An AI-aided Solution to Open-angle Glaucoma Screening in Developing and Rural Countries

Jiacheng Dang Secondary diploma, Yvonne Chan PHD ‘Iolani School

Address: 563 Kamoku St. Honolulu, Hi, 96826

Tel: 808-683-5982

E-mail: jd2101@iolani.org

Précis

This study developed a deep learning algorithm that detects open-angle glaucoma (AUC=0.994). It also discussed how it can be implemented with previous work to address glaucoma screening in developing countries.

Purpose: The purpose of this study was to develop a feasible deep learning algorithm that diagnose open-angle glaucoma on digital fundus images

Methods: 3870 digital fundus images (DFIs) from RIGA and OGIGA were used to train the object detection deep learning architecture You Only Look Once v3(YoloV3) to detect the area of Optic Disc (OD) from a DFI. 901 healthy and 761 glaucomatous optic discs from ACRIMA, RIM-ONE, Shoi86-HRF, and Drishti-GS were used to train the image classification architectures to detect glaucomatous change on the optic disc. Seven of the most popular deep learning architectures, 5 of the optimizers and 4 different learning rates were tested in various combinations to optimize the classification accuracy between healthy and glaucomatous optic discs. Multiple training strategies were applied to minimize the overfitting of the deep learning model. The area under the receiver operating characteristic curve (AUC) was calculated and compared other deep learning algorithms to evaluate the accuracy of discrimination for each algorithm.

Results: In the test dataset, this deep learning system achieved an accuracy of 97.9%(AUC=0.994), a sensitivity of 98.2% and specificity 97.6%, which was significantly larger than the AUCs of all the other deep learning models with different structures. A website interface was built to offer public access to the diagnostic algorithm.

Conclusion: A deep learning system with a website interface that can detect glaucomatous changes with high sensitivity and specificity.

Key Words: glaucoma, deep learning, optic disc

Introduction

Research Background and Significance

Glaucoma is the second leading cause of blindness in the world. By 2020, 80 million people are predicted to be living with glaucoma (Quigley, 2006). Open-angle glaucoma is the most common form of glaucoma. It is a chronic eye disease that leads to elevated eye pressure, optic atrophy, and visual field defect. The treatment for terminal open-angle glaucoma is limited, and the resulting blindness is irreversible. The visual loss and blindness caused by open-angle glaucoma can affect patients' quality of life in multiple ways. Not only do the patients face difficulties living independently, but they also have roughly double the incidence of anxiety and depression (Skalicky, 2008). The prevention of the progression of open-angle glaucoma is possible; however, given proper treatment at its early stage. The most significant difficulty in its diagnosis is that open-angle glaucoma is asymptomatic at its early stage. Patients are often unaware of its existence until the severe visual loss has occurred. In China, approximately 90 percent of glaucoma patients are unaware of their diseases and thus receive no treatment. If an annual glaucoma test were deployed, this incidence could be cut in half (Ding, 2005). However, the current detection of early glaucoma is problematic. The current glaucoma examination process is time consuming, expensive, and based on a visual

examination (Khouri, 2015). The considerable expense and the cumbersome process of the annual glaucoma exam prevent many people from taking the glaucoma exam. More importantly, glaucoma examination is barely accessible in developing countries. Due to the cost of retina scanning equipment and the required expertise for operating such equipment and making diagnosis, glaucoma examination is unaffordable or not offered in places with limited medical resources. Furthermore, such detection methods of visual examination are subjective and prone to errors. Hence, the purpose of the study is to develop an early-stage glaucoma diagnosis system that is accurate, user-friendly, and low-cost, enabling glaucoma tests to be conducted by people with minimal expertise to prevent the progression of glaucoma for those with limited access to screening.

Diagnosis on Digital Fundus Images

Besides eye pressure measurement and other glaucoma examinations based on numerical measurements of patients, image analysis on Digital Fundus Images (DFIs) of patients are also widely applied in glaucoma diagnosis. DFIs are images of patients' retina taken by widefield retinal imaging devices. DFIs are one of the major sources of reference during glaucoma examination by ophthalmologists. During the visual examination on DFI, various indicators can be used to diagnose glaucoma: Optic Disc dilation, vertical cup to disc ratio (Hancox OD, 1999), the ISNT rule (disc rim of thickness of inferior ≥ superior ≥ nasal ≥ tempura) (Harizman, 2006), and the occurrence of parapapillary atrophy (Jonas, 1992). However, detections based on these indicators are either difficult to be applied to computer-aid diagnosis or be proved to have no causation with glaucoma (Bailey, 2017) (James, 2017). One of the other detection methods is based on the Optic Disc (OD) assessment through ophthalmoscopy. Compared to other indicators, assessment on the Optic Disc is more applicable to deep learning. OD is the retina area where retinal ganglion cells form the optic nerve while exiting the eye, which transmits visual information from the retina to the brain. Optic atrophy, as one indicator of glaucoma, is the loss of some or most nerve fibers in the optic nerve, which changes the structure of the OD. Since a large number of healthy and glaucoma DFIs are accessible on public databases, analyzing DFIs is especially useful for training the glaucoma diagnosing program-- an system that determines whether any potential glaucoma features are present in a DFI. In this paper, open-angle glaucoma is diagnosed based on the classification of Optic Discs.

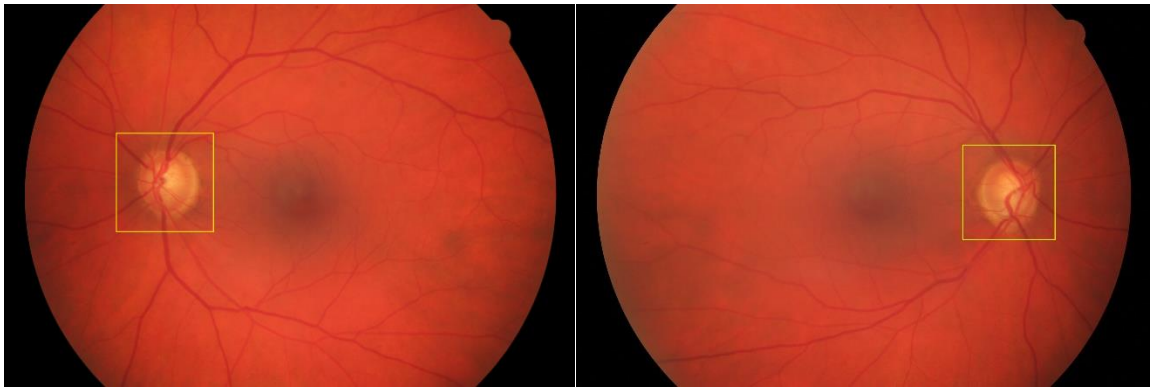


Fig 1 Two digital fundus images with optic discs indicated in the yellow boxes.

Deep Learning

Deep learning is a class of machine learning algorithms that uses multiple layers consisting of multiple linear and non-linear transformation data to extract higher-level and abstract features progressively from the raw input (Bengio, 2013). Convolutional neural networks (CNNs) are a class of deep learning commonly applied to analyze and classify images by assigning importance (or weights and biases) to various features in the image. Eventually, CNNs can differentiate between these features automatically. Deep learning architectures contain multilayer neural networks, including different configurations of networks and training strategies that aim to optimize the outcome. With enough training, CNNs will be able to acquire a set of best-fit weights for each feature to reach optimal classification results via updating the weights of each feature based on the loss function (Ren, 2017). A CNN is composed of four significant parts: convolutional layers, activation functions such as Rectified Linear Unit (ReLU), pooling layers, and fully connected layers. This research proposed an innovative diagnostic system which combines object detection neural networks (Yolov3) and image classification neural networks (Vgg16) to optimize the diagnostic performance. This structure allows the system to bypass manual preprocessing of the datasets, which consequently eliminates the errors associated with such operations. Consequently, this structure allows the system to be trained on relatively larger datasets to achieve a more stable and reliable accuracy. Furthermore, to acquire the optimal performance, seven most outperforming deep learning architectures and five optimizers were compared with tuned hyperparameters. By utilizing seven pre-trained deep learning architectures, this research also tested the validity of pre-trained ImageNet neural network models by fine-tuning the glaucoma datasets. All of the pre-trained models significantly improve the performance of the model and primarily lowers the training time even being applied to a distinctive dataset. To offer public access to glaucoma testing, a website interface (www.glaucomark .com) was developed using the model introduced in this paper with an accuracy of 97.9% (AUC= 0.994). After uploading the optic disc image, the website interface will return the diagnostic result of glaucoma numerically under 30 seconds of processing time.

Methods

Convolutional Neural Networks Layers

The convolutional layers are the major structure of CNN, and they are accountable for the feature extraction. The convolutional layer utilizes a set of learnable filters with different parameters to perform operations on the original input image. Each of the filters will eventually produce a 2-dimensional activation map which will be stacked to form an output volume for further operation. Using a CNN allows the system to capture relevant features from an image at a relatively low computational cost because of weight sharing (CS231n, 2019). In this study, Vgg16 uses sixteen convolutional layers to emphasize the depth of the feature mapping. Since optic discs with and without glaucoma look relatively identical, a larger number of parameters from the deeper neural networks allows the deep learning model to capture the subtle differences during Optic Disc classification tasks. Activation functions numerically achieve the weighted sum of the input values and decide whether it should be kept. ReLU (Rectified Linear

Unit) is one type of activation function that computes the function f(x)=max (0, x). ReLU is less computationally expensive than other activation functions since it involves fewer calculations. This leaves the data unchanged and makes it less susceptible to Vanishing Gradients. The simplicity of this function also solves the saturation problem as long as the input exceeds zero. ReLU activation functions were used after each convolutional layer in VGG16. SoftMax is another type of activation function that turns the output of a fully connected layer into probability that sums to one. It outputs a vector that indicates the potential output classes together with the probabilities associated with each (Sergey, 2015). In the proposed model, the SoftMax function was placed at the end of the deep learning model and output a numerical value from 0 to 1 corresponding to the likelihood of glaucoma on the input digital fundus image. Pooling layers in a CNN architecture are used for abstracting image features (CS231n, 2019). The common pooling layer (max-pooling and mean-pooling) employs a 2x2 size filter with a stride of 2, selecting 25% of the activations on an activation map from convolutional layers. Every max-pooling operation would select the most significant number from a 2x2 region. Pooling layer largely lowers the capacity of parameters, which reduces the computation in the network and controls overfitting, improving the overall performance of the network. A fully connected layer works as a means of mapping features with full connections to all activations from the previous layers (CS231n, 2019). Those activations will be calculated based on a weighted multiplication and biased addition in the next hidden layer and eventually get to the output layer. The usage of fully connected layers helps to select representative features for further classification.

Binary Cross-Entropy Loss Function

In deep learning, the loss function is a method to evaluate how well the model output fits the actual results. The loss function will return a vast number of predictions based on the trained model deviating too much from actual results. In this paper, Binary Cross-Entropy Loss was used to evaluate the loss of this binary prediction model. It consists of a Sigmoid activation and a Cross-Entropy Loss.

$$CE = -\sum_{i=1}^{C'=2} t_i log(f(s_i)) = -t_1 log(f(s_1)) - (1 - t_1) log(1 - f(s_1))$$

Each of the C classes would process through a Sigmoid Activation function first and calculate a Loss based on the Cross-Entropy Loss function. The loss is summed up to indicate the loss of each epoch.

Deep Learning Model

A deep learning model consists of deep learning architecture and model parameters. Deep learning architectures determine the arrangement of different functional layers that apply weighted calculations to the input data. The model parameters determine the weight of each calculation. While training the deep learning model, the model adjusts its parameters by comparing its prediction on training datasets with the actual results until the model prediction is accurate on the testing dataset (not included in training). To optimize the deep learning model, a series of techniques were used. And the validation of the model, as well as the model application, was also completed in this research.

Two-part Diagnostic System

In order to diagnose glaucoma from digital fundus images, two deep learning algorithms were used for two separated tasks. The first part of the system adopts the object detection network YoloV3 to detect the area of optic disc from the retina. Since the change of optic disc structure has been a major symptom of glaucoma, excluding the analysis on areas outside the optic disc will greatly improve the performance of the diagnostic system. An image classification network VGG16 was used in the second part, allowing the system to detect glaucoma features from the optic disc area.

One unique difference of the proposed diagnostic system in this paper is that it utilizes a two-parts deep learning architecture, aiming to avoid the complex and error-prone manual pre-processing of the input digital fundus images. Data augmentation was applied to enhance the capacity of datasets during training. The system proposed in this paper is constructed using TensorFlow Library in Python.

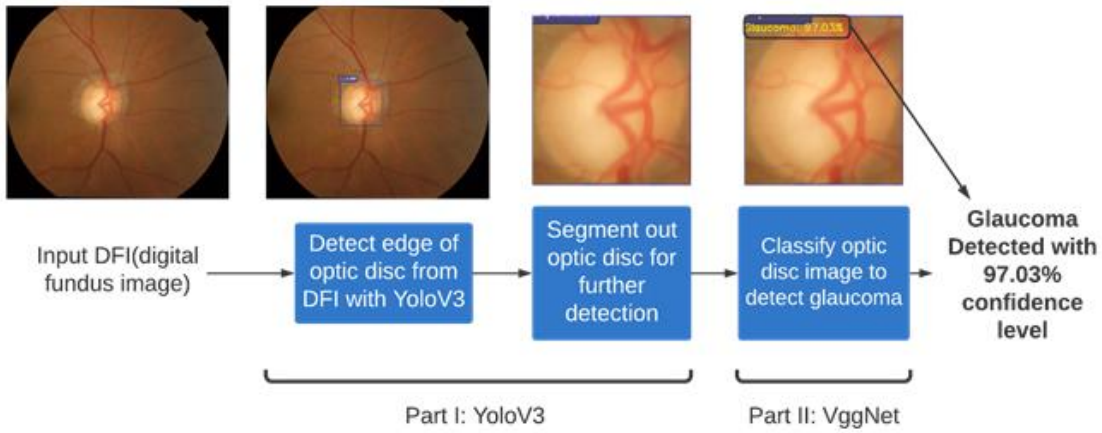


Fig 2. A flowchart displaying the structure of the detection system

To offer public and compatible open-angle glaucoma classification access, a website application was designed based on HTML, CSS, JavaScript, and Python with frameworks and packages such as React, TensorFlow, and Flask. The website has been deployed on GitHub pages with its Flask API being served on Alibaba Cloud Server. The website is now active at **www.glaucomark.com**. After the user uploads the photo of an optic nerve, the website will send the data to the cloud server and fetch back the diagnosing result in under 45 seconds. No data will be saved.

You Only Look Once V3 (YoloV3)

YoloV3 is one of the state-of-art object detection networks. YoloV3 has extremely fast processing speed and a relatively smaller size compared to other object detection architectures. These features enable the deep learning model to be easily embedded into a website interface for public access, making fast diagnosis possible. In one evaluation, It predicts bounding boxes locations and class probabilities by one single neural network directly from full images. Based on the mAP measured on the COCO dataset, YoLoV3(mAP=51.5, time=22) achieved similar performance to Single Shot MultiBox Detector (mAP=50.4, time=125), but six times faster (Redmon, 2018). The loss function was the sum of the loss of four features of detection ((x, y), (w, h), class, and confidence).

The Yolo loss function starts with the loss of localization, where (x, y) represents the of the bounding box coordinates, and (w, h) represents the width and height of the bounding box when there is an object:

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Then added with the loss of two prediction confidence including the bounding box with an object:

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

And the prediction confidence of the bounding box without an object:

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

Finally added with loss of the class probability when there is an object:

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2$$

Where $\mathbb{1}_i^{\text{obj}}$ represents whether the detected object appears in the cell I and $\mathbb{1}_{ij}^{\text{obj}}$ represent if the j th bounding box in the i th cell is responsible for the object (Redmon, 2018).

Visual Geometry Group Neural Network (VggNet)

VggNet is a CNN architecture emphasizing the aspect of depth to its performance by utilizing relatively more convolutional layers with significantly more parameters. Vgg16 was used in the diagnostic system due to its large numbers of parameters (10^7). Since the visual difference between healthy and glaucoma digital fundus images is insignificant, the depth of Vgg16 allows the model to capture the more subtle differences between a healthy retina and a retina with glaucoma. Using tiny (3x3) convolution filters in all layers, it steadily increases the depth of the network by constructing 11-19 convolutional layers, which significantly increases the accuracy of the network by allowing the networks to extract more representational and advanced features. Using multiple tiny (3x3) convolutional layers instead of one larger convolutional layer(7x7) makes the decision function more discriminative and reduces the number of parameters by 55% with the same Receptive Field (CS231n, 2019). In this paper, Vgg16 was compared with six other object classification CNN architectures and outperformed

the other five. However, its limitation of a large number of parameters at the last three fully connected layers leads to severe overfitting due to the high similarity of the data. Such overfitting was solved using multiple techniques.

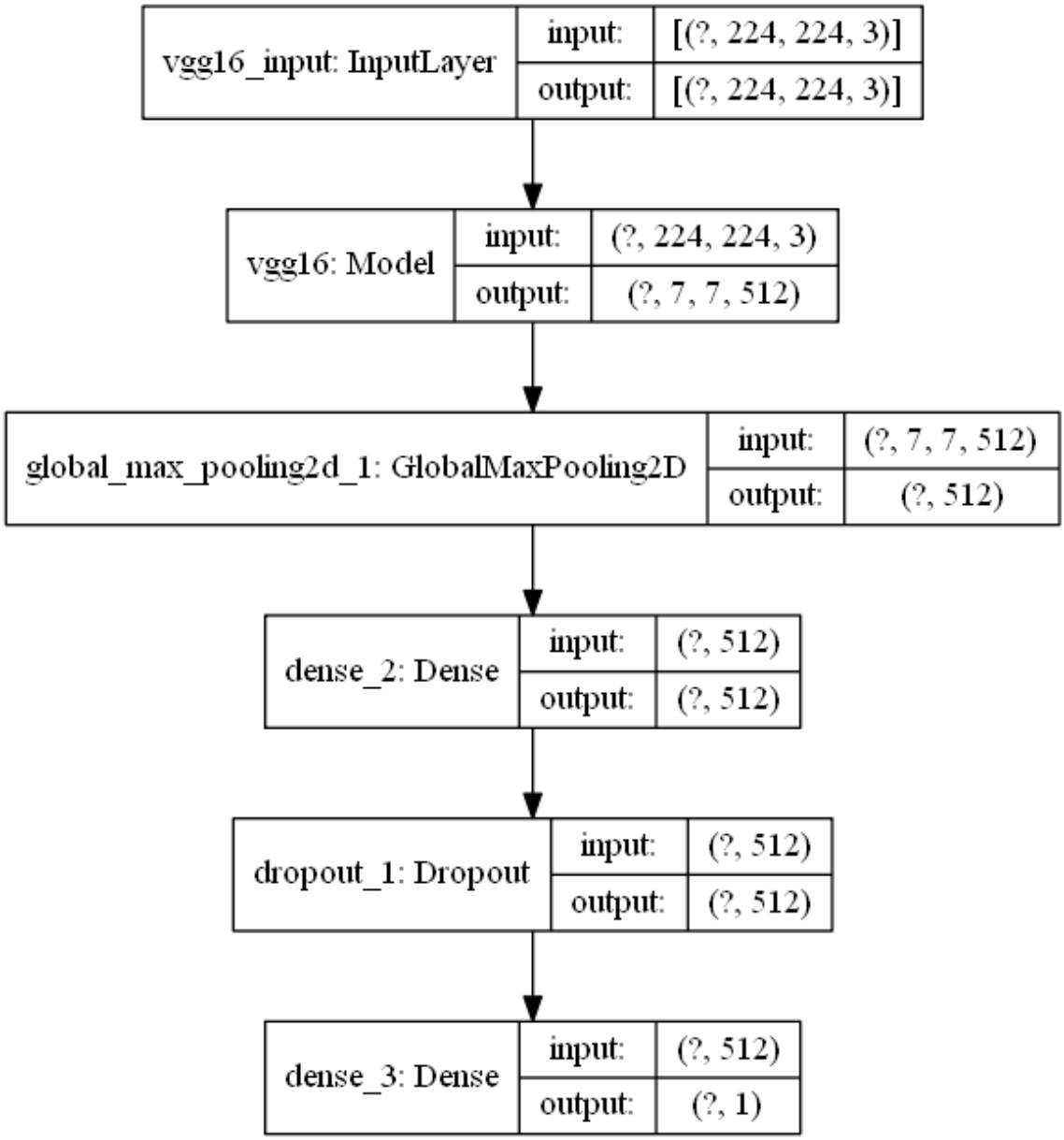


Fig 3 Model structure of the modified Vgg16 used in this paper.

Data Augmentation.

Data augmentation was utilized in this system to enlarge the datasets. A larger dataset increases the diagnostic accuracy of the model in return for extra training time. During this process, original images were transformed into new images with different sizes, viewpoints, brightness, orientations, and translations. Due to the high similarity and symmetry of the fundus image datasets, such operation is exceptionally effective, augmenting the dataset by at least five times.

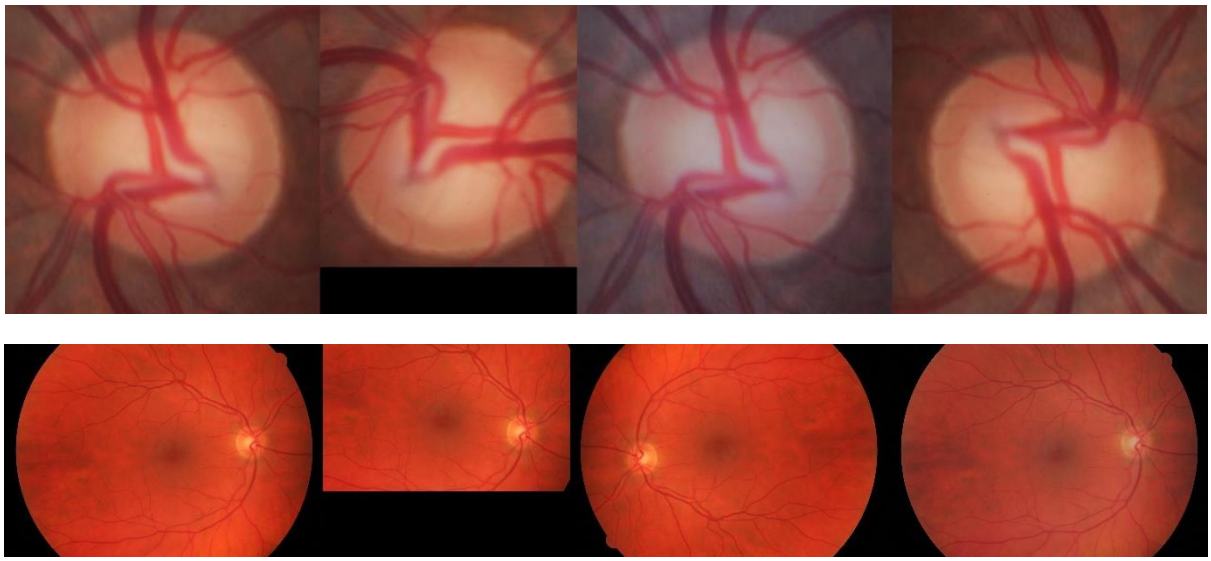


Fig 4,5 Augmented Optic Disc image for image classifier and object detector training

Sources of DFI Datasets

DFIs for testing and training were obtained from RIGA (Almazroa, 2005) , OGIGA(NCBL, 2015), ACRIMA(Diaz-Pinto, 2019), RIM-ONE(RIM-ONE, 2015), Shoi86-HRF (cvblab, 2016), and Drishti-GS (Drishti-GS, 2016). The RIGA and the OGIGA databases contain 3870 DFIs with optic discs labeled by ophthalmologists and were used to train the YoloV3. The ACRIMA, RIM-ONE, Shoi86-HRF, and Drishti-GS databases consist of 901 healthy and 761 glaucoma optic discs for Vgg16 training. The composition of all the data is summarized in Table 1.

Table 1 List of public glaucoma databases and number of images acquired.

	No. of images used to identify optic disc	**No. of images used to diagnose glaucoma**		
Database	**OD-Labeled**	**Normal**	**Glaucoma**	**Total**
Drishti-GS		31	70	101
SJ86-HRF		300	101	401
RIM-ONE		261	194	455
ACRIMA		309	396	705
OGIGA	650			*650*
SRIGA	3220			*3220*
	3870	901	761	**5532**

The Glaucoma/Normal Datasets (Column 2, 3) consist of optic disc images that is specified either glaucoma or healthy. The OD-Labeled Datasets (Column 4) consist of digital retina images labeled with the location of optic discs.

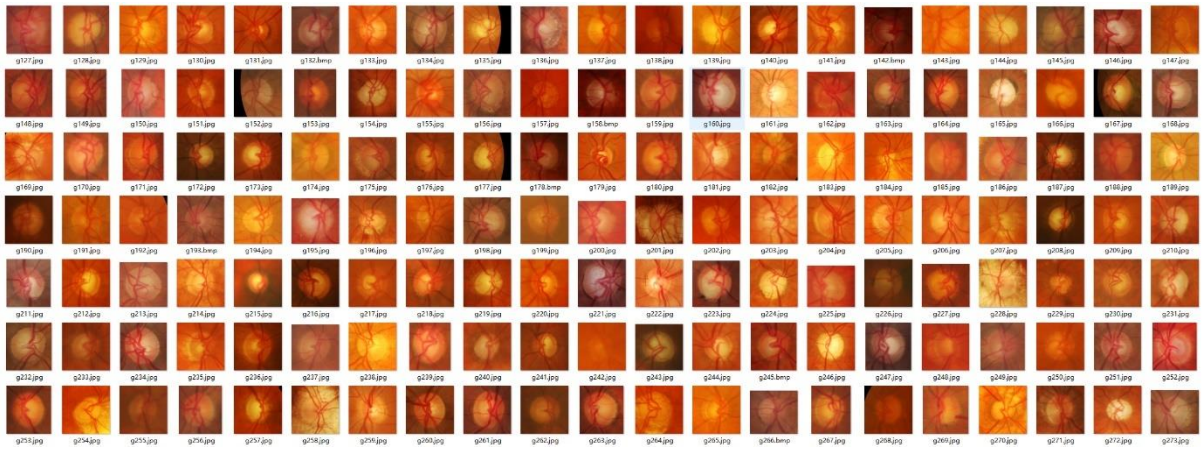


Fig 6 Part of object classfication networks training data

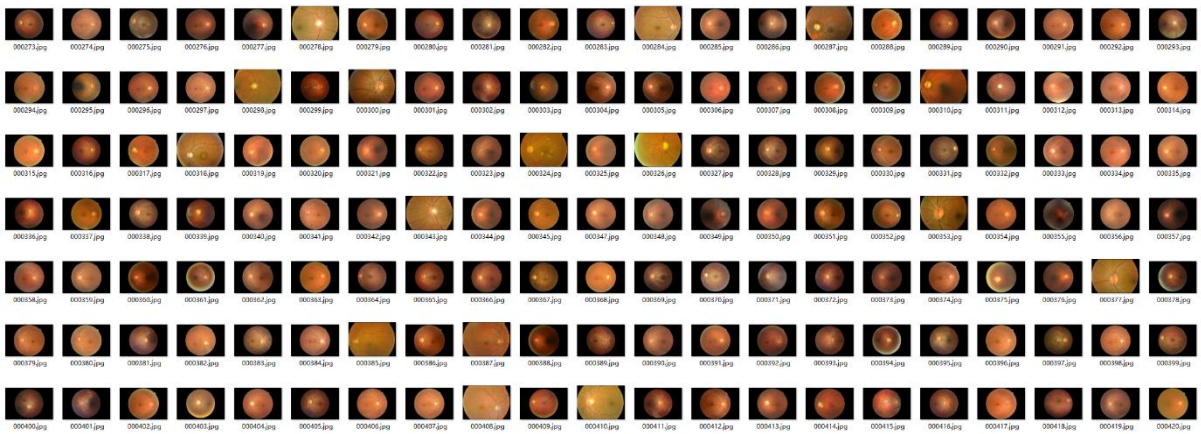


Fig 7 Part of objection detection networks training data

Training Strategy to Avoid Overfit

Overfit is one of the most common problems during model training. An overfitting deep learning model lacks the ability to give sensible outputs to sets of input that it has never seen before. In other words, overfitting occurs when a model has learned noise and/or unrepresentative features from the training datasets rather than representative features, in such a way that it may fail to fit new datasets. In this paper, multiple techniques had been deployed to reduce overfitting.

Early Stopping is a training technique such that the model will be monitored by a function that stops training when the accuracy of the model does not increase for 40 continuous epochs (for each epoch, the model runs through all the training data). This process can prevent overfitting by stopping the model's iteration before its iterative convergence to the training data.

Data Augmentation improves the diagnostic ability of the model by increasing the sample size. The model's ability to generalize is mostly dependent on the size and versatility

of the datasets. By adopting data augmentation to the training datasets, the model will have a reduced chance of overfitting.

Regularization was used to reduce the complexity of the model. The model's complexity increases with more parameters. A complex model overfits more commonly. Regularization reduces the chance of overfitting by adding new information, which simplifies the model.

Deep learning architectures were selected based on their size. The model's learning capacity and complexity determine whether the model generalizes well. In this paper, most architectures were selected based on their number of parameters (fewer than 3e7) to reduce overfitting.

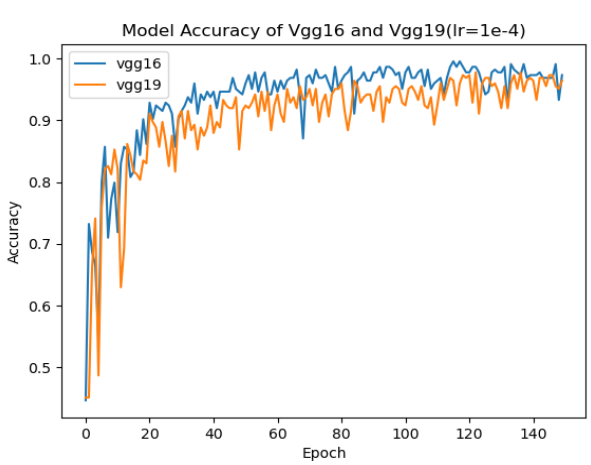
Results

Model Comparison

In this paper, the Vgg16, Vgg19, ResNet50V2, DenseNet121, Xception, InceptionResnetV2, and Inception architectures were applied the same glaucoma diagnosis task using their pre-trained ImageNet weights available in the Keras core. They were trained on 1.28 million images from the ImageNet database and acquired outstanding accuracy. These pre-trained models were beneficial for the system to extract features from Digital Fundus Images (DFIs). These architectures were modified to fit the datasets such that their last fully connected layers were changed to a global average pooling layer (GlobalAveragePooling2D) followed by a Dense layer with ReLU activation, a dropout layer of rate 0.5, a two-node fully connected layer representing two classes (glaucoma and healthy) and a sigmoid classifier. These modified pre-trained models were trained on 1329 training and 332 testing DFIs with learning rates (lr) of 1e-4 and 1e-5. They were trained for 150 epochs, each on a device using GPU RTX-2080Ti 11GB in Ubuntu-16.11. Table 2 presents the validation accuracy of each model on both learning rates. Figures 8-13 show how the models' accuracies changed during training.

Table 2 Performance of Architectures on Learning Rates 1e-4 and 1e-5.

Architecture	Depth	Parameters	Optimizer	Accuracy(lr=1e-4)	Accuracy(lr=1e-5)
VGG16	23	1.38e8	Adam	0.963	0.979
VGG19	26	1.43e8	Adam	0.967	0.974
ResNet50V2	50	2.6e7	Adam	0.956	0.906
DenseNet121	121	8e6	Adam	0.909	0.922
Xception	126	2.3e7	Adam	0.944	0.963
InceptionResNet	572	5.6e7	Adam	0.932	0.935
InceptionV3	159	2.4e7	Adam	0.947	0.919



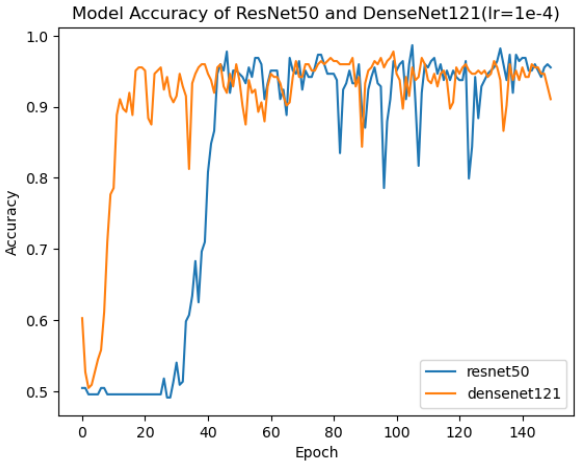
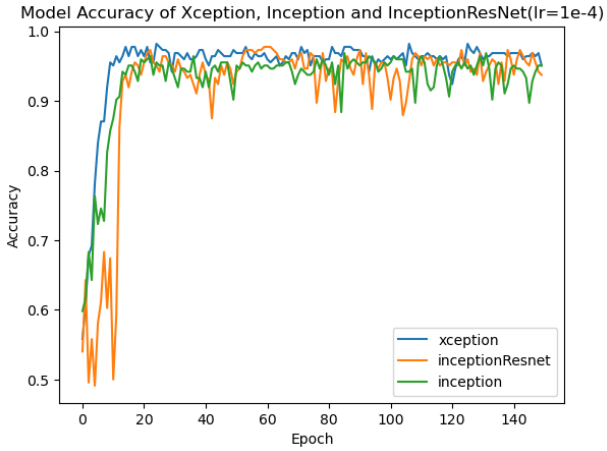


Fig 8. Model Accuracy of Vgg16 and Vgg19 (lr=1e-4). **Fig 9.** Model Accuracy of ResNet50 and DenseNet 121 (lr=1e-4)



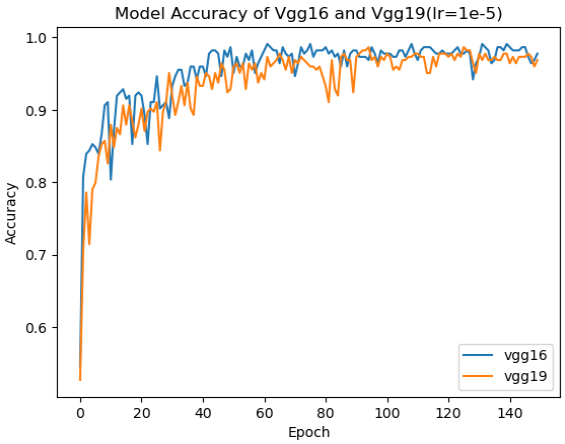
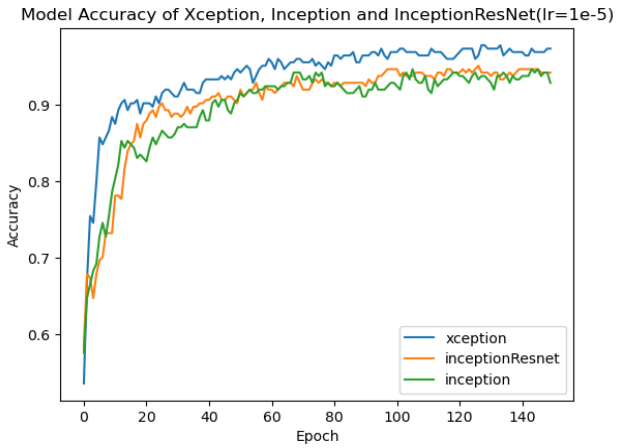


Fig 10. Accuracy of Xception, Inception, and InceptionResNet (lr=1e-4) **Fig 11.** Accuracy of Vgg16 and Vgg19 (lr=1e-5)



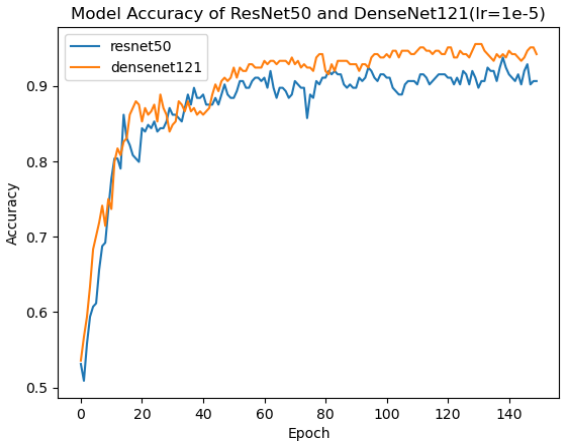


Fig 12. Accuracy of Xception, Inception, and InceptionResNet (lr=1e-5) **Fig 13.** Accuracy of ResNet50 and DenseNet 121(lr=1e-5)

Optimizer Comparison

The error in predictions made by the model after each iteration is measured using a loss function. During the training process, parameters of the model are tweaked to minimize the loss function for optimal performance. Optimizers play the role of improving the model

according to the loss function. In this paper, the five most popular optimizers were deployed on the same VGG16 model on 1329 training and 332 testing DFIs with a learning rate of 1e-5. Results showed that Adam optimizer outperformed other optimizers. Details are shown in Table 3. The Adam optimizer with binary cross-entropy loss function and a learning rate of 1e-5 acquires the accuracy of 0.979. Figure 18 presents the model's performance with different optimizers over 150 epochs.

Table 3 Performance of Optimizers on Vgg16

Optimizer	Loss Function	Momentum	Learning Rate	Accuracy
Adadelta	binary cross-entropy	N/A	1e-5	0.759
SGD	binary cross-entropy	0.9	1e-5	0.817
RMSprop	binary cross-entropy	0.9	1e-5	0.929
Adam	binary cross-entropy	N/A	1e-5	0.979
Adagrad	binary cross-entropy	N/A	1e-5	0.895

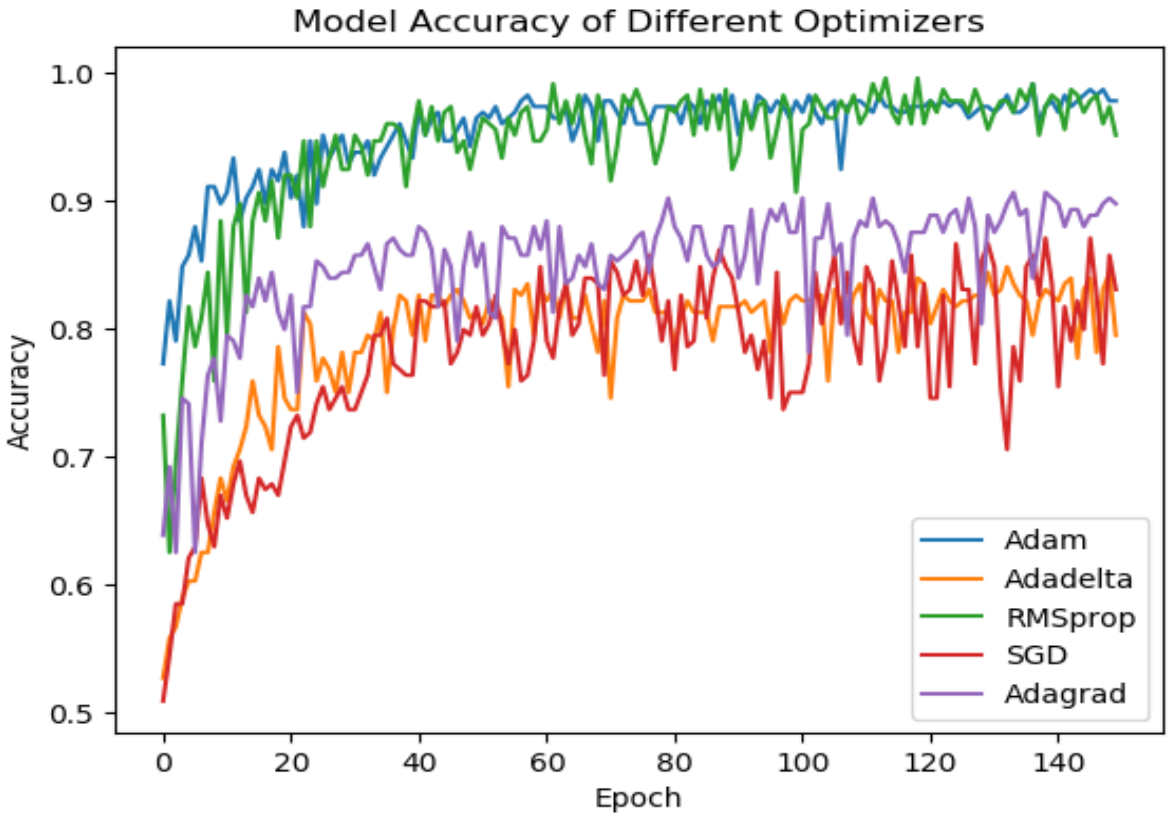


Fig 14. Validation accuracy of the model, using different optimizers.

Hyperparameter Tuning

For optimal performance, multiple learning rates were tested. The learning rate is a crucial hyperparameter in CNN model training that controls the amount of change to improve the model after each update according to the estimated error. A huge learning rate will lead to a quick convergence to a suboptimal solution, whereas a minimal learning rate can significantly slow down the training process. The system is trained with a learning rate of 1e-3,1e-4,1e-5 and 1e-6. The highest performance is obtained with a learning rate of 1e-5. The system achieved an accuracy of 97.9%, as shown in Table 2. To generalize the experiments and result, the training was repeated 150 times (150 epochs) until the system's accuracy was stable.

Table 4 Performance of the proposed system

Learning rate	Accuracy (%)
1e-3	70.1
1e-4	76.7
1e-5	97.9
1e-6	97.1

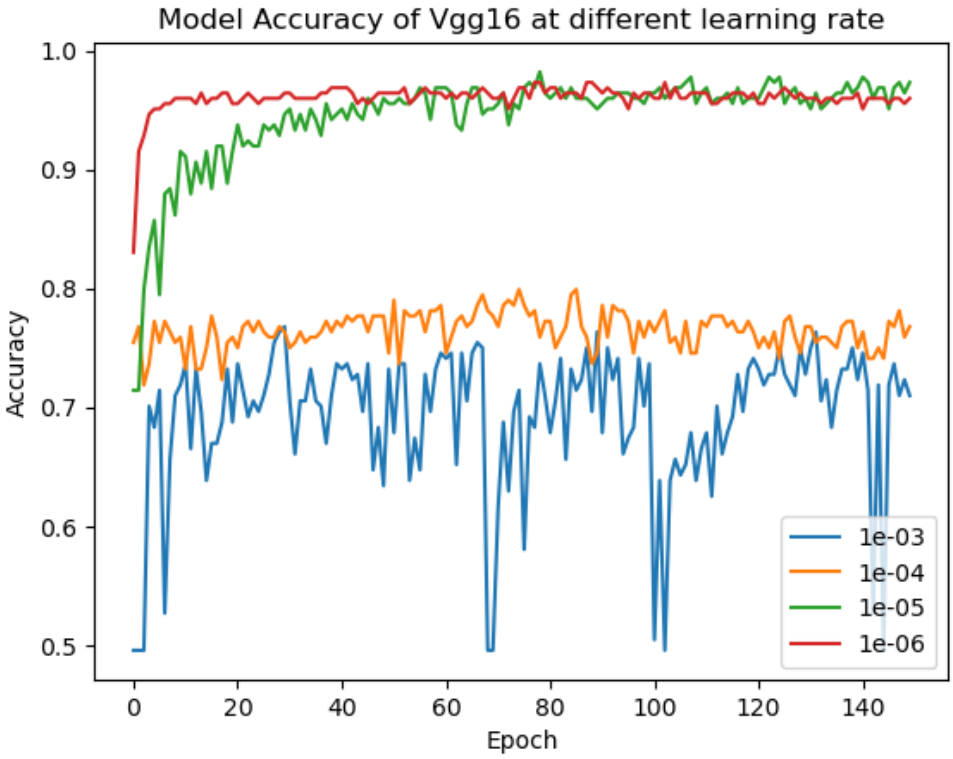


Fig 15. Validation accuracy of the model, using different learning rates.

Model Evaluation

To numerically measure the diagnostic performance of the deep learning model, the Receiver Operating Characteristic (ROC) curves and the area under ROC curve (AUC) were reported. Moreover, multiple evaluating criteria and a confusion matrix were employed to measure the performance of the model. A ROC curve is a graph that presents the diagnostic

ability of a binary classifying model with varying discrimination threshold. The ROC curve is created by plotting the true positive rate against the false positive rate. AUC (area under the ROC curve) denotes the measure of separability: capability of distinguishing between classes.

Confusion Matrix

n=332	**Predicted: Positive**	**Predicted: Negative**	
Actual: Positive	**TP**=162	**FN**=4	166
Actual: Negative	**FP**=3	**TN**=163	166
	165	167	

where true positive (TP) denotes the image has glaucoma and the prediction is positive, false positive (FP) denotes the image does not have glaucoma, but the prediction is positive, false negative (FN) denotes the image has glaucoma, but the prediction is negative, and true negative (TN) denotes the image does not have glaucoma and the prediction is negative.

Measure	Value	Derivations
Sensitivity	0.9818	TPR = TP / (TP + FN)
Specificity	0.9760	SPC = TN / (FP + TN)
Precision	0.9759	PPV = TP / (TP + FP)
Negative Predictive Value	0.9819	NPV = TN / (TN + FN)
False Positive Rate	0.0240	FPR = FP / (FP + TN)
False Discovery Rate	0.0241	FDR = FP / (FP + TP)
False Negative Rate	0.0182	FNR = FN / (FN + TP)
Accuracy	0.9789	ACC = (TP + TN) / (P + N)
F1 Score	0.9789	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient	0.9578	TP*TN - FP*FN / sqrt((TP+FP)* (TP+FN)*(TN+FP)*(TN+FN))

Fig 16. Measurements based on confusion matrix.

A ROC curve is a graph that presents the diagnostic ability of a binary classifying model with varying discrimination threshold. The ROC curve is created by plotting the true positive rate against the false positive rate. AUC (area under the ROC curve) denotes the measure of separability: capability of distinguishing between classes.

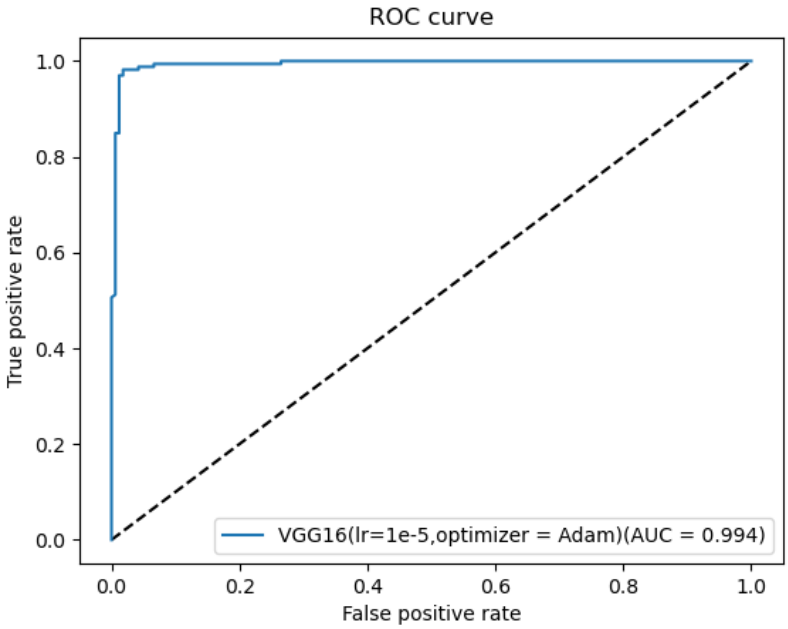


Fig 17. ROC curve of the proposed model with AUC=0.994

User Interface Demonstration

To offer public and compatible open-angle glaucoma classification access, a website application was designed based on HTML, CSS, JavaScript, and Python with frameworks and packages such as React, TensorFlow, and Flask. The website has been deployed on GitHub pages with its Flask API being served on Alibaba Cloud Server. The website is now active at www.glaucomark.com.

Given this classification website, patients can receive personal ophthalmic care through any network-connected technology. The process takes 45 seconds based on the device and no data will be saved.

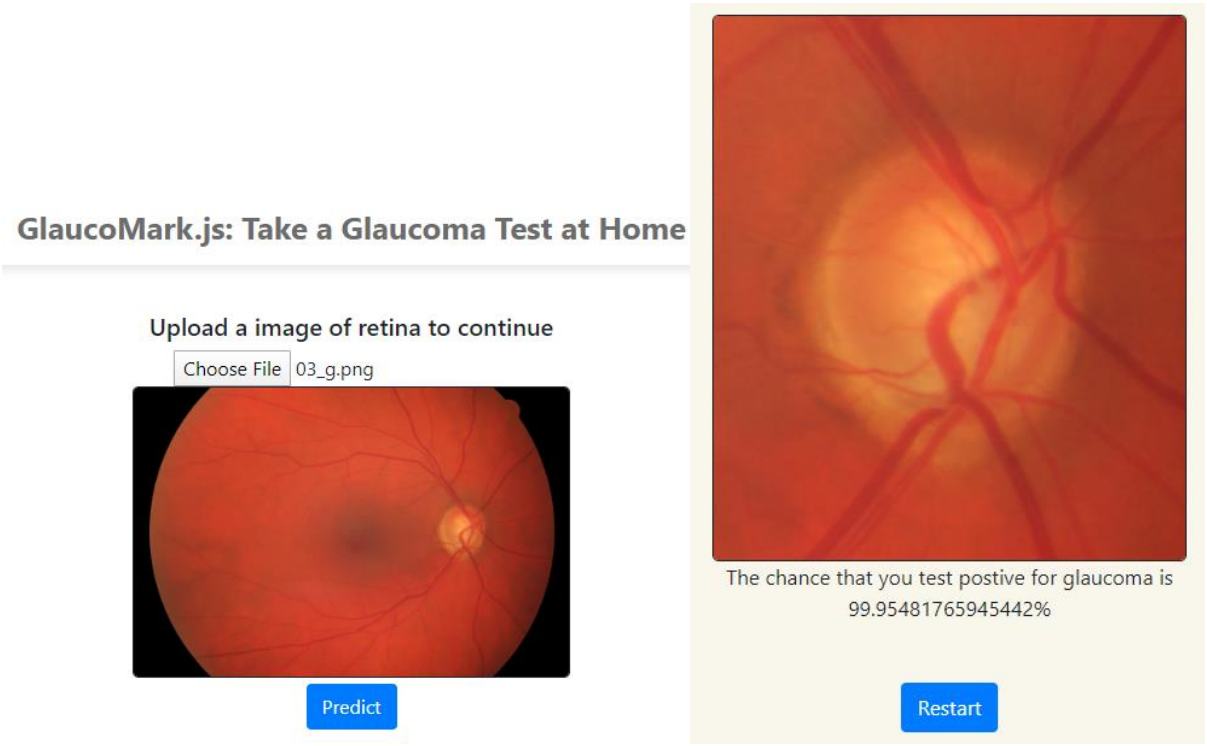


Fig 17, 18 The uploading page and result page of the web interface

Discussion

Model Training

The training process was suggested that the combination of object detection and image classification neural networks largely reduces the excessive computation of the detection, boosting the training speed and the diagnostic accuracy. In this study, the effectiveness of using two DL architectures in combination was investigated. The input image was preprocessed using an object detection network, following an image classification network to classify images based on the area of interest (optic disc). The powerful discrimination observed with the proposed method also reinforces the claim that glaucoma could be diagnosed by graphically analyzing the optic discs area on the retina.

Compared to other developed several recognition methods on automatic open-angle glaucoma detection, the AUC obtained in the current study - with the combined deep learning classifier - suggests far better discriminating and processing speed may be possible. Singh et al. proposed an innovative technique using features from segmented optic discs (with blood vessels removed) using 44 DFI and achieved an accuracy of 94.7% [8]. Chen et al. utilized a six-layer CNN based deep learning architecture and data augmentation and achieved accuracies of 83.1% and 88.7% on two different datasets [9]. Chakrabarty et al. used a newly proposed feature extraction technique, which reported an accuracy of 79.2% based on datasets of 2252 DFIs [10].

Vgg16 and Vgg19 approached relatively better performance than other architectures in this study. This could be explained by their significantly higher number of parameters. With more parameters, CNN will be able to capture more subtle information of the input image. Generally, training datasets with larger capacity would produce better performance during

CNN training. In this work, 3870 fundus images labeled with optic disc area were used to train the object detection model and 1662 fundus images were used to develop the object classification model which consists of sixteen layers. Although all of the seven deep learning architectures acquire similar classifying performance on ImageNet dataset, they acquired different performance specifically on the glaucoma datasets. Since the image of a healthy and glaucoma optic disc is hardly distinguishable by people without expertise, CNN structures with more parameters will outperform other CNN structures on capturing those subtle differences. In other words, DL architecture with more parameters performs well on classification between images with subtle differences. However, the extra amount of computation from the increased parameter size should also be considered.

Retina Imaging Device

To find a retina imaging method to incorporate with this glaucoma diagnosing system, a Raspberry Pi Based camera is being built based on the research of Shen et al (Shen, 2017). This camera utilizes both infrared and white lighting sources to locate and capture the photos through the naturally dilated pupils in a dark environment. This camera is pocket-sized, non-mydriatic, and inexpensive, costing about $1000 less than the typical retina imaging devices. The major advantage of the camera is that it's non-mydriatic, meaning it does not require pupil dilation before using. Dilating drops are inconvenient to use on oneself and can be dangerous in the much less common glaucoma patient that has narrow angles (shallow anterior chamber) * and dilation can cause a total blockage of the trabecular meshwork, the aqueous drainage system. This uncommon situation causes angle-closure glaucoma which leads to permanent loss of vision. Thus, by excluding the process of dilation, this camera can be operated safely by people without expertise (Shen et al ,2018).

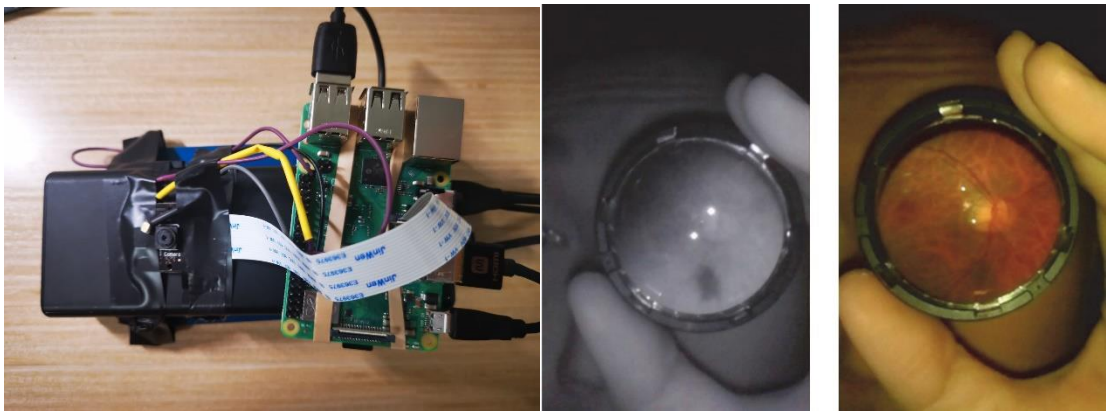


Fig 18 Demo of the Retina Camera: The raspberry-pi based camera(left) is able to take retina images without pupil dilation in dark where pupil naturally dilates. The image on the right shows the image captures under infrared lighting and white lighting. This design allows people with minimal expertise to operate the camera (Shen, 2018).

Application

Open-angle glaucoma is easy to treat but asymptomatic at its early stage. In China, 90 percent of patients with this disease are unaware of it and thus receive no treatment. Although the annual examination of open-angle glaucoma can prevent blindness, the costly and time-consuming natures of glaucoma screening undermine people's initiative and spontaneity of taking the examination annually, which consequently prevents glaucoma detection. Other than

that, in developing countries, ophthalmologists are hardly accessible, and most of the blind people there are blind from entirely preventable diseases, one of which is glaucoma.

In this study, a new glaucoma screening method was proposed by combining the diagnosing website with the low-cost and portable retina image camera. The proposed solution has the potential to implement large-scale glaucoma screening in developing countries. With a shallow learning curve, the camera enables volunteers without expertise to take a retina image of patients. The glaucoma examination can then be performed simply by uploading the retina image to the Glaucomark website.

The usage of these two technologies could enlarge the volunteers of only ophthalmologists to people who generally care about vision care. Doctors and nurses at local clinics in developing/rural counties could also be trained to use this camera and the website. With a large-scale glaucoma screening and the retina scans taken, ophthalmologists will be able to perform telemedicine to prevent the progression of glaucoma from leading towards blindness.

In the future, we intend to add a semantic segmentation network to determine the progression of glaucoma at different stages to facilitate glaucoma treatment. Furthermore. the model will be retrained on a dataset covering people of different ages, genders, and nationalities to ensure the reliability of the diagnosis.

Conclusion

Glaucoma is the second leading cause of blindness in the world, mostly in developing countries with limited access to vision care. The difficulty of open-angle glaucoma is centralized on early detection since open-angle glaucoma is asymptomatic until blindness but treatable at its early stage. Although the method of annual glaucoma examinations can effectively prevent its progression, this method is relatively unapproachable due to the lack of medical resources in developing/rural countries. This research developed a highly accurate and fast deep learning model to detect glaucoma with 97.9% accuracy. A website interface is developed to offer public glaucoma examination service that requires no expertise for use. A raspberry-pi based camera was built based on the research of Shen et al to take retina images as the input of the Glaucomark website. An AI-aided solution was also proposed to facilitate the glaucoma screening process in developing/rural countries.

References

[1] Quigley H A 2006 The number of people with glaucoma worldwide in 2010 and 2020 *British Journal of Ophthalmology* **90** 262–7

[2] Khouri A S and Fechtner R D 2015 Primary Open-Angle Glaucoma *Glaucoma* 333–45

[3] Oddone F, Virgili G, Parravano M, Brazzelli M, Novielli N and Michelessi M 2010 Optic nerve head and fibre layer imaging for diagnosing glaucoma *Cochrane Database of Systematic Reviews*

[4] O.d. M D H 1999 Optic disc size, an important consideration in the glaucoma evaluation *Clinical Eye and Vision Care* **11** 59–62

[5] Harizman N 2006 The ISNT Rule and Differentiation of Normal from Glaucomatous Eyes *Archives of Ophthalmology* **124** 1579

[6] Jonas J B 1992 Glaucomatous Parapapillary Atrophy *Archives of Ophthalmology* **110** 214

[7] Bengio Y, Courville A and Vincent P 2013 Representation Learning: A Review and New Perspectives *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** 1798–828
[8] Singh A, Dutta M K, Parthasarathi M, Uher V and Burget R 2016 Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image *Computer Methods and Programs in Biomedicine* **124** 108–20
[9] Chen X, Xu Y, Wong D W K, Wong T Y and Liu J 2015 Glaucoma detection based on deep convolutional neural network *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*
[10] Chakrabarty L, Joshi G D, Chakravarty A, Raman G V, Krishnadas S and Sivaswamy J 2016 Automated Detection of Glaucoma From Topographic Features of the Optic Nerve Head in Color Fundus Photographs *Journal of Glaucoma* **25** 590–7
[11] Anon *CS231n Convolutional Neural Networks for Visual Recognition*
[12] Sergey 2015 Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift *arXiv.org*
[13] Almazroa Retinal fundus images for glaucoma analysis: the RIGA dataset *Deep Blue Data*
[14] Anon www.ncbi.nlm.nih.gov
[15] Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi J M and Navea A 2019 CNNs for automatic glaucoma assessment using fundus images: an extensive validation *BioMedical Engineering OnLine* **18**
[16] Anon *RIM-ONE: An open retinal image database for optic nerve evaluation - IEEE Conference Publication*
[17] Anon cvblab - Overview *GitHub*
[18] Anon *Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation - IEEE Conference Publication*
[19] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y and Berg A C 2016 SSD: Single Shot MultiBox Detector *Computer Vision – ECCV 2016 Lecture Notes in Computer Science* 21–37
[20] Simonyan, Karen, Zisserman and Andrew 2015 Very Deep Convolutional Networks for Large-Scale Image Recognition *arXiv.org*
[22] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell. 2017;39(6):1137-1149.
[23] Skalicky S, Goldberg I. Depression and quality of life in patients with glaucoma: a cross-sectional analysis using the geriatric depression scale-15, assessment of function related to vision, and the glaucoma quality of life-15[J]. J Glaucoma, 2008, 17 (7): 546–551.
[24] Ding, Xiangyuan. "Dr.Wang's discussion regarding the situation of glaucoma diagnosis." N.p., 2015. Web. 20 July 2020. [25] "Cup-to-disc ratio not crucial to glaucoma documentation" Primary Care Optometry News, June 2005 Murray Fingeret, OD https://www.healio.com/optometry/glaucoma/news/print/primary-care-optometry-news/%7B431519a3-4481-4633-9567-2734ebc4bac5%7D/cup-to-disc-ratio-not-crucial-to-glaucoma-documentation
[25] Bailey Y. Shen, Shizuo Mukai. **A Portable, Inexpensive, Nonmydriatic Fundus Camera Based on the Raspberry Pi® Computer**. *Journal of Ophthalmology*, 2017; 2017: 1
DOI: 10.1155/2017/4526243

[26] "High Eye Pressure and Glaucoma", James C. Tsai, MD https://www.glaucoma.org/gleams/high-eye-pressure-and-glaucoma.php
[27] Redmon, Joseph, and Ali Farhadi. "YOLOv3: An Incremental Improvement." *ArXiv.org*. N.p., 08 Apr. 2018. Web. 20 July 2020.