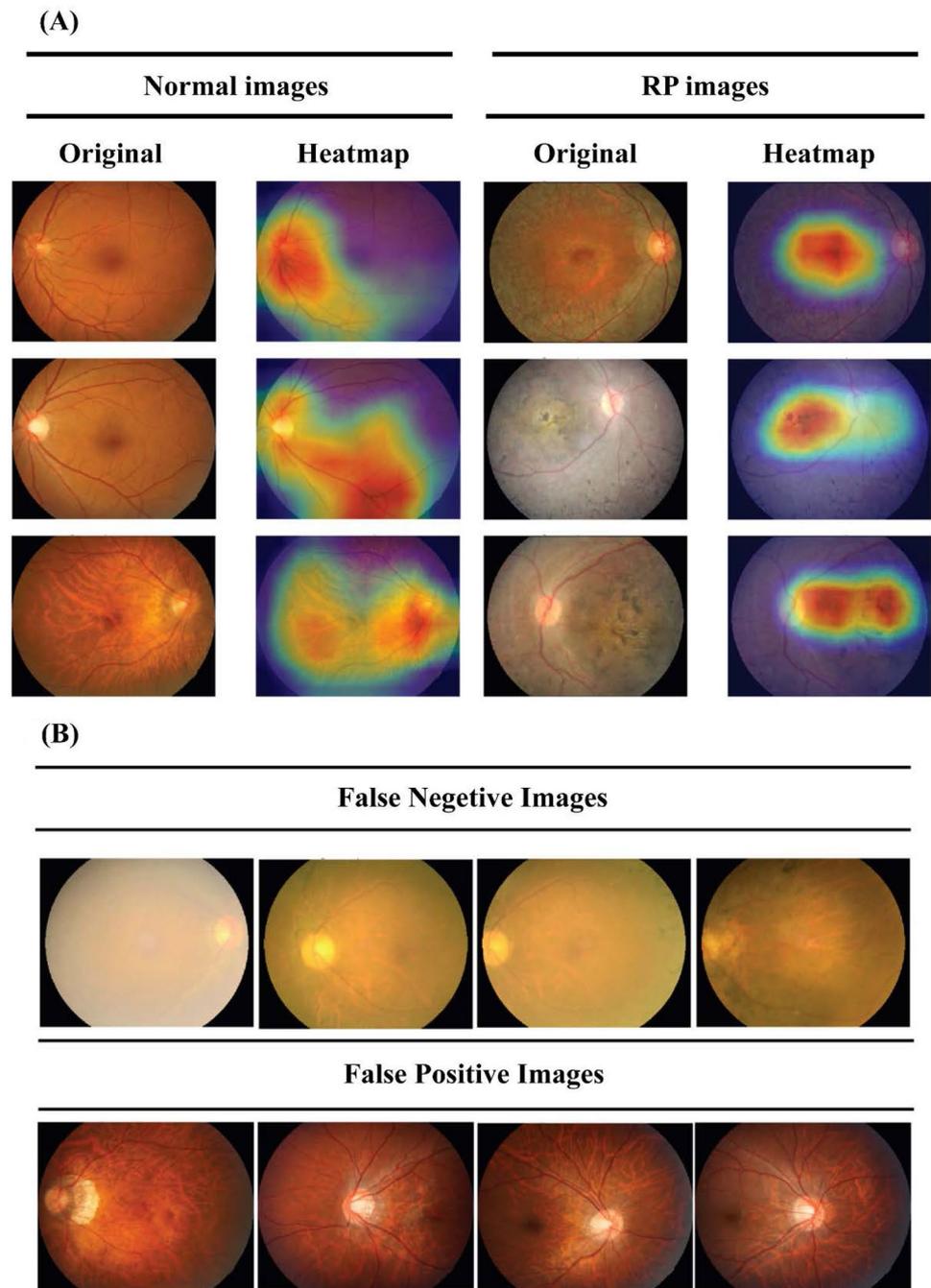


Fig. 5 Overall dataset performance. **A** Area under the receiver operating characteristic (AUROC) results in the training, validation, and test datasets. The test dataset AUROC is slightly lower than that of the validation or training datasets but still is more than 95%. **B** Confusion

matrix of the training, validation, and test datasets. **C** Bar chart compares the AUROC, accuracy, and F3 among the training, validation, and test datasets

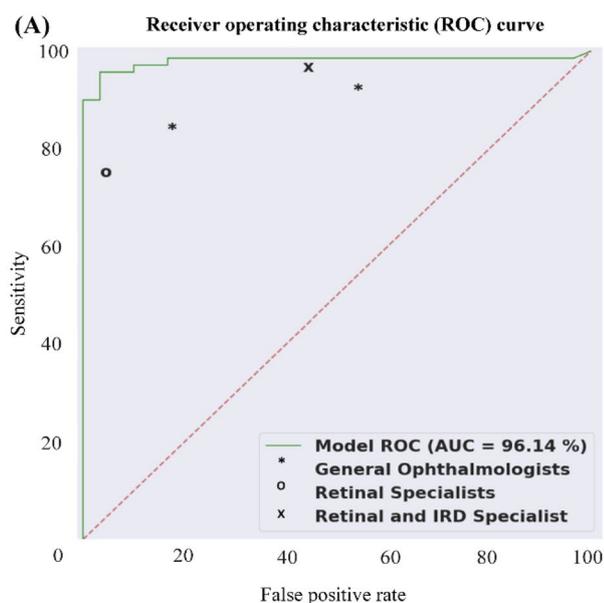
Fig. 6 Model interpretation and error analysis. **A** Class activation heat maps show retinitis pigmentosa (RP) features. **B** High-myopia images are likely to be falsely predicted as RP, and RP patients with cataract have a higher chance to be classified as normal



[17]. Similarly, the Xception model is the latest modification in the Inception model series; it adds a residue connection into the model and improves the Inception architectures by replacing Inception modules with depthwise separable convolutions [18]. The separable convolution used in Xception has a major impact on its improvement as it can potentially decouple the learning of channel-wise and space-wise features. We assume this to be the case because RP color fundus images show neurodegenerative patterns both space-wise and channel-wise. The separate learning of channel-wise and space-wise features might result

in better outcome of our model. Therefore, the Xception model performs better on our dataset.

This study validates the value of applying deep learning to the detection of RP from color fundus images. This should be of importance as RP is a disease that may lead to irreversible blindness, and early detection could help the patients to seek further consultation and potential treatments. Furthermore, RP is an inherited disease that can affect other members in the patient's family. Early awareness may also assist with the family planning. However, our algorithm is

**(B)**

	Accuracy, %	Recall, %	False Positive Rate, %	Precision, %	F3, %
G1	83.00	82.86	16.67	92.06	83.69
G2	78.00	91.43	53.33	80.00	90.14
R	81.00	74.29	3.33	98.11	76.13
R+IRD	84.00	95.71	43.33	83.75	94.37
Model	96.00	95.71	3.33	98.53	95.99

G: General Ophthalmologists
 R: Retinal Specialists
 R+IRD: Retinal and IRD specialist

Fig. 7 Performance comparison of our AI model and ophthalmologists. **A** Receiver operating characteristic curve of the randomly selected 100 images. **B** Table shows all evaluation performance metrics on the AI model and the results of four experts

limited as it only detects the presence of RP and does not differentiate between RP and different causative genes; in addition, it cannot identify the RP stage. This limitation is due to the relatively low number of images, by deep learning standards, used to train this model. Compared to other ocular diseases, RP is rare, and thus data are limited. Hence, a small dataset was used. In the future, we will consider international collaboration to attain enough images to cover different RP subtypes, thus expanding our algorithm scope of performance.

Our study has the following strengths. First, we utilized the data from TIP, a comprehensive data source that has long-term follow-up and observational data for most IRD patients in Taiwan. Most fundus photographs used in this study were captured using the same camera and graded by a fixed team of retina specialists in Department of Ophthalmology, National Taiwan University, Taiwan. The inherent problem of CNN-based classification systems is that algorithms may use peculiarities in image acquisition and grading to make predictions. In this study, using images taken from the same camera to train the algorithm mitigates the risk of the RP classification based on imaging anomalies. However, the homogeneity of our testing data increases the risk of overfitting; hence, the model may perform worse when tested with images from other sources. Further validation is still warranted in the future. Based on our results, cross-institutional collaboration to collect more RP images may potentially expand the algorithm capacity for detection of different RP genotypes.

Conclusion

In this study, a deep learning-based algorithm trained using color fundus images was demonstrated. It achieved high sensitivity and specificity in identifying eyes with RP. To the best of our knowledge, this is the first study to evaluate the utility of deep learning in automating the detection of RP from fundus photographs. Further research is needed to explore the practicality of clinical applications of this algorithm.

Abbreviations AI: Artificial intelligence; AMD: Age-related macular degeneration; AUROC: Area under the receiver operating characteristic; CNN: Convolutional neural network; DR: Diabetic retinopathy; grad-CAM: Gradient class activation map; IRD: Inherited retinal disease; OCT: Optical coherence tomography; ROC: Receiver operating characteristic; RP: Retinitis pigmentosa; TIP: Taiwan inherited retinal degeneration project

Acknowledgements We thank Dr. Chia-Yi Cheng, Dr. Mei-Chi Tsui, and Dr. Hsuan-Chieh Lin for the help with collecting data in this study.

Funding This study is supported by the research grants: NTU Medical Genie — AI Decision Support System for Precision Medicine (Sub-project 4: AI Technologies for Precision Medicine) from National Taiwan University Hospital, Taipei, Taiwan.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Hartong DT, Berson EL, Dryja TP. Retinitis pigmentosa. *Lancet*. 2006; 368(9549): 1795-1809.
2. Prado DA, Acosta-Acero M, Maldonado RS. Gene therapy beyond luxturna: a new horizon of the treatment for inherited retinal disease. *Curr Opin Ophthalmol*. 2020; 31(3): 147-154.
3. Miraldi Utz V, Coussa RG, Antaki F, Traboulsi EI. Gene therapy for RPE65-related retinal disease. *Ophthalmic Genet*. 2018; 39(6): 671-677.
4. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019; 103(2): 167-175.
5. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016; 316(22): 2402-2410.
6. Lee CS, Tying AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017 ;8(7): 3440-3448.
7. 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(11): 1170-1176.
8. Camino A, Wang Z, Wang J, Pennesi ME, Yang P, Huang D, et al. Deep learning for the segmentation of preserved photoreceptors on en face optical coherence tomography in two inherited retinal diseases. *Biomed Opt Express*. 2018; 9(7): 3092-3105.
9. Fujinami-Yokokawa Y, Pontikos N, Yang L, Tsunoda K, Yoshitake K, Iwata T, et al. Prediction of causative genes in inherited retinal disorders from spectral-domain optical coherence tomography utilizing deep learning techniques. *J Ophthalmol*. 2019; 2019: 1691064.
10. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018; 172(5): 1122-1131.
11. Jin K, Lu H, Su Z, Cheng C, Ye J, Qian D. Telemedicine screening of retinal diseases with a handheld portable non-mydratric fundus camera. *BMC Ophthalmol*. 2017; 17(1): 89.
12. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015;1–9.
13. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*. 2005; 73–81.
14. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision. 2017;618–626.
15. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In Advances in Neural Information Processing Systems 2014; 3320–3328.
16. 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. *Scientific reports*. 2018; 8(1): 1-13.
17. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI Conference on Artificial Intelligence 2017.
18. Chollet F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017; 1251–1258.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.