

# Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models

Yidong Chai<sup>\*,a</sup>, Hongyan Liu<sup>a</sup>, Jie Xu<sup>b</sup>

<sup>a</sup> Research Center for Contemporary Management, School of Economics and Management, Tsinghua University, Beijing, 100084, China

<sup>b</sup> Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, 100084, China

## ARTICLE INFO

### Keywords:

Deep learning  
Disease diagnosis  
Convolutional neural networks  
Domain knowledge  
Glaucoma diagnosis  
Medical image analysis

## ABSTRACT

Glaucoma is one of the leading causes of blindness in the world and there is no cure for it yet. But it is very meaningful to detect it early as earlier detection makes it possible to stop further loss of visions. Although deep learning models have proved their advantages in natural image analysis, they usually rely on large datasets to learn to extract hidden features, thus limiting its application in medical areas where data is hard to get. Consequently, it is meaningful and challenging to design a deep learning model for disease diagnosis with relatively fewer data. In this paper, we study how to use deep learning model to combine domain knowledge with retinal fundus images for automatic glaucoma diagnosis. The domain knowledge includes measures important for glaucoma diagnosis and important region of the image which contains much information. To make full use of this domain knowledge and extract hidden features from image simultaneously, we design a multi-branch neural network (MB-NN) model with methods to automatically extract important areas of images and obtain domain knowledge features. We evaluate the effectiveness of the proposed model on real datasets and achieve an accuracy of 0.9151, sensitivity of 0.9233, and specificity of 0.9090, which is better than the state-of-the-art models.

## 1. Introduction

Glaucoma is a group of eye diseases that damage eye optic nerves and harm peripheral visions. Glaucoma can result in vision loss and blindness irreversibly [1,2]. In 2010, there were about 44.7 million people diagnosed with glaucoma around the world, and that figure is expected to rise up to 58.6 million by 2020 [3]. Open-angle glaucoma, the most common type of glaucoma, develops slowly without pain, making it difficult to detect before deterioration. Hence, it is called “silent theft of eyesight” by some experts [4]. The disease has some common symptoms including high intraocular pressure, optic nerve damage, large cup-to-disc ratio, and vision loss [5,6]. Currently, there is no cure for glaucoma, but early detection and treatment is essential to prevent more serious consequences (e.g., blindness). With early detection, an operation can be done to control intraocular pressure and avoid further vision deterioration [7].

In the hospital, images such as retinal images or fundus photographs are usually taken and interpreted manually to diagnose glaucoma. As the manual image interpretation is costly and prone to error, automatic glaucoma detection is an interesting and meaningful research topic [7]. Many scholars adopt computer vision algorithms for diagnosis. The

mainstream of these studies can be divided into two types. In the first type of the methods, researchers manually design features such as energy-based features [8], local configuration pattern (LCP) features [9], and higher order spectra (HOS) features [10]. Then classifiers are built based on these features for diagnosis. The second type of the methods extracts medical measures such as cup-to-disc ratio for diagnosis [7]. These features are calculated based on a segmentation procedure, where manually designed features are used to segment glaucoma-related tissues such as optic disc and optic cup [7].

Although many studies have been done for glaucoma diagnosis, manually designing features requires a considerable amount of engineering skill and domain expertise [11]. What is more, these features may oversimplify the problem and be ad hoc, for even experts may omit some important hidden patterns. Hence, deep learning models, which have the advantages to capture complicated hidden patterns from high-dimensional data and show superiority in many studies [11–15], have been introduced for glaucoma diagnosis [16,17]. These studies can also be divided into two streams. The first type of studies takes the entire retinal image as input and totally relies on deep learning models such as convolutional neural network (CNN) to extract features from images for glaucoma diagnosis [16,17]. For the second type, it makes use of

\* Corresponding author.

E-mail addresses: [chaiyd@yeah.net](mailto:chaiyd@yeah.net) (Y. Chai), [liuhy@sem.tsinghua.edu.cn](mailto:liuhy@sem.tsinghua.edu.cn) (H. Liu), [fionahsu920@foxmail.com](mailto:fionahsu920@foxmail.com) (J. Xu).

<https://doi.org/10.1016/j.knosys.2018.07.043>

Received 17 December 2017; Received in revised form 29 July 2018; Accepted 31 July 2018

Available online 31 July 2018

0950-7051/ © 2018 Elsevier B.V. All rights reserved.

domain knowledge and medical measures such as cup-to-disc ratio for diagnosis. Deep learning models are adopted to segment tissues in retinal images in order to calculate these measures [18,19]. These studies have achieved better results than traditional computer vision methods, and have shown meaningful explorations of using deep learning in glaucoma diagnosis. However, the first type of studies extracts features purely from data, which makes it rely on large dataset to learn useful features. What is more, these models omit domain knowledge and cannot output medical measures (e.g., cup-to-disc ratio), which doctors care a lot, because these measures can give reasons for diagnosis results. For the second type, though domain knowledge is used, diagnosis purely relies on medical features (e.g., cup-to-disc ratio). Useful hidden features that could be learned from deep learning models are neglected.

In this paper, we design a multi-branch neural network model to utilize both domain knowledge and hidden features automatically learned for glaucoma diagnosis. The domain knowledge lies in two ways. First, rather than only taking each global image as input, the model also incorporates a local important region. Second, not only structured features such as intraocular pressure, eyesight, age and symptoms are considered, but also features extracted from images such as cup-to-disc ratio, retinal nerve fiber layer defects (RNFLD), peripapillary atrophy (PPA), optic disc size and optic cup size in superior, nasal, inferior and temporal parts are exploited. In this way, we incorporate domain knowledge about glaucoma in this model, and to some extent, we also combine two types of glaucoma automatic detection methods. Based on traditional CNN models, we design a new deep learning model which we call multi-branch neural network model to combine different components together. The model outperforms both classical CNNs and traditional computer vision methods. Our proposed model can help doctors to make diagnosis faster and more accurate. In this sense, unlike other deep learning models which act like a black box and only tell you yes or no, our model can also generate some measures (e.g., cup-to-disc ratio, RNFLD and PPA) that are helpful for doctors to understand results and make decisions for diagnosis.

The novelty of work is that we propose a new deep learning model to combine domain knowledge and features learned from deep learning models to obtain better diagnosis results. To make use of domain knowledge, we proposed a framework to extract informative areas and obtain domain knowledge features automatically. Besides commonly used features such as cup-to-disc ratio, we also propose to extract RNFLD, PPA features for glaucoma diagnosis.

The remainder of the paper is organized as follows. In Section 2, we give a review of related work. Section 3 introduces our model in detail. Section 4 describes the data used. Then, we report the experiment setting and results in Section 5 and discuss the experiments in Section 6. We conclude the paper in Section 7.

## 2. Related work

### 2.1. Glaucoma diagnosis using traditional computer vision methods

Many researchers have studied how to diagnose glaucoma automatically based on retinal images. Early studies usually use traditional computer vision methods to extract manually designed features. These studies can be divided into two types.

The first type of methods is non-segmentation based one, which designs various features such as local configuration pattern (LCP) features [9], features from GIST descriptor [12], entropy-based features [20], energy-based features [8,21], fractal dimension (FD) features [22], higher order spectra (HOS) features [10,23,24], correntropy features [25], fast Fourier transform (FFT) features [26], and wavelet-based features [24,27]. Acharya et al. [23] calculated texture and HOS features from retinal images and trained an SVM classifier to predict glaucoma. Mookiah et al. [24] proposed an automated diagnosis system to distinguish normal between glaucomatous cases using HOS and

discrete wavelet transform features. Dua et al. [27] trained classifiers based on wavelet-based energy features, which were calculated from various wavelet filters. Yadav et al. [28] calculated texture features of the area around optic cup and trained a neural network classifier based on these features for glaucoma classification.

The second one is the segmentation-based approach, which generates common measures for glaucoma diagnosis based on a segmentation procedure, where researchers design features to segment important areas such as cup and disc. Nayak et al. [29] proposed methods to segment optic disc and cup through morphological operations to calculate the cup-to-disc ratio for glaucoma diagnosis. Babu and Shenbagadevi [30] proposed to use fuzzy c-means clustering on wavelet transformed green plane images to calculate the cup-to-disc ratio. Lee et al. [31] detected intensity decrease around optic disc for RNFLD, which is also a common measure of glaucoma. The segmentation-based approach usually needs to detect optic disc first, which is another research problem attracting many scholars. Chaudhuri et al. [32] proposed to detect optic disc region by thresholding out pixels with intensity values below a certain level. Lalonde et al. [33] proposed an optic disc detection method by combining a Hausdorff-based template matching technique and pyramidal decomposition for large-scale object tracking.

Although these studies achieve good results and show the feasibility of automatic glaucoma diagnosis, they rely heavily on manually designed features, which costs much energy and resource. What is more, the features are usually hard to design and ad hoc due to a variety of appearances. This problem can be solved by deep learning models.

### 2.2. Glaucoma diagnosis using deep learning models

Deep learning models have the ability to automatically learn complicated hidden patterns from high-dimensional data [11]. Since AlexNet was proposed in 2012 [12], many powerful models such as VGG16 net [13], GoogLeNet [14] and Deep Residual Net [15] have been proposed and demonstrated their advantages over traditional methods in image classification. Besides, deep learning models have also achieved state-of-the-art performance in a wide range of tasks such as object detection [34], semantic segmentation [35], natural language processing [36], and speech recognition [37].

Studies using deep learning models for glaucoma diagnosis can be separated into two types. The first type outputs the glaucoma diagnosis results directly through deep learning models. Chen et al. [16] trained a CNN on ORIGA and SCES datasets for glaucoma diagnosis and got better results than state-of-the-art algorithms. Raghavendra et al. [17] proposed an eighteen-layer CNN for glaucoma diagnosis. This type of work extracts features purely based on data, which makes it rely on large datasets to learn useful features. What is more, domain knowledge is neglected and medical measures are not available, which can be complementary to features from deep learning models and can give reasons for diagnosis results.

The second type of studies uses deep learning models to segment the glaucoma related tissues such as optic disc and optic cup, and then calculates medical measures (e.g., cup-to-disc ratio) for diagnosis [18,19]. Zilly et al. [18] proposed to segment optic cup and disc from retinal images using ensemble learning based on a CNN architecture and calculated the cup-to-disc ratio for glaucoma diagnosis. Although this type of work extracts some medical features (e.g., cup-to-disc ratio), there are still some other unexploited features (e.g., PPA size, RNFLD), which can improve diagnosis further. What is more, although deep learning features are adopted for segmentation, these useful features are omitted for diagnosis.

In addition, existing deep learning work evaluates their proposed methods based on a dataset which includes only glaucomatous cases and normal cases, while in real situations, doctors need to distinguish glaucoma from all other eye diseases, which is more difficult to achieve high accuracy. Therefore, more research should be conducted to

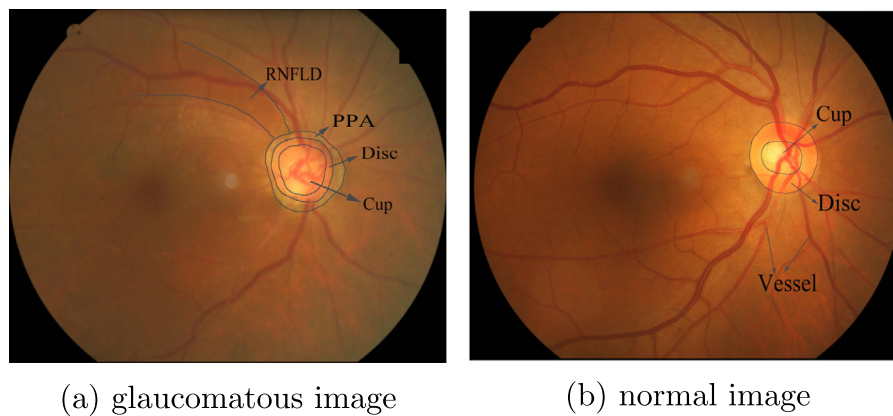


Fig. 1. Comparison between glaucoma and normal case.

explore the ability of deep learning models in distinguishing glaucoma from non-glaucoma.

In this paper, we aim to solve these issues and propose models to take the advantage of both deep learning models and domain knowledge. We conduct experiments on datasets collected from hospital containing cases of various eye diseases to evaluate the performance of deep learning models in glaucoma diagnosis.

### 3. Multi-branch neural network model

Glaucoma has some common performances including larger cup-to-disc ratio, RNFLD and PPA. Measure cup-to-disc ratio is the ratio of the diameter of optic cup region to the diameter of optic disc [38]. The optic disc is a light vertical oval in retinal images and the optic cup is a white, cup-like area in the center of the optic disc [38]. These two areas are shown in Fig. 1. The normal cup-to-disc ratio is between 0.3 and 0.5. When the ratio is above 0.5, it is very likely to have glaucoma [5]. Different from normal people, glaucoma and myopia patients usually have PPA, which is a green area around the optic disc, as shown in Fig. 1. RNFLD, a roughly wedge shape region starting from optic disc, is another major feature of glaucomatous images [6,39]. In Fig. 1, image (a) is a glaucomatous image, and image (b) is a normal case. As we can see, image (a) has a larger cup-to-disc ratio, which is around 0.8. Meanwhile, PPA and RNFLD also appear in the glaucomatous image.

The disc region is defined as the smallest rectangular box covering the optic disc and the cup. This region, though relatively small, contains much more discriminative information than others. Meanwhile, we cannot discard other regions as they also contain some useful information such as vasculature hemorrhage [1]. This inspires us to design a model that analyzes the entire image, while more attention should be paid to the optic disc region at the same time.

To utilize domain knowledge features and to analyze the disc area and the entire image simultaneously, we propose a multi-branch neural network model for automatic glaucoma detection. Fig. 2 shows the whole framework of the proposed model. Under this framework, glaucoma diagnosis is achieved in a fully automatic way. The first branch takes the entire image as input, and extracts features through a CNN. To get the input of the second branch, we use Faster-RCNN to obtain optic disc region. Then a CNN is used to extract local important features. For the third branch, we use fully convolutional network (FCN) model to segment disc area, cup area and PPA area, and then calculate measures. We also input RNFLD from a CNN and non-image features from case reports to the third branch.

#### 3.1. Optic disc region extraction

To get the optic disc region, we need to segment it from retinal images. Object detection techniques can be used to do that. Rather than

choosing traditional approaches [32,33], we adopt a deep learning model, Faster-RCNN [34], to do this. Faster-RCNN is a well-known object detection method proposed in 2015 [34]. It introduces a Region Proposal Network (RPN) to generate high-quality region proposals. The model RPN is a CNN that shares full-image convolutional features with the detection network [34]. Faster-RCNN is one of the state-of-the-art methods in the field of object detection. It accelerates the runtime by taking advantage of GPU for region proposals and the real-time object detection goal can be achieved [34].

#### 3.2. Domain knowledge features

##### 3.2.1. Image features

We design features including cup-to-disc ratio, PPA, RNFLD, disc and cup sizes in superior, nasal, inferior and temporal parts.

To obtain values of features about optic disc, cup and PPA, we need to segment these areas first. We adopt FCN [40], which is a powerful deep learning framework for image segmentation. FCN includes two parts: down-sampling and up-sampling, which aim to extract image features and output pixel labels for each class respectively [40].

To get the cup-to-disc ratio, we estimate the value through the root of the ratio of the two area sizes. The size is measured by the number of pixels in the area.

For optic disc and optic cup sizes, we divide disc and cup area into superior, nasal, inferior and temporal parts, and then obtain cup size and disc size of each part as features. The superior, nasal, inferior and temporal parts are shown in Fig. 3.

For the PPA measure, we use the size of PPA area as a feature, named as *PPA size*.

For the RNFLD feature, we use a CNN to detect if there is any RNFLD in a retinal image. We define this feature as *RNFLD*.

We summarize these features in Table 1.

##### 3.2.2. Non-image features

Features such as age, intraocular pressure, eyesight and symptoms are also extracted. We regard them as non-image features for simplicity.

For age, intraocular pressure and eyesight features, we just use the raw numerical values from case reports.

For symptoms, we process the patient complaint text and extract binary features such as “falling eyesight”, “eye hurts”, “swollen eye”, “blurred vision”, and “headache”, as shown in Table 2.

We deal with complaint text by designing a procedure to extract each symptom feature. If the complaint text includes a symptom, then the corresponding value is 1. Otherwise, it is 0. The procedure to extract the “headache” feature is shown as an example below.

For symptom features, we process them in a similar way. Finally, we generate 25 variables indicating various symptoms.

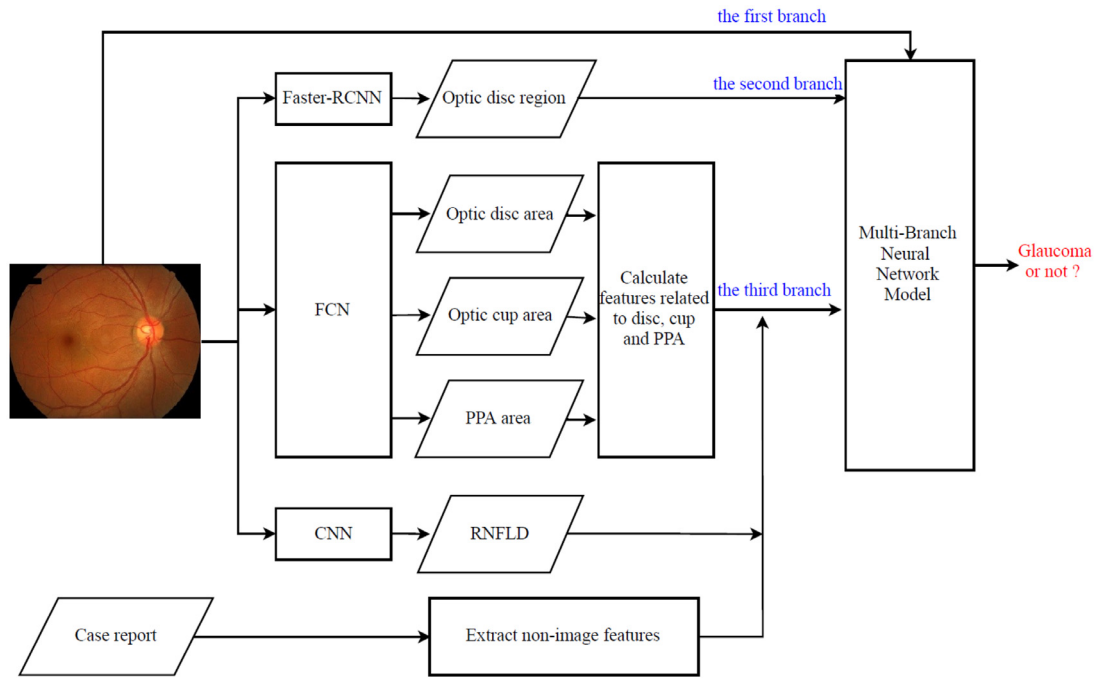


Fig. 2. Framework of our proposed model.

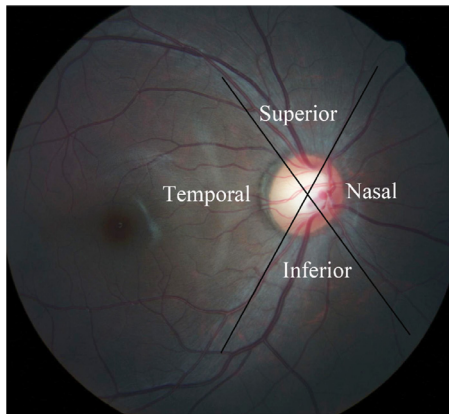


Fig. 3. Superior, nasal, inferior and temporal parts.

Table 2

Non-image features.

Feature	Data type	Description
age	Numerical	The age of the patient
intraocular pressure	Numerical	The intraocular pressure of the patient
eye sight	Numerical	The eye sight of the patient
falling eyesight	Binary	If the eyesight is falling
eye hurts	Binary	If the eye hurts
swollen eye	Binary	If the eye is swollen
swollen eyelid	Binary	If the eyelid is swollen
blurred vision	Binary	If the vision is blurred
headache	Binary	If headaches
relatives with glaucoma	Binary	If any of the patient's relatives has the glaucoma
red eye	Binary	If the eye is red
red eyelid	Binary	If the eyelid is red
itchy eye	Binary	If the eye is itchy
eye grinds	Binary	If the eye grinds
lacrimation	Binary	If the patient has the lacrimation
photophobia	Binary	If the patient has the photophobia
shadow	Binary	If there appears shadow in the vision
cataract	Binary	If the patient has the cataract
secretion	Binary	If the patient has the eye secretion
scieropia	Binary	If the patient has the scieropia
eye trauma	Binary	If the patient has the trauma in eye
diplopia	Binary	If the patient has the diplopia
impaired vision	Binary	If the vision is impaired
strabismus	Binary	If the patient has the strabismus
amblyopia	Binary	If the patient has the amblyopia
eye fatigability	Binary	If the eye feels fatigable
blindness	Binary	If the patient is blind
diabetes	Binary	If the patient has the diabetes

Table 1

Image features designed based on domain knowledge.

Feature	Type	Model
Cup-to-disc ratio	Numerical	FCN
Disc sizes in superior, nasal, inferior and temporal parts	Numerical	FCN
Cup sizes in superior, nasal, inferior and temporal parts	Numerical	FCN
RNFLD	Binary	CNN
PPA size	Numerical	FCN

### 3.3. Multi-branch neural network

To analyze the entire image, optic disc region and domain features simultaneously, we design a multi-branch neural network model. The first and the second branch of the model are CNNs to analyze the entire image and the disc region image respectively, and the third branch is a fully connected forward neural network to deal with domain knowledge features. Then we concatenate the outputs of these branches and add two fully connected layers (FC layers) for final prediction. The structure of the model is illustrated in Fig. 4. “C1: Feature maps 32, 5\*5\*3” means there are 32 kernels of size 5\*5\*3 in the first convolutional layer.

“FC (128)” means a fully connected layer with 128 nodes.

#### 3.3.1. Branches of the model

The first two branches are both CNN models. Each CNN model contains five convolutional layers, which applies a convolution operation to the input and then performs max pooling operation. The size of the kernel in the second branch model is 3\*3, for a small kernel size can reduce parameters and avoid overfitting [41]. On the other hand, for the first branch CNN model, the size of kernels in the first two



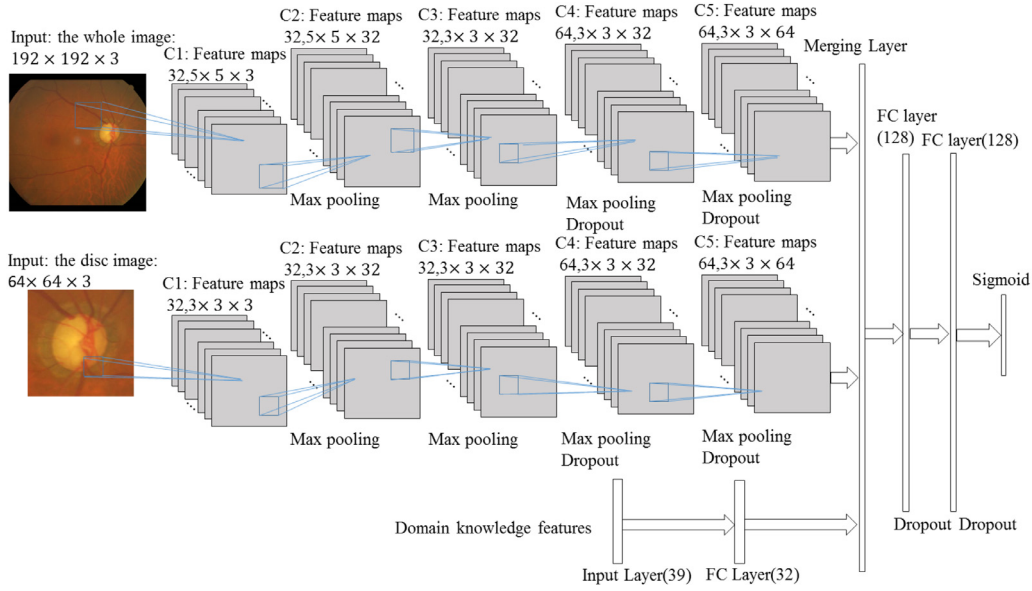


Fig. 4. Structure of the multi-branch model.

convolutional layers is  $5 \times 5$ , which is a relatively larger receptive field so that it can extract more global information. We also add dropout layer, where each node is “dropped out” of the net with probability 0.5. Max-pooling is a non-linear down-sampling technique that can reduce the number of parameters and the amount of computation in the network, and hence can control overfitting [12]. The stride size in the pooling operation is (2, 2) as a large size would lead too much information loss. In our designed network, Rectified Linear Units (ReLU) activation function is applied, which performs better than sigmoid function [19].

The third branch is a fully connected neural network to process the domain knowledge features including image features and non-image features described above. We also use ReLU activation function in this layer.

### 3.3.2. Merging layer

At the top of each branch, a merging layer is added to combine features extracted from different branches. The features from each branch are concatenated and reformed as a vector, and the merging layer concatenates these three vectors to form a new vector  $x_{merge}$  as shown in Eq. (1),

$$x_{merge} = Merge(x_1, x_2, x_3) \quad (1)$$

where  $x_1$ ,  $x_2$  and  $x_3$  represent the outputs of the first, second and third branch respectively.

### 3.3.3. Fully connected layer and logistic regression classifier

Two fully connected layers are added on the top of the merging layer, and the merged vector is used as input. The fully connected layers also have ReLU as activation function. We use dropout technique with dropout ratio 0.3 to overcome overfitting. The outputs of this two layers  $x_{fc}^1$ ,  $x_{fc}^2$  are calculated through Eq. (2) and (3) respectively.

$$x_{fc}^1 = dropout(max(W_{fc}^1 \cdot x_{merge} + b_{fc}^1, 0)) \quad (2)$$

$$x_{fc}^2 = dropout(max(W_{fc}^2 \cdot x_{fc}^1 + b_{fc}^2, 0)) \quad (3)$$

After that, sigmoid function is used to get the probability of label 0 and label 1, as shown in Eq. (4) and (5),

$$q(y = 0|x_{fc}^2; \theta) = \frac{1}{1 + \exp(-\theta^T \cdot x_{fc}^2)} \quad (4)$$

$$q(y = 1|x_{fc}^2; \theta) = 1 - q(y = 0|x_{fc}^2; \theta) = \frac{1}{1 + \exp(\theta^T \cdot x_{fc}^2)} \quad (5)$$

where  $q(y = 0|x_{fc}^2; \theta)$  is the probability of the label being 0, given input  $x_{fc}^2$  and the parameter  $\theta$ ;  $q(y = 1|x_{fc}^2; \theta)$  is for label 1. Then we calculate cross-entropy [42] as shown in Eq. (6),

$$H(p, q) = -p(0)\log q(0) - p(1)\log q(1) \quad (6)$$

where  $p$  is the true probability and  $q$  is the predicted probability. We use  $q(0)$  to represent  $q(y = 0|x_{fc}^2; \theta)$  and  $q(1)$  to represent  $q(y = 1|x_{fc}^2; \theta)$  for simplicity. We define loss function as the average of all cross-entropies in the sample. Suppose we have  $N$  samples, then the loss function is calculated through Eq. (7),

$$Loss = -\frac{1}{N} \sum_{i=1}^N (p_i(0)\log q_i(0) + p_i(1)\log q_i(1)) \quad (7)$$

We use back-propagation to get the derivatives of loss with the model parameters [43]. The back-propagation is a commonly used algorithm to train the model in neural networks. We follow the same procedure as used in other research [44]. In addition, we use Adadelata as the optimizer, for it is robust to the selection of hyper-parameters [45].

## 4. Data

The data we use to conduct experiments contains two parts. The first part is used to extract medical knowledge. The second one is used to train model for glaucoma diagnosis.

### 4.1. Data for domain knowledge feature generation

To extract the optic disc region, we trained Faster-RCNN model on a manually labeled dataset containing 2000 images.

To segment optic disc area, optic cup area and PPA area, we trained three FCN models based on three manually labeled datasets including 1000, 134 and 1000 retinal images respectively.

To obtain RNFLD feature value, we trained a CNN based on 2000 pre-labeled images from a hospital, including 831 images with RNFLD and 1169 without. As initial images were taken from different equipment, they have different sizes such as  $1924 \times 1556$ ,  $1572 \times 1308$ , and  $3048 \times 2432$ . Therefore, we need to equalize the size of these images. We first cropped images into squares, which were then resized to  $192 \times 192$ .

**Table 3**  
Dataset summary.

Dataset	Description	Number of instances	Task
D1	Retinal images with optic disc region labeled	2000	Optic disc region detection to get the input for the second branch
D2	Retinal images with labeled optic disc and PPA areas	1000	Optic disc and PPA area segmentation
D3	831 images with RNFLD labels, and 1169 without.	2000	To classify whether the image has RNFLD
D4	Retinal images with labeled optic cups	134	Optic cup segmentation
D5	Retinal images with diagnosis result, intraocular pressure, eye sight and complaint text	2554	To train the multi-branch neural network model

#### 4.2. Data for multi-branch neural network

We obtained 2554 retinal fundus images from 1542 patients, among which 1023 images are from glaucoma patients and 1531 are from people without glaucoma. These images had corresponding intraocular pressure, case reports and diagnosis results. Each patient might have various kinds of eye diseases such as cataract and diabetic retinopathy. For glaucomatous cases, there were 255 cases with cataract, 173 with diabetic retinopathy, 427 with myopia, and 142 with macular degeneration. For non-glaucomatous cases, there are 334 cases with cataract, 188 with diabetic retinopathy, 365 with myopia, and 137 with macular degeneration. We did not distinguish exact stages of glaucoma.

Similarly, we first cropped each image into a square according to the shorter side of the image. Then we resized the square image to 192\*192. The details of the datasets are summarized in Table 3.

### 5. Experiment

All the experiments were conducted on a workstation with Ubuntu 14.04.5 LTS, a 2.00 GHz Intel(R) Xeon(R) E5-2620 CPU, and a NVIDIA GeForce GTX 1080 GPU.

#### 5.1. Optic disc detection

We use D1 for optic disc detection task. These images are divided into training set and test set randomly with a ratio of 4:1. Then, we train Faster-RCNN based on framework of TensorFlow, which is a popular framework in the field of deep learning [46]. We adopt transfer learning to train the model. We use the VGG16 weights trained by the renowned Visual Geometry Group in ILSVRC 2014 for initialization [13]. We adopt the step learning policy for learning rate (basic learning rate is 0.001; gamma is 0.1; step size is 30000). We use SGD as optimizer with momentum being 0.9. We achieve a good performance on the test set. Among all the 400 retinal fundus images in the test set, only 8 images are not correctly detected because of relatively poor image quality. In Fig. 5, rectangles show optic disc regions extracted using our trained Faster-RCNN model.

As we can see, in the first three images, the red rectangles cover the disc region precisely and completely. Especially, the third image is not very clear, but the algorithm still works well. However, the algorithm can not find any disc region in the last image. In fact, this is a severe cataract case where the disc region is impossible to find even for experts. For the failed cases, we choose a region randomly as disc image for the input of the second branch.

#### 5.2. Generating domain knowledge features

To obtain RNFLD feature, we separate 2000 images of D3 randomly by a ratio of 4:1 to form the training set and the test set. After that, we fine-tune a CNN model, VGG16 [13], on TensorFlow. We adopt step learning policy for learning rate (basic learning rate is 0.000001; gamma is 0.1; step size is 1000) and SGD as optimizer with momentum 0.9. The accuracy of RNFLD detection is 90.44%.

For feature PPA size, we split the 1000 images of dataset D2 by a

ratio of 0.8 to form training set and test set. Then we train the FCN model based on TensorFlow. We adopt the Adam optimizer to train the model ( $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ). The learning rate is set as 0.00001. We achieve a satisfying result with an average precision of 84.15%. After that, we calculate the size of the PPA area as the value of feature PPA size.

For measures about optic disc and optic cup, datasets D2 and D4 are used to train FCN models [47] for optic disc area and optic cup area segmentation respectively. Both datasets are split by a ratio of 0.8 to form training set and test set. Then FCN models are trained based on TensorFlow. The parameters are the same as that in PPA area segmentation. We achieve a satisfying result with an average precision of 90.29% for optic disc and 84.15% for optic cup. Then we calculate cup-to-disc ratio, optic disc size and cup size in 4 parts.

Fig. 6 shows examples of optic disc area, optic cup area and PPA area segmentation results. The first column shows the initial images. The middle one shows the labels. The right one shows the results, where the green areas are the targets.

#### 5.3. Glaucoma diagnosis

We implement the multi-branch neural network (MB-NN) model based on Keras framework [48] with TensorFlow as backend. We adopt 5-fold cross-validation to assess the model. We also use data augmentation techniques including random crop, flipping images upside down, flipping them left to right and rotating them randomly. Adadelta optimizer is adopted (learning rate is 1.0; rho is 0.95; epsilon is None; decay is 0) to train the model. Early stopping technique (minimal delta is 0.0001; patience is 20) is used to avoid overfitting. For other deep learning models to be compared with, we adopt the same tricks.

To evaluate the performance of our proposed model, multi-branch neural network (MB-NN), we compare it with three benchmark sets. The first benchmark set is composed of traditional computer vision algorithms including logistic regression model based on manually designed features. The second benchmark set is composed of deep learning models including Alexnet [12], VGG16 [13] and inception\_v3 [41]. We include the third benchmark set including models with only two branches of the proposed model to test the usefulness of each branch.

##### 5.3.1. Benchmark set 1: Comparison with traditional computer vision algorithms

For traditional glaucoma diagnosis models, we trained logistic regression models based on rotation-invariant local binary patterns (LBP), higher order spectra (HOS), discrete wavelet transform (DWT) and scale-invariant feature transform (SIFT) features [24,27,49] respectively. We denote them as LR-LBP, LR-HOS, LR-DWT, and LR-SIFT respectively. The performances are listed in Table 4.

As we can see from Table 4, MB-NN achieves the best performance with an accuracy of 0.9151, sensitivity of 0.9233, and specificity of 0.9090, among models in benchmark set 1. We find that the results of models in the benchmark set are not good, with accuracies, sensitivities, and specificities all around 0.6. This may be because features in these models cannot represent patterns of glaucomatous and non-

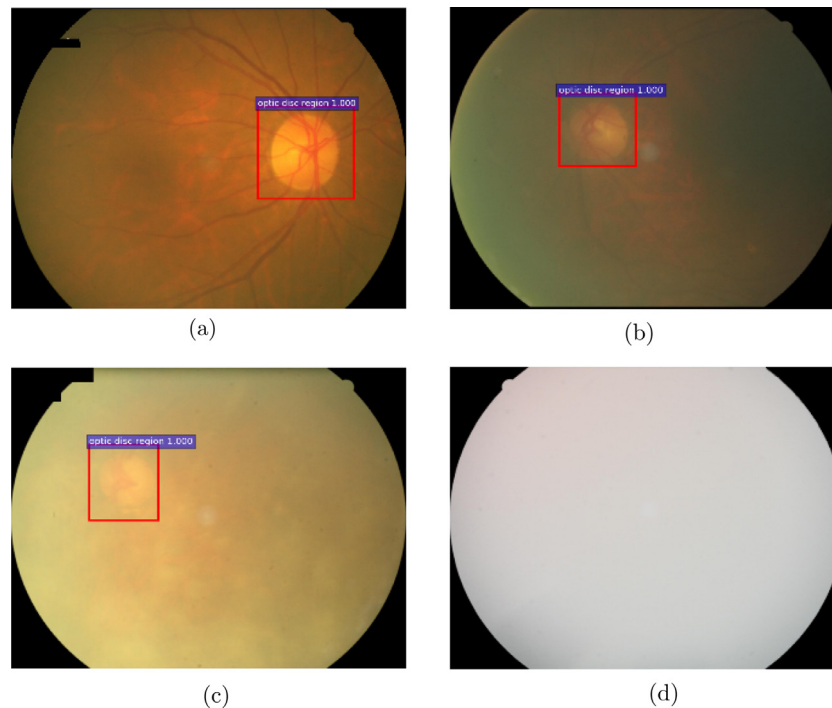


Fig. 5. Examples of extracted optic disc regions .

glaucomatous cases well.

### 5.3.2. Benchmark set 2: Comparison with other deep learning models

To compare the performance of our proposed model with other existing deep learning models, we implement four famous models: AlexNet, VGG16, inception\_v3, and a CNN model proposed by Raghavendra et al. [17]. The results are shown in Table 5.

In Table 5, we find the proposed model has achieved the best performance in accuracy, sensitivity and specificity among all the deep learning models in benchmark set 2. As the difference between our model and models in the benchmark set is that we adopt domain knowledge, this demonstrates the effectiveness of using domain knowledge for glaucoma diagnosis.

### 5.3.3. Benchmark set 3: Sensitivity analysis

To test the usefulness of each branch of the proposed model, we implement three two-branch models: MB-NN without inputting entire image, MB-NN without inputting disc area, and MB-NN without inputting domain knowledge features. For convenience, we call them MB-NN\_0, MB-NN\_1, and MB-NN\_2 respectively. The results are in Table 6.

As we can see from Table 6, the performances of all the two-branch models are inferior to the complete model. The MB-NN\_1 achieves an accuracy of 0.8631, sensitivity of 0.8597, specificity of 0.8657, obviously lower than MB-NN, which demonstrates the usefulness of features from the optic disc region. MB-NN\_2 performs the poorest among them with an accuracy of 0.8012, sensitivity of 0.8175, specificity of 0.7840. This shows that domain knowledge features can help improve performance obviously.

## 6. Discussion

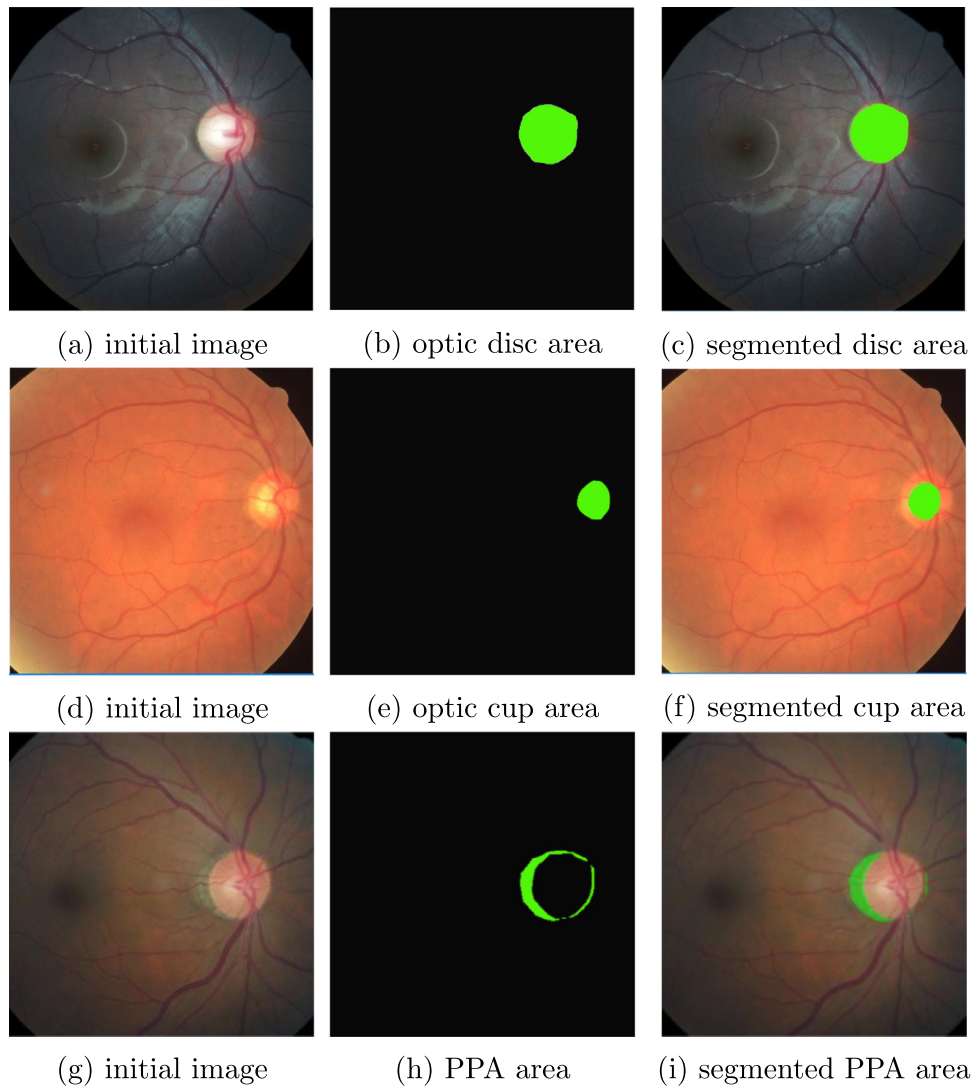
### 6.1. Model performance

Through experiments, we see that our proposed model performs better than the traditional computer vision models and other deep learning models. The diagnosis accuracies, sensitivities and specificities of traditional computer vision methods are only about 60%, much lower than the results obtained in some other papers that adopt the

same models [24,27]. This may be because the data is different. We collect the data from Tongren hospital, one of the best hospitals in China. The dataset includes many challenging and confusing glaucomatous cases, for Tongren hospital often receives patients with various intractable diseases from all over the country. Besides, not like datasets in other research which only contain glaucomatous and normal cases, our dataset includes non-glaucomatous cases which cover various diseases, thus making the patterns of non-glaucomatous cases more complicated for the model to learn. What is more, some of the non-glaucomatous cases share some similar symptoms with glaucoma, which means glaucomatous cases and non-glaucomatous cases share some hidden patterns. We show some non-glaucomatous cases in Fig. 7.

As we can see from Fig. 7, though these images are all non-glaucomatous cases and belong to the same label in the dataset, there are obvious differences between them on one hand. On the other hand, these negative cases share similarity with the positive cases in the meantime. For example, PPA appears in the high myopia case, which is also a common symptom for glaucoma, thus making the classification task harder. Besides, we cannot see some images clearly due to severe cataract. All these factors make the learning task hard, and thus compromise the performances of the model.

Meanwhile, we find the traditional deep learning models also perform poorly. Compared to the learning task of distinguishing dogs and cats, the learning task studied in this paper is more difficult, for the hidden medical patterns are more complicated. We need a much larger dataset to extract useful features and to achieve good results. Therefore, although the number of images in each category (2554 images for two categories) is larger than the ILSVRC 2012 (roughly 1000 images in each category) [12], the features learned are still not powerful enough to get the accuracy achieved in many open image classification tasks. Hence, for the practical glaucoma diagnosis, more images are necessary. However, it is much harder to get the labeled medical images than nature images, especially for rare diseases or scenarios where the medical images can only be taken using special or expensive equipment and labels can only be generated by few specialists. Therefore, how to utilize a relatively small dataset to obtain satisfying results is a common problem in medical research. In this circumstance, domain knowledge is useful in that we give more attention to the important area and we



**Fig. 6.** Examples of PPA, optic disc and optic cup segmentation results.

**Table 4**

Comparison between MB-NN and models in benchmark set 1.

Method	Accuracy	Sensitivity	Specificity
LR-LBP	0.6127	0.6233	0.6032
LR-HOS	0.5874	0.5837	0.5905
LR-DWT	0.6232	0.6152	0.6311
LR-SIFT	0.6078	0.6210	0.5950
<b>MB-NN</b>	<b>0.9151</b>	<b>0.9233</b>	<b>0.9090</b>

**Table 5**

Comparison between MB-NN and models in benchmark set 2.

Method	Accuracy	Sensitivity	Specificity
AlexNet	0.7522	0.7424	0.7626
VGG16	0.7832	0.7638	0.8033
Inception_v3	0.7791	0.7769	0.7814
CNN [17]	0.7032	0.7124	0.7011
<b>MB-NN</b>	<b>0.9151</b>	<b>0.9233</b>	<b>0.9090</b>

**Table 6**

Comparison between MB-NN and models in benchmark set 3.

Method	Accuracy	Sensitivity	Specificity
MB-NN_0	0.9013	0.9019	0.9008
MB-NN_1	0.8631	0.8597	0.8657
MB-NN_2	0.8012	0.8175	0.7840
<b>MB-NN</b>	<b>0.9151</b>	<b>0.9233</b>	<b>0.9090</b>

combine the domain knowledge features with deep learning models to achieve better results.

## 6.2. Medical implications

First, the model can help doctors to conduct the glaucoma risk assessment among citizens. Glaucoma is one of the main threats to human eyes [1,2]. Open-angle glaucoma, the most common type, develops slowly without pain, making it difficult to detect before deterioration. Therefore, it is necessary to conduct the glaucoma risk assessment among citizens to find suspicious cases. However, it is impossible to afford such activity due to the limitation of medical resources. The proposed model can perform automatic glaucoma diagnosis, which makes the assessment feasible.

Second, the proposed model can also assist doctors to provide better



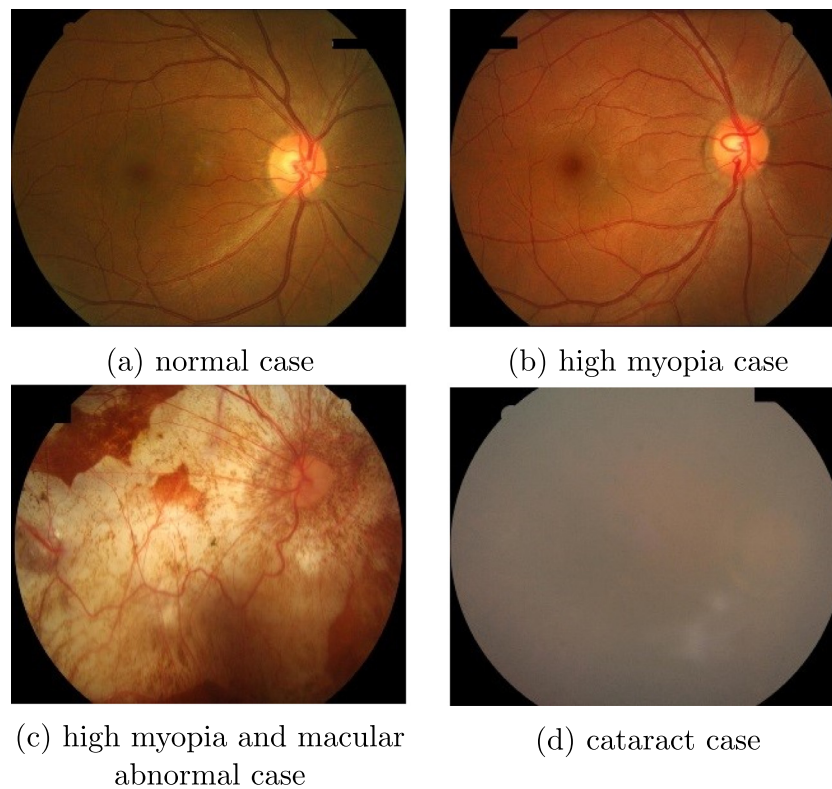


Fig. 7. Four non-glaucoma examples.

services. In the hospital, doctors need to manually interpret medical images, which is costly and prone to error, especially when they are tired. In this case, the proposed model can alleviate doctors' pressure in two ways. First, the diagnosis results of the proposed model can assist doctors to make accurate and fast decisions. Meanwhile, domain knowledge features extracted by our model can help doctors understand the diagnosis results and make better decisions. Therefore, the proposed model can offer an efficient and reliable assistance to doctors so that they can provide better services.

### 6.3. Research limitations

Research limitations lie in two ways. First, although we have got good results in distinguishing glaucoma and non-glaucoma, there is still room for further improvement. Second, the proposed framework needs to train several deep learning models to extract domain knowledge features, which needs several labeled datasets to train the model.

## 7. Conclusions

In this paper, we make use of medical domain knowledge to design a multi-branch neural network model for glaucoma diagnosis. As the optic disc region contains much more important information than other regions, we extract this region to deal with it separately. We also notice that doctors often use some measures to diagnose glaucoma. Therefore, we incorporate domain knowledge features such as cup-to-disc ratio, RNFLD, PPA size and symptoms from images and texts in the deep learning model. In this way, the proposed model combines latent features automatically extracted from images with domain knowledge features to improve both accuracy and interpretability. To automatically generate these domain knowledge features, we propose a framework to segment related areas, calculate feature values and then input them to the multi-branch model in a seamless way. Experiments conducted on real datasets show that our model outperforms models based on the traditional computer vision algorithms and the classical

convolutional neural network models.

Future research can be done in two ways. First, we plan to enlarge the experimental dataset to build a deeper model to improve the performance further. Second, we plan to design an FCN that can segment optic disc, cup and PPA at the same time, which can save computation resources further.

## Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (NSFC) with grant numbers 71432004, 71771131, 71490724, and 81700882. It is also supported by the priming scientific research foundation for the senior researcher in Beijing Tongren Hospital with grand number 2017-YJJ-GGL-009.

## References

- [1] P.N. Schacknow, J.R. Samples, *The glaucoma book: A practical, evidence-based approach to patient care*, Springer Science & Business Media, 2010.
- [2] N.V.M. Anand V Mantravadi MD, *Glaucoma, Primary Care: Clinics in Office Practice* 42 (3) (2015) 437–449.
- [3] H.A. Quigley, A.T. Broman, The number of people with glaucoma worldwide in 2010 and 2020, *Br. Ophthalmol.* 90 (3) (2006) 262–267.
- [4] J.C. Javitt, Preventing blindness in americans: the need for eyehealth education, *Surv. Ophthalmol.* 40 (1) (1995) 41–44.
- [5] J.B. Jonas, W.M. Budde, S. Panda-Jonas, *Ophthalmoscopic evaluation of the optic nerve head*, *Surv. Ophthalmol.* 43 (4) (1999) 293–320.
- [6] R.R.B.A.M.B.R.R.S.P.-J. Jost B Jonas Tin Aung, *Glaucoma, The Lancet logo* 390 (2017) 2083–2093.
- [7] M.S. Haleem, L. Han, J. van Hemert, B. Li, Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: a review, *Computerized medical imaging and graphics* 37 (7) (2013) 581–596.
- [8] U. Raghavendra, S.V. Bhandary, A. Gudigar, U.R. Acharya, Novel expert system for glaucoma identification using non-parametric spatial envelope energy spectrum with fundus images, *Biocybernetics and Biomedical Engineering* (2017).
- [9] U.R. Acharya, S. Bhat, J.E. Koh, S.V. Bhandary, H. Adeli, A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images, *Comput. Biol. Med.* 88 (2017) 72–83.
- [10] K.P. Noronha, U.R. Acharya, K.P. Nayak, R.J. Martis, S.V. Bhandary, Automated classification of glaucoma stages using higher order cumulant features, *Biomed*

- Signal Process Control 10 (2014) 174–183.
- [11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436.
  - [12] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, (2012), pp. 1097–1105.
  - [13] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
  - [14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 1–9.
  - [15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 770–778.
  - [16] X. Chen, Y. Xu, S. Yan, D.W.K. Wong, T.Y. Wong, J. Liu, Automatic feature learning for glaucoma detection based on deep learning, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 669–677.
  - [17] U. Raghavendra, H. Fujita, S.V. Bhandary, A. Gudigar, J.H. Tan, U.R. Acharya, Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images, *Inf. Sci. (Nij)* 441 (2018) 41–49.
  - [18] J. Zilly, J.M. Buhmann, D. Mahapatra, Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation, *Computerized Medical Imaging and Graphics* 55 (2017) 28–41.
  - [19] S.M. Shankaranarayana, K. Ram, K. Mitra, M. Sivaprakasam, Joint Optic Disc and Cup Segmentation Using Fully Convolutional and Adversarial Networks, *Fetal, Infant and Ophthalmic Medical Image Analysis*, Springer, 2017, pp. 168–176.
  - [20] S. Maheshwari, R.B. Pachori, V. Kanhangad, S.V. Bhandary, U.R. Acharya, Iterative variational mode decomposition based automated detection of glaucoma using fundus images, *Comput. Biol. Med.* 88 (2017) 142–149.
  - [21] U.R. Acharya, E. Ng, L.W.J. Eugene, K.P. Noronha, L.C. Min, K.P. Nayak, S.V. Bhandary, Decision support system for the glaucoma using gabor transformation, *Biomed. Signal Process Control* 15 (2015) 18–26.
  - [22] R. Kolář, J. Jan, Detection of glaucomatous eye via color fundus images using fractal dimensions, *Radioengineering* 17 (3) (2008) 109–114.
  - [23] U.R. Acharya, S. Dua, X. Du, C.K. Chua, et al., Automated diagnosis of glaucoma using texture and higher order spectra features, *IEEE Trans. Inf. Technol. Biomed.* 15 (3) (2011) 449–455.
  - [24] M.R.K. Mookiah, U.R. Acharya, C.M. Lim, A. Petznick, J.S. Suri, Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features, *Knowl. Based Syst.* 33 (2012) 73–82.
  - [25] S. Maheshwari, R.B. Pachori, U.R. Acharya, Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images, *IEEE J. Biomed. Health Inform.* 21 (3) (2017) 803–813.
  - [26] R. Bock, J. Meier, L.G. Nyúl, J. Hornegger, G. Michelson, Glaucoma risk index: automated glaucoma detection from color fundus images, *Med. Image Anal.* 14 (3) (2010) 471–481.
  - [27] S. Dua, U.R. Acharya, P. Chowriappa, S.V. Sree, Wavelet-based energy features for glaucomatous image classification, *IEEE Trans. Inf. Technol. Biomed.* 16 (1) (2012) 80–87.
  - [28] D. Yadav, M.P. Sarathi, M.K. Dutta, Classification of glaucoma based on texture features using neural networks, *Contemporary Computing (IC3)*, 2014 Seventh International Conference on, IEEE, 2014, pp. 109–112.
  - [29] J. Nayak, R. Acharya, P.S. Bhat, N. Shetty, T.-C. Lim, Automated diagnosis of glaucoma using digital fundus images, *J. Med. Syst.* 33 (5) (2009) 337.
  - [30] T.G. Babu, S. Shenbagadevi, Automatic detection of glaucoma using fundus image, *Eur. J. Scientific Res.* 59 (1) (2011) 22–32.
  - [31] S. Lee, K. Kim, J. Seo, D. Kim, H. Chung, K. Park, H. Kim, Automated quantification of retinal nerve fiber layer atrophy in fundus photograph, *Engineering in Medicine and Biology Society*, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, 1 IEEE, 2004, pp. 1241–1243.
  - [32] S. Chaudhuri, Automatic detection of the optic nerve in retinal images, *Proc. IEEE Int. Conf. Image Processing*, 1989, 15 (1989).
  - [33] M. Lalonde, M. Beaulieu, L. Gagnon, Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching, *IEEE Trans. Med. Imaging* 20 (11) (2001) 1193–1200.
  - [34] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems*, (2015), pp. 91–99.
  - [35] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1915–1929.
  - [36] H. Li, M.R. Min, Y. Ge, A. Kadav, A context-aware attention network for interactive question answering, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2017, pp. 927–935.
  - [37] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, *Acoustics, speech and signal processing (icassp)*, 2013 IEEE international conference on, IEEE, 2013, pp. 6645–6649.
  - [38] Y. Hatanaka, A. Noudo, C. Muramatsu, A. Sawada, T. Hara, T. Yamamoto, H. Fujita, Automatic measurement of cup to disc ratio based on line profile analysis in retinal images, *Engineering in Medicine and Biology Society, EMBC*, 2011 Annual International Conference of the IEEE, IEEE, 2011, pp. 3387–3390.
  - [39] V. Algazi, J. Keltner, C. Johnson, Computer analysis of the optic cup in glaucoma. *Investigative ophthalmology & visual science* 26 (12) (1985) 1759–1770.
  - [40] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 3431–3440.
  - [41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), pp. 2818–2826.
  - [42] P.-T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, *Ann. Oper. Res.* 134 (1) (2005) 19–67.
  - [43] Y. Chauvin, D.E. Rumelhart, *Backpropagation: Theory, architectures, and applications*, Psychology Press, 1995.
  - [44] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, Backpropagation applied to handwritten zip code recognition, *Neural Comput.* 1 (4) (1989) 541–551.
  - [45] M.D. Zeiler, Adadelta: an adaptive learning rate method, *arXiv preprint arXiv:1212.5701* (2012).
  - [46] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467* (2016).
  - [47] M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, R. Urtasun, Multinet: real-time joint semantic reasoning for autonomous driving, *arXiv preprint arXiv:1612.07695* (2016).
  - [48] F. Chollet, et al., *Keras*, 2015.
  - [49] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.